

Pascal Degiovanni,
Natacha Portier, Clément Cabart,
Alexandre Feller et
Benjamin Roussel

Physique quantique, information et calcul

Des concepts aux applications

SAVOIRS ACTUELS

EDP Sciences/CNRS ÉDITIONS

Dans la même collection

Calculs et algorithmes quantiques

David Mermin

Relativité restreinte

Ericourgoulhon

Physique quantique - Tomes 1 et 2

Michel Le Bellac

Comprenons-nous vraiment la mécanique quantique ? 2^e édition

Franck Laloë

Mécanique Quantique - Tomes 1, 2 et 3 - Nouvelle édition

Claude Cohen-Tannoudji, Bernard Diu et Franck Laloë

La théorie statistique des champs

François David

Retrouvez tous nos ouvrages et nos collections sur
<http://laboutique.edpsciences.fr>

Imprimé en France

© 2020, **EDP Sciences**, 17, avenue du Hoggar, BP 112, Parc d'activités de Courtabœuf, 91944 Les Ulis Cedex A

et

CNRS Éditions, 15, rue Malebranche, 75005 Paris.

Tous droits de traduction, d'adaptation et de reproduction par tous procédés réservés pour tous pays. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (art. L. 122-4, L. 122-5 et L. 335-2 du Code de la propriété intellectuelle). Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre français d'exploitation du droit de copie, 3, rue Hautefeuille, 75006 Paris. Tél. : 01 43 26 95 35.

EDP Sciences,

ISBN (papier) : 978-2-7598-1910-2, ISBN (ebook) : 978-2-7598-2413-7

CNRS Éditions,

ISBN (papier) : 978-2-271-13054-9, ISBN (ebook) : 978-2-271-13055-6

Préface

Nous vivons dans un monde d'information. Immédiatement disponible au point d'en être envahissante, immédiatement répandue à l'échelle planétaire, matière première et produit de l'économie des intelligences artificielles, elle contribue plus que toute autre chose à l'évolution de notre société. Pour le meilleur ou pour le pire ? Sans doute pour le meilleur et pour le pire, un « en même temps » typiquement quantique.

Les prouesses techniques des systèmes de transmission et de traitement de l'information reposent sur des travaux mathématiques remarquables qui, au milieu du siècle dernier, ne furent pas étrangers aux besoins de codage et, surtout, de décodage nécessités par la guerre. Les réflexions de Shannon sur la quantité d'information contenue dans un message, sur la possibilité de comprimer ce message sans (trop) perdre d'information expliquent comment la Télévision Numérique Terrestre parvient à faire passer 40 chaînes en haute définition là où la transmission analogique, sans compression, peinait à en faire passer 6. Les travaux de Turing et de ses successeurs sur la calculabilité sont à la base des systèmes de cryptographie (codes de cartes de crédit, par exemple). Il suffit d'imaginer un problème simple dont l'inverse est si complexe qu'il soit pratiquement insoluble (multiplication et factorisation, par exemple).

Mais ces avancées mathématiques n'auraient été que conceptuelles sans le développement exponentiel du *hardware*, de machines capables de réaliser des calculs complexes, infiniment plus puissantes que le balbutiant et encombrant ordinateur à lampes ENIAC (1945). Les circuits intégrés de ces machines reposent eux-mêmes sur une compréhension fine de la conduction électrique dans le silicium, que seule peut permettre la physique quantique.

Cette physique est née il y a près d'un siècle, de l'effort exceptionnel de quelques chercheurs pour comprendre la Nature au niveau des atomes et de molécules. Elle est devenue la grande aventure scientifique du XX^e siècle. Sous des formes diverses, elle couvre toute la physique connue, à la notable exception de la gravitation qui, avec la relativité générale, résiste encore et toujours au quantique. Elle s'applique aux particules élémentaires et aux plus grandes structures, restes des fluctuations quantiques de l'univers primordial. Elle atteint un niveau de précision sans précédent, avec un accord théorie-expérience sur 12 chiffres significatifs.

Elle est aussi à l'origine d'applications et de technologies qui ont modelé la société et son économie. La compréhension quantique de la conduction dans les solides conduit au transistor, puis au circuit intégré et enfin au smartphone. Des impulsions laser sont les véhicules de l'information dans les fibres optiques du réseau mondial. Les horloges atomiques du GPS nous localisent à quelques mètres près. Leurs descendantes sont si précises qu'elles n'auraient pas encore pris une minute de retard si nous avions pu les démarrer au moment du Big Bang. L'IRM, enfin, qui nous a fait gagner des années d'espérance de vie, repose sur la danse quantique des moments magnétiques nucléaires dans le champ d'un aimant supraconducteur, encore un phénomène authentiquement quantique. Une fraction importante de l'économie repose sur des « technologies quantiques », un exemple frappant de l'impact à très long terme, souvent méconnu hélas, de la science fondamentale.

En dépit de ses succès, la physique quantique reste difficile à « comprendre », pour paraphraser Feynman. Sa logique échappe en effet à notre esprit, forgé dans un monde où le quantique est voilé. Superpositions d'états (le « en même temps » quantique, illustré par le chat de Schrödinger, mort et vivant à la fois), intrication (deux objets quantiques qui ont interagi deviennent en quelque sorte inséparables) nous semblent étranges parce que nous ne pouvons les appréhender pleinement que par l'intermédiaire d'un formalisme parfaitement complet et cohérent mais parfois contraire au bon sens.

Les technologies issues de la physique quantique, lasers, ordinateurs, ont heureusement donné aux physiciens de nouveaux outils pour explorer le quantique, pour en comprendre les mécanismes les plus intimes. Ce renouveau d'intérêt pour la physique quantique a débuté avec la réalisation de quelques-unes des expériences de pensée proposées par les pères fondateurs pour asseoir leur interprétation du formalisme naissant. Interféromètres à fente mobile, illustrant la complémentarité chère à Bohr, paires de particules intriquées dans la situation du paradoxe Einstein Podolsky Rosen, chatons et chats de Schrödinger, mesures quantiques idéales, exploration de la frontière classique-quantique ont permis d'illustrer directement les postulats fondamentaux de la physique quantique et de confirmer les intuitions géniales de ses fondateurs.

Le degré de sophistication croissant de ces expériences a permis d'entrevoir, depuis une vingtaine d'années, ce qui est appelé — de façon peut être prématurée — la « seconde révolution quantique ». Ne peut-on utiliser la logique très particulière du monde quantique pour inventer de nouvelles méthodes de transmission et de traitement de l'information ? La physique quantique et la théorie de l'information ne peuvent-elles converger à nouveau et se renouveler l'une l'autre ?

La logique quantique permet par exemple de communiquer de façon inconditionnellement sûre par partage d'une clé dont le secret est garanti par les lois mêmes de la mesure quantique (on ne peut mesurer un système sans en changer l'état). Elle permet en principe aussi de calculer. Un ordinateur quantique manipulant au lieu des bits habituels (0 ou 1) des qubits (0 et 1

« en même temps ») pourrait faire interférer plusieurs calculs menés « en même temps » afin de résoudre certains problèmes comme la recherche au sein d'une liste non structurée ou bien la factorisation plus efficacement que n'importe quel algorithme classique connu. Toute la théorie de la calculabilité, toute la cryptographie se trouvent renouvelées par ce nouveau paradigme, même si la réalisation pratique, qui se heurte au terrible problème de la décohérence, est sans doute encore lointaine.

Le développement de ces nouvelles applications du quantique est l'objet de programmes majeurs (incluant le fameux *flagship* sur les technologies quantiques de l'Union Européenne). Des centaines de groupes, des milliers de chercheurs travaillent directement sur ce sujet. De façon remarquable, il réalise la convergence vers un but commun de domaines jusque-là aussi disparates que les pièges à ions, les atomes froids, les circuits supraconducteurs ou les centres colorés. En résulte une fertilisation, une émulation sans précédent entre champs de la physique et entre physique et théorie de l'information.

Maîtriser les concepts de l'information quantique, c'est donc maîtriser en même temps les concepts quantique et ceux de l'information, faire la synthèse entre Schrödinger et Shannon, Heisenberg et Turing. Bien sûr, chacune de ces disciplines est décrite séparément dans une pléthore de manuels. Mais aucun ne réalisait pleinement cette synthèse, jusqu'au présent volume.

Issu d'un enseignement inter-disciplinaire donné à l'ENS de Lyon, le livre de Pascal Degiovanni, Natacha Portier, Clément Cabart, Alexandre Feller et Benjamin Roussel réussit à présenter dans un cadre unique physique quantique et théories de l'information et de la calculabilité et à tisser les liens logiques entre elles. Chaque domaine est présenté de façon très claire, avec une approche pédagogique souvent originale. Dans l'exposé de la physique quantique, par exemple, le rôle essentiel donné aux intégrales de chemins et à la fonctionnelle d'influence pour la dynamique des systèmes ouverts, est suffisamment novateur pour qu'un (vieux) spécialiste du sujet puisse trouver un vif intérêt à sa lecture. Le lecteur sera bien sûr confronté avec des domaines bien au-delà de sa compétence habituelle. La qualité de l'exposé, les exemples clairs, la lecture à plusieurs niveaux (certains paragraphes techniques peuvent être omis en première approche) lui permettront de progresser rapidement et d'acquérir une réelle familiarité avec le sujet.

Au-delà d'un cadre théorique détaillé, ou rien de la nécessaire rigueur mathématique n'est oublié, l'avant dernier chapitre donne une introduction utile aux avancées expérimentales les plus significatives en information quantique. Ce manuel aura sûrement une place de choix dans la bibliothèque de tous ceux qui pratiquent au quotidien l'information quantique, mais aussi de tous ceux qui s'y intéressent tout simplement. Il sera aussi fort utile pour les enseignants, en leur proposant de nouvelles approches et de nouveaux exemples d'application.

La lecture attentive, crayon en main, que mérite ce livre nous permettra d'attendre le deuxième volume qui sera consacré à un programme encore plus ambitieux. Si la physique quantique fertilise la théorie de l'information, en

lui fournissant un nouveau paradigme logique, se peut-il que la théorie de l'information nous permette de formuler, de comprendre le quantique et le lien entre formalisme et monde réel d'une façon plus claire ? Et quelles pourraient en être les conséquences pour les problèmes encore ouverts comme la limite entre monde quantique et monde classique ou l'unification du quantique et de la gravitation ?

Jean-Michel Raimond
Professeur émérite, Sorbonne Université

Avant-propos

Do. Or do not. There is no try.

Yoda

L'histoire de ce livre est un peu comme une de ces séries américaines produites et diffusées par Netflix : une sorte de roman à tiroirs 2.0 avec beaucoup de personnages, d'intrigues et de rebondissements.

Tout commence par deux enseignants, Pascal D. et Natacha P., délivrant un cours en L3 à l'École normale supérieure de Lyon, il y a déjà de cela une dizaine d'années (2005-2006 et de 2011 à nos jours pour être précis). En apparence, rien de bien palpitant. Comme dans tout bon prologue sur Netflix, on y découvre la banalité confortable du quotidien : ici, la routine universitaire d'une École dite « normale » et prétendue « supérieure » selon l'expression consacrée. Les choses se compliquent quand on franchit la porte de l'amphi et que l'on découvre l'objectif du cours : « Enseignement interdisciplinaire en licence de Sciences de la Matière et en licence de Mathématiques et d'Informatique. L'objectif est de présenter les relations entre physique quantique, théorie de l'information et théorie du calcul qui sont à l'origine du développement de l'information et du calcul quantique. Il s'agit aussi de montrer comment ces relations interdisciplinaires ont conduit à une meilleure compréhension de la physique quantique et à une réflexion sur les fondements même de la théorie quantique. »

Vaste programme... Pour les curieux, l'intérêt d'un tel cours est évident. Tel enseignant vous dirait par exemple que les physiciens y trouveront l'occasion d'approfondir et de manipuler la théorie quantique sur un terrain un peu différent de celui auquel ils sont habitués. Tel autre que, pour les mathématiciens et les informaticiens, c'est l'occasion de découvrir la théorie quantique non pas à travers le prisme des manuels de physique mais avec une perspective plus proche de leurs propres intérêts. L'alléchant synopsis vous vend la possibilité de voir concrètement comment un nouveau champ disciplinaire a émergé à partir de l'apport de plusieurs domaines mais aussi de progrès technologiques et de questionnements fondamentaux comme plus appliqués.

L'étudiant, bien que jeune, n'en est pas pour autant tombé de la dernière pluie et comprend donc vite que derrière ces paroles rassurantes se cachent beaucoup d'enthousiasme, d'euphémismes et énormément de sous-entendus. La

souffrance sera certaine, la route longue et difficile, mais il entrevoit aussi une façon différente de faire et de penser la science, loin des cours rébarbatifs où le professeur monologue au tableau et où les étudiants bavardent entre eux. Plus tard, certains découvriront avec stupeur et tremblements la petite annonce passée par Shackleton pour ses expéditions polaires : « Hommes requis pour voyage périlleux, bas salaire, froid intense, longs mois de ténèbres, dangers constants, retour incertain. Honneur et célébrité en cas de succès. »

Mais nous n'en sommes qu'au début du pilote et il nous manque encore trois personnages à introduire. Pendant que les enseignants professaient, ce qui, selon la sagesse universitaire, consiste à éclairer de leur savoir comme d'une douce lumière tamisée des troupeaux de chèvres qui pensent juste à manger ou dormir (le cours est de 18 h à 20 h), trois étudiants du susnommé établissement, Clément C., Alexandre F. et Benjamin R., ont décidé d'écrire des notes de cours. Erreur de jeunesse ou égarement momentané causé par l'ennui considérable que leur procuraient certains des cours qu'ils suivaient à l'époque¹, cette idée est à l'origine des difficultés qui vont suivre. Comme pour le péché originel, l'histoire ne nous dit pas qui fut le fruit et qui fut le serpent.

Dans tout bon scénario, il doit y avoir à un moment un élément déclencheur, quelque chose qui démarre l'intrigue elle-même. Sur Netflix, c'est un meurtre sanglant, ici c'est une petite idée apparemment bien innocente : « Et si on écrivait un livre à partir des notes de cours ? ». Et là, c'est le début de la chute.

Comme dans toute première saison qui se respecte, la motivation des personnages est à son plein, leur naïveté totale. On part des notes de cours, on se dit qu'avec un peu de travail elles pourront être mises en forme pour faire un ouvrage sympathique et moderne. La rencontre avec Michèle Leduc lors d'une conférence à Lyon donne le coup d'envoi officiel de notre projet. On est en marche !

Le vrai drame a lieu au début des thèses de nos plus jeunes protagonistes. Un jour ensoleillé comme tant d'autres, avachis sur les sièges quelque peu spartiates du laboratoire, une discussion a lieu sur la structure de l'ouvrage. Une idée saugrenue vint alors : « Ce serait une vraie valeur ajoutée au livre que de faire une troisième partie sur les fondements de la théorie quantique (partie à l'époque assez embryonnaire), en mettant cela en relation avec les théories de l'information et du calcul. On pourra la structurer en mettant en évidence l'idée d'émergence à différents niveaux de complexité, détailler le côté relationnel et mettre cela en contexte avec d'autres domaines de la physique. » On ne marche plus, on court, on plonge du haut de la falaise. Avec le sourire, toujours.

De ce qui suivit, l'Histoire ne retiendra qu'une chose : il est étonnant que des scientifiques arrivent à faire des prédictions aussi éloignées de la réalité. Compléter cet ouvrage a pris six années de travail. Certes, les prédictions initiales étaient très optimistes mais pour aboutir au livre que vous tenez

1. À l'exception du cours susmentionné bien sûr...

entre vos mains, il a fallu une discipline de fer : lectures de papiers, groupe de travail hebdomadaire tous les mercredis, multiples *book camps*² chez Pascal D. pendant les week-ends, relectures, conversations téléphoniques, changements de perspectives, nouvelles relectures, rendez-vous avec l'éditeur, relectures, changement de dernière minute, etc.

De ces joies et souffrances naquirent deux volumes de quatre cents pages (à peu près) au lieu d'un seul. Dans le paysage de la littérature scientifique, son statut est quelque peu exotique : ce n'est ni un livre de cours complet sur la théorie quantique ou sur l'informatique quantique, ce n'est pas non plus un livre sur les fondements. C'est un livre de physique qui aborde un certain nombre de questions en traversant les barrières disciplinaires. C'est une introduction aux technologies quantiques et, en même temps, une réflexion sur les concepts qui sous-tendent la théorie quantique, l'espace, le temps, la connaissance.

Comment se finit cette histoire sachant que beaucoup de questions restent ouvertes ? Connaissant le penchant naturel des protagonistes de cette histoire pour les visions grandioses dignes des plus beaux trips sous substances psychoactives, le suspense est total. En tout cas, c'est une bonne question et nous vous remercions de l'avoir posée. D'ailleurs, si vous lisez les deux volumes de cet ouvrage, vous découvrirez peut-être une réponse.

Remerciements

Bien entendu, cet ouvrage n'aurait pu voir le jour sans le soutien d'un certain nombre de personnes.

Naturellement, Michèle Leduc et Michel Le Bellac, dont l'enthousiasme pour notre projet ne s'est jamais démenti, alors même que nous accumulions mois de retard et pages en trop. Avec constance, ils nous ont fait bénéficier de leurs conseils avisés et de leur grande expérience scientifique. Nous espérons avoir été à la hauteur des espoirs que nous avons suscités chez eux. Souhaitons que les ventes soient inversement proportionnelles aux abîmes de perplexité dans lesquels nous les avons sûrement plongés.

Plusieurs collègues ont accepté la difficile tâche de relecteur de ce volume. Un grand merci à Jean-Michel Raimond, Michel Le Bellac et Michel Fruchart pour leurs relectures complètes et approfondies de la totalité du manuscrit. En plus d'avoir relu de la première à la dernière page ce manuscrit, Jean-Michel Raimond nous a fait l'honneur de préfacer ce volume : qu'il en soit chaleureusement remercié.

Nous remercions aussi Daniel Estève, Walter Appel, Omar Fawzi, Charles Grenier, Frédéric Magnez et Pierre Rouchon qui nous firent part de leurs retours et corrections sur diverses parties du manuscrit. À la lumière de leurs expertises complémentaires, ils nous suggérèrent de nombreuses corrections

2. Séance d'écriture intensive sur deux jours entrecoupée de raclette et autres plats pour tenir la distance sur un doux fond de musique Eurobeat.

qui nous ont permis d'améliorer cet ouvrage tant sur le plan du contenu scientifique que sur le plan de la pédagogie. Nous tenons à les en remercier chaleureusement.

Bien sûr, nous avons bénéficié de conseils éclairés de nombreux autres collègues à divers stades du projet. Même s'il est impossible d'en faire la liste exhaustive, nous remercions particulièrement Scott Aaronson, Jean Dalibard et Mark Wilde pour leurs suggestions et pour nous avoir encouragés à mener à bien ce projet dans toute son ambition.

L'écriture étant une nourriture exclusivement spirituelle, nous remercions nos employeurs en France (CNRS, ENS de Lyon et Université Claude Bernard) de nous avoir financés et fourni d'excellentes conditions de travail durant la rédaction de cet ouvrage. Enfin, lorsque deux d'entre nous s'envolèrent vers d'autres cieux, l'Agence spatiale européenne a permis à cette aventure de continuer au sein de l'Advanced Concepts Team à Noordwijk où de nouvelles problématiques de recherches directement issues du travail sur cet ouvrage purent éclore et se développer.

Bien avant que l'idée de faire un livre ait germé, Sébastien Manneville et Éric Fleury apportèrent leur soutien au projet d'un cours optionnel commun aux Licences de Sciences de la Matière et de Mathématique et Informatique de l'École normale supérieure de Lyon. Ce fut l'étincelle initiale sans laquelle ce projet n'aurait jamais vu le jour.

Nous ne saurions d'ailleurs que trop remercier les générations d'étudiants de L3, mais aussi de M1, de M2 qui suivirent cet enseignement avec abnégation (toujours), curiosité (souvent) et passion (parfois), ainsi que les doctorants chargés des tutorats. Leurs questions souvent profondes, leurs critiques évidemment impitoyables et leur participation enthousiaste aux exposés que nous propositions en validation de cet enseignement ont fortement contribué à faire avancer notre réflexion et le développement de cet ouvrage.

Enfin, même si un proverbe dit que beaucoup remercier signifie secrètement demander davantage, un grand merci à nos familles et à nos proches pour leur soutien et leur accompagnement au cours de ce long voyage.

Volume I

Préface	iii
Avant-propos	vii
Introduction	1
Structure de l'ouvrage	5
1 Théorie quantique	9
1.1 La formulation de Feynman	10
1.1.1 Statut de la théorie quantique	10
1.1.2 Les règles de Feynman	11
1.2 Applications et conséquences physiques	18
1.2.1 Le système à deux niveaux	18
1.2.2 La particule libre	20
1.2.3 La limite classique	22
1.2.4 L'effet Aharonov-Bohm	23
1.2.5 Particules indiscernables	27
1.2.6  Physique quantique et topologie	30
1.3 La formulation hamiltonienne	32
1.3.1 Notion d'état quantique et linéarité	32
1.3.2 Produit scalaire et espace de Hilbert	35
1.3.3 Évolution temporelle	37
1.3.4 Retour sur le système à deux niveaux	41
1.4 La mesure en physique quantique	42
1.4.1 Mesures projectives	42
1.4.2 Observables et indétermination de Heisenberg	44
1.5 Perspectives	46
2 Théorie de l'information	49
2.1 Les systèmes de communication	50
2.2 L'entropie de Shannon	52
2.2.1 La mesure de l'information	52
2.2.2 L'entropie de Shannon	54

2.2.3	Premier théorème de Shannon sur le codage	55
2.3	Propriétés de l'entropie de Shannon	62
2.3.1	Extraction d'information	62
2.3.2	Cas d'une source composée	63
2.3.3	Multiplexeur	65
2.3.4	✂ La divergence de Kullback-Leibler	67
2.3.5	✂ Entropie et dynamique stochastique	68
2.4	Bayes et la théorie de l'information	68
2.4.1	Une autre conception des probabilités	68
2.4.2	Exemples d'utilisation de l'inférence bayésienne	70
2.4.3	Information mutuelle et inférence bayésienne	73
2.5	Capacité de transmission d'un canal	78
2.5.1	Canal de transmission	78
2.5.2	Capacité de transmission d'un canal	80
2.5.3	Second théorème de Shannon sur le codage	81
2.5.4	✂ La borne de Slepian-Wolf	82
3	Calculabilité et complexité	85
3.1	Avec quoi calcule-t-on ?	86
3.1.1	Du transistor à la fonction	87
3.1.2	Machine de Turing	91
3.2	Calculable et incalculable	97
3.2.1	Problèmes indécidables	98
3.2.2	Variantes des machines de Turing	102
3.2.3	Autres modèles de calcul	105
3.2.4	✂ Hypercalcul	106
3.3	Les thèses de Church, Turing, Landauer et Deutsch	110
3.3.1	La thèse de Church-Turing	111
3.3.2	La thèse de Landauer	111
3.3.3	La thèse de Church-Turing-Deutsch	112
3.3.4	La question des ressources	113
3.4	Complexité	114
3.4.1	Complexité algorithmique	114
3.4.2	Classes de complexité	117
3.4.3	P <i>versus</i> NP	124
3.4.4	L'aléatoire ou le rôle du hasard en complexité	127
3.4.5	Thèse de Church-Turing étendue	133
4	Systèmes quantiques simples	135
4.1	Systèmes à deux niveaux	136
4.1.1	De l'état au vecteur de Bloch	136
4.1.2	Du vecteur de Bloch à l'état	138
4.1.3	Les transformations du qubit	139
4.1.4	La polarisation du photon	140
4.1.5	Le spin 1/2	146

4.2	Oscillateur harmonique quantique	150
4.2.1	États propres et énergies propres	151
4.2.2	États quasi classiques	152
4.2.3	Photons dans une cavité	155
4.3	Tomographie quantique du qubit	158
4.3.1	Reconstruire $ \psi\rangle$ avec une infinité de réalisations	158
4.3.2	Le cas d'une source imparfaite	159
4.3.3	Nombre fini de réalisations	160
5	États intriqués : définition et propriétés générales	167
5.1	Systèmes composés	168
5.1.1	Qu'est-ce qu'un système composé ?	168
5.1.2	Les états dans un système composé	169
5.2	Les états intriqués et leurs symétries	176
5.2.1	La décomposition de Schmidt des états intriqués	177
5.2.2	La décomposition de Schmidt est-elle unique ?	178
5.2.3	États intriqués et corrélations de mesure	182
5.3	Intrication et mélange	183
5.3.1	Sources imparfaites et purification	183
5.3.2	Théorème de Gisin, Hughston, Jozsa et Wootters	184
5.3.3	Mélanges et théorie de l'information	187
5.4	Mesures et intrication	188
5.4.1	Qu'est-ce qu'une mesure généralisée ?	189
5.4.2	Mesures faibles <i>versus</i> mesures projectives	192
5.4.3	Quelle information est extraite du système ?	195
5.4.4	✗ Approche perturbative	199
6	Impossibilités quantiques	203
6.1	Indiscernabilité des états non orthogonaux	203
6.1.1	Indiscernabilité de deux états non orthogonaux	204
6.1.2	Ce que l'indiscernabilité ne dit pas	205
6.2	Théorème de non-clonage	205
6.2.1	Le théorème de non-clonage	206
6.2.2	Ce que le non-clonage ne dit pas	207
6.2.3	Non-clonage et indiscernabilité	208
6.3	Théorème de non-effacement	209
6.3.1	Le théorème de non-effacement	209
6.3.2	Le problème de l'effacement	211
6.4	Impossibilités non locales	212
6.4.1	Le théorème de non-communication	213
6.4.2	Le théorème d'inconvertibilité	214
6.4.3	Le théorème de non-dissimulation	215
6.5	Discernabilité d'états non orthogonaux	217
6.5.1	Discrimination sans erreur	217
6.5.2	Discrimination sans ambiguïté mais avec erreur	219

6.5.3	Comparaison des deux stratégies	221
6.5.4	Application aux télécommunications	221
7	Communiquer en utilisant des qubits	227
7.1	Communiquer quantiquement	228
7.1.1	Motivation	228
7.1.2	Le canal quantique	229
7.2	Les protocoles quantiques élémentaires	230
7.2.1	Le partage de secret au moyen d'états quantiques . . .	230
7.2.2	Utiliser l'intrication pour communiquer	237
7.2.3	Échange d'intrication	242
7.2.4	Ressources et protocoles	244
7.3	Contenu en information d'un qubit	250
7.3.1	Bits classiques <i>versus</i> qubits	251
7.3.2	La borne de Holevo	252
7.3.3	Le théorème de codage de Schumacher	256
7.4	Propriétés de l'entropie de von Neumann	259
7.5	Conditionnement en théorie quantique	262
7.5.1	Conditionnement classique en théorie quantique . . .	262
7.5.2	Entropie conditionnelle quantique	263
7.6	La théorie de Shannon quantique	267
7.6.1	Information conditionnelle et communication	269
7.6.2	Le transfert d'état quantique	274
7.6.3	Conséquences du protocole de transfert	277
7.6.4	Protocoles duaux	282
7.6.5	Capacité d'un canal quantique	285
7.6.6	Perspectives	290
8	Calculer en utilisant des qubits	293
8.1	Oracles et accélération quantique prouvée	295
8.1.1	Trouver si une fonction est constante	295
8.1.2	Chercher dans une base de données non triée	302
8.2	Une accélération exponentielle ?	311
8.2.1	Le problème de Simon	312
8.2.2	La transformée de Fourier quantique	313
8.2.3	L'estimation de phase quantique	317
8.2.4	Décomposer un nombre en facteurs premiers	322
8.2.5	Au-delà de l'algorithme de Shor	326
8.3	Complexité et théorie quantique	329
8.3.1	Machine de Turing quantique	330
8.3.2	La classe BQP	331
8.3.3	Thèse de Church-Turing étendue quantique	334
8.3.4	Les problèmes dans BQP ont-ils un intérêt ?	335
8.4	Circuits quantiques	336
8.4.1	Circuits classiques et complexité	337

8.4.2	Circuits quantiques et complexité	338
8.4.3	Codes correcteurs d'erreurs	346
8.5	✂ La simulation quantique	355
8.5.1	✂ Enjeux et difficultés de la simulation quantique	355
8.5.2	✂ Le calcul adiabatique	359
9	Dynamique des systèmes quantiques ouverts	365
9.1	Les interférences quantiques en pratique	366
9.2	Empreintes quantiques et décohérence	368
9.2.1	Empreintes et intrication	369
9.2.2	La fonctionnelle d'influence de Feynman-Vernon	370
9.3	Trajectoires quantiques	371
9.3.1	Motivations	371
9.3.2	Les états relatifs	372
9.4	Le cas markovien	375
9.4.1	Le modèle d'environnement markovien	375
9.4.2	L'équation de Lindblad	377
9.4.3	L'équation de Schrödinger stochastique	378
9.5	Super-opérateurs et markoviannité	380
9.5.1	Le théorème de Kraus	380
9.5.2	Cas markoviens et non markoviens	383
9.5.3	Conditions de validité de l'approche markovienne	388
9.5.4	✂ Vers une équation de type Lindblad	389
10	Vers une ingénierie quantique	395
10.1	Atomes en cavité	395
10.1.1	Dispositif expérimental et modélisation	396
10.1.2	Les trajectoires quantiques	399
10.1.3	Tomographie d'un champ en cavité	406
10.1.4	Fuites de photons	408
10.1.5	Études expérimentales de la décohérence	415
10.2	Électrodynamique quantique sur circuit	417
10.2.1	Circuits électriques quantiques	417
10.2.2	Circuits supraconducteurs simples	419
10.2.3	Créer des qubits supraconducteurs	426
10.2.4	L'électrodynamique sur circuit	429
10.2.5	Les trajectoires quantiques du qubit	433
10.2.6	Les expériences de suivi des trajectoires	444
10.2.7	✂ Applications des circuits supraconducteurs	445
11	Caractériser les corrélations quantiques	449
11.1	Le raisonnement EPR	450
11.1.1	Hypothèses	450
11.1.2	Le raisonnement	450
11.2	L'inégalité de Bell	451

11.2.1	Établissement de l'inégalité	452
11.2.2	Prédictions de la théorie quantique	453
11.3	Les tests expérimentaux du réalisme local	456
11.3.1	Historique	456
11.3.2	Les difficultés expérimentales	456
11.4	Compléments sur les inégalités	459
11.4.1	Le cas d'un état mélange	459
11.4.2	Le cas d'un état pur général	460
11.4.3	Théorème de Tsirelson	461
11.4.4	Contradiction GHZ	463
11.5	Conséquences conceptuelles	464
Conclusion		469
Bibliographie		473
Index		489

Introduction

À la veille du centenaire de la théorie quantique, son succès ne s'est pas démenti : cette théorie décrit un vaste éventail de phénomènes qui vont de la physique des particules élémentaires aux fluctuations du rayonnement cosmologique en passant par les échelles atomiques, moléculaires et même macroscopiques avec les superfluides, les supraconducteurs et bien d'autres phases de la matière. Comme Serge Haroche l'a très justement souligné [192], la théorie quantique ne nous fournit pas seulement un outil pour décrire la Nature avec une grande précision, elle nous a également permis d'accéder à des technologies de calcul, d'exploration et d'action sur la matière comme sur le rayonnement inimaginables avant sa découverte. Parmi ces révolutions technologiques, les semi-conducteurs et avec eux le transistor, puis le laser et les fibres optiques ont été des éléments essentiels dans le développement des technologies de la communication et de l'informatique. Ainsi, depuis sa naissance jusqu'à la fin du vingtième siècle, la théorie quantique connut un succès écrasant sans pour autant que sa compréhension fondamentale ait beaucoup changé depuis les travaux de ses pères fondateurs.

En effet, durant cette première période, la communauté des physiciens, sans doute déroutée par le caractère contre-intuitif de la théorie quantique et de son formalisme, adopta dans sa grande majorité une attitude très pragmatique face à cette nouvelle manière de décrire la Nature, faisant sienne la philosophie de Cocteau : « Puisque ces mystères nous dépassent, feignons d'en être les organisateurs ». Pendant plusieurs décennies, les physiciens renoncèrent donc à vraiment comprendre la théorie quantique mais organisèrent son déploiement dans l'ensemble des champs disciplinaires de la physique.

C'est à la marge de cette attitude dominante, résumée par l'adage *Shut up and calculate*, que se développèrent un ensemble de travaux visant à questionner les fondements même de la théorie quantique, à manipuler des systèmes quantiques uniques (quand le dogme enseigné dans les amphithéâtres insistait sur le caractère statistique de son interprétation), ou encore à utiliser la théorie quantique de manière essentielle pour calculer ou communiquer. Ce furent les germes de ce qu'on appelle maintenant la seconde révolution quantique, qui prit son plein essor lorsque de nombreux laboratoires s'avérèrent capables de manipuler des objets quantiques individuels de manière reproductible, contrôlée et précise. Sur un plan plus conceptuel, cette seconde révolution donna lieu à

un rapprochement spectaculaire entre la physique quantique et les sciences de l'information et du calcul, rapprochement que le grand public connaît sous la forme de la quête de l'ordinateur quantique mais qui va bien au-delà.

Ainsi, peu à peu, la perspective scientifique sur la théorie quantique d'un certain nombre de chercheurs glissa donc d'un regard d'ingénieur, qui l'utilise sans trop se poser de questions en s'appuyant sur des concepts (classiques) pour disposer d'un cadre interprétatif, à un véritable questionnement sur les implications du fait que nous vivons dans un univers décrit par la théorie quantique et, de manière conjointe et encore très embryonnaire, sur les limites de ce cadre. Même si nous n'avons encore à ce jour aucune indication expérimentale que la théorie quantique n'est pas la fin de l'histoire, une nouvelle manière d'appréhender la théorie en lien avec les théories de l'information et du calcul est en train d'émerger. Ce courant de pensée vise à explorer les questions fondamentales suivantes : comment la théorie quantique peut-elle constituer une description complète de la physique à toutes les échelles, de la structure la plus fondamentale de l'espace-temps aux échelles cosmologiques ? Et, surtout, quelles informations peut-on tirer du lien entre théorie quantique, théorie de l'information et théorie du calcul pour mieux comprendre la théorie quantique elle-même et, peut-être, en cerner les limites ?

L'objectif de cet ouvrage est précisément de faire le point sur ces développements et de permettre aux étudiants et aux chercheurs d'en avoir une vue synthétique pour pouvoir suivre les progrès en cours.

Une première perspective scientifique de cet ouvrage est donc de présenter ce qui, pour nous, constitue le bagage scientifique minimal pour rentrer dans le domaine des technologies quantiques. Celles-ci visent à utiliser les ressources que constituent les états d'un système quantique et le principe de superposition pour faire des choses que l'on ne pouvait pas ou que l'on peut moins bien faire dans un monde régi par des lois classiques ou quasi classiques. Tel est le grand changement apporté par la seconde révolution quantique : non seulement la théorie quantique permet de décrire la Nature avec une précision et une universalité jamais égalée, mais ce qui en constitue le cœur permet également de calculer, de communiquer, de mesurer et enfin de simuler la Nature d'une manière radicalement nouvelle et très prometteuse. Au cours du dernier demi-siècle, ces nouvelles possibilités ont donné naissance à un foisonnement d'activités tant fondamentales qu'appliquées.

Pour mieux appréhender l'étendue de ce vaste domaine d'activité, les concepts importés d'autres disciplines comme les notions de ressources, d'information ou encore de complexité algorithmique s'avèrent extrêmement utiles. En premier lieu, ils permettent de quantifier les spécificités et les atouts de la théorie quantique en termes d'efficacité ou de rapidité dans des situations expérimentales rencontrées au laboratoire ou de manière plus générale. Mais au-delà de cette utilité, leur confrontation permet d'appréhender le domaine des technologies quantiques avec un regard qui transcende les disciplines traditionnelles et qui, comme la courte Histoire de ce domaine le montre déjà, est

un puissant moteur d'inspiration. Cette hybridation entre théorie quantique, théories de l'information et du calcul constitue le langage naturel des technologies quantiques. Un des objectifs de cet ouvrage, principalement décliné dans son premier tome, est donc de montrer comment s'enchevêtrent théorie quantique, sciences de l'information et du calcul pour donner au lecteur les clés de la seconde révolution quantique.

Mais cet ouvrage a également une perspective plus fondamentale : comment reformuler et comprendre la théorie quantique d'une manière qui nous permette de valider ou d'infirmer l'hypothèse selon laquelle elle constitue une description complète de l'Univers ? Ce questionnement va au-delà de la démarche habituelle du physicien qui confronte modèles et expériences. Il transcende également le domaine des technologies quantiques pour aborder la question de notre compréhension de la Nature. Tenter d'y répondre nécessite d'approfondir les conséquences de cette hypothèse en les poussant dans leurs derniers retranchements. Sans prétendre répondre à la question, c'est la démarche que nous adopterons dans le second tome de cet ouvrage.

Pour nous aider dans cette tâche, nous verrons que, là aussi, les concepts importés se révèlent d'une aide précieuse. La fertilisation croisée entre les différentes disciplines dont il sera question ici ne fait pas que nous apporter de simples outils quantitatifs mais nous permet de cerner au mieux ce que la théorie quantique nous apprend sur le monde.

Nous pensons en effet qu'une théorie physique doit naturellement indiquer sa propre interprétation. En pratique, celle-ci apparaît plus ou moins naturellement selon le point de vue adopté. Dans le cas de la théorie quantique, la formulation orthodoxe que l'on trouve dans les manuels ou la formulation de Feynman par l'intégrale de chemins, qui se sont toutes les deux avérées très fructueuses pour les applications, ne constituent pas une clé évidente pour en déverrouiller l'interprétation naturelle. En particulier, ce qui constitue la grande rupture philosophique apportée par la théorie quantique, à savoir l'absence d'état objectif pour un système physique réel qui est nécessairement un sous-système de l'Univers, est pour ainsi dire passée sous les radars durant des décennies à la notable exception des travaux d'Everett [151].

Comme en relativité générale, la notion d'état relatif fait ressortir le rôle de l'observateur. Avec la physique quantique, nous sommes cependant amenés à en préciser sa structure : comme Bohr et Everett le soulignèrent, il doit évidemment disposer d'instruments de mesure — tels que règles et horloges en relativité générale — mais aussi d'une capacité de traitement de données (et donc de calcul) ainsi que de communication pour communiquer avec d'autres observateurs. Cependant, si la théorie quantique est complète, tout observateur est aussi un sous-système de l'Univers et doit pouvoir être considéré comme un système quantique. Cela pose immédiatement le problème de l'émergence du classique au sein d'un univers purement quantique, un problème important dont nous discuterons de manière approfondie dans le second tome de cet

ouvrage en nous appuyant sur les concepts et méthodes introduits dans le premier tome.

Au travers de ce volume, l'importance de la notion d'observateur, le lien entre théorie physique, calcul et probabilités, ainsi que l'émergence du classique dans un univers où tout observateur et tout système d'acquisition et de traitement de données est lui aussi un système quantique nous conduiront naturellement vers une interprétation relative de la théorie quantique dans laquelle la prédictibilité et la comparaison théorie/expérience n'ont de sens que pour des sous-systèmes suffisamment complexes pour pouvoir être considérés comme des observateurs au sens où nous l'avons dégagé. Cette démarche nous amène à souligner une similitude profonde entre la relativité générale et la théorie quantique : ni l'une ni l'autre ne visent à décrire une réalité absolue mais toutes deux fournissent une description complète et cohérente des corrélations entre sous-systèmes physiques. Pour la théorie quantique, la théorie de Shannon, la théorie des probabilités, ainsi que la formalisation du calcul apparaissent comme la « couche logicielle » qui permet de tester cette cohérence pour des sous systèmes assez complexes. Dans cette optique, la théorie quantique de la communication est l'expression des relations entre ces notions de « haut niveau » et la couche « bas niveau » constituée par leurs réalisations physiques sous forme de systèmes quantiques.

Bien entendu, nous ne prétendons pas que ces modestes réflexions constituent la fin de l'histoire. En effet, même si ce point de vue nous paraît naturel, il est plus proche d'une esquisse que d'un édifice formel achevé. Plusieurs questions importantes restent en suspens : outre les mystères phénoménologiques que constituent la matière et l'énergie sombres, d'un point de vue conceptuel, le concept fondamental d'espace-temps échappe encore à une modélisation microscopique parfaitement convaincante dans le cadre de la théorie quantique. Peut-être que cela n'a pas grande importance et qu'une approche de nature effective conviendra, mais le problème reste encore ouvert malgré des décennies de propositions plus ou moins heureuses ou malheureuses. Dans le même périmètre scientifique, la relation entre très grandes (cosmologiques) et très petites échelles (planckiennes) n'est pas encore comprise. Enfin, la place de la notion de calcul et de complexité (classique comme quantique) est très loin d'être claire même si ces concepts commencent à apparaître dans la physique des trous noirs.

Quelles que soient les surprises que nous réserve l'avenir, nous avons la naïveté de croire que le point de vue et l'aventure intellectuelle que nous avons essayé de partager au travers de cet ouvrage nous auront permis de comprendre un peu mieux la théorie quantique que lorsque nous l'avons apprise. Même si le chemin fut aussi long que la Pan Américaine, parfois aussi douloureux que celui de Saint-Jacques en espadrilles, nous y avons pris beaucoup de plaisir. Puisse le lecteur apprécier ce *road trip* autant que nous.

Structure de l'ouvrage

Le premier tome, que vous tenez entre vos mains, présente en détail trois disciplines scientifiques, théorie quantique, théorie de l'information et théories de la calculabilité et complexité, ainsi que les relations qui sont récemment apparues entre elles. Il décrit une bonne partie des outils couramment utilisés par les chercheurs. Nous espérons qu'il permettra, d'une part à un étudiant intéressé d'acquérir les bases de ces domaines et, d'autre part, à l'expert, de revoir ces concepts et outils sous un jour nouveau.

Le second tome, quant à lui, s'attache à montrer comment les relations entre ces trois disciplines nous permettent de revenir sur l'interprétation physique de la théorie quantique, ainsi que ce qui en découle pour notre compréhension du monde. Plus prospectif, il présente un panorama de plusieurs thèmes de recherche actuels autour de ces questions conceptuelles.

La logique du découpage en deux tomes s'inscrit dans une perspective simple : là où le premier tome contient plusieurs histoires se déroulant en parallèle, centrées sur les applications technologiques de la théorie quantique, ce n'est véritablement que dans le second tome que ces histoires convergeront vers une synthèse qui, nous espérons, éclairera les fondements de la théorie quantique elle-même.

Pour constituer une passerelle entre les différentes communautés scientifiques auxquelles ce livre s'adresse, nous avons abandonné le format en silo des manuels de cours habituels pour développer notre propos selon cinq axes majeurs. Chacun de ces axes développe une histoire cohérente qui fournit un point de vue utile pour comprendre la théorie quantique.

L'histoire principale qui ouvre ce premier volume est celle de la théorie quantique. À partir de l'approche pragmatique basée sur les amplitudes à la Feynman (chapitre 1), ainsi que de l'étude des deux systèmes fondamentaux pour les technologies quantiques, le qubit et l'oscillateur harmonique (chapitre 4), le cœur de la théorie quantique sera révélé par l'étude de la notion d'intrication (chapitre 5). Étrangeté quantique par excellence, l'intrication sera le fil d'Ariane structurant l'ouvrage dans son ensemble, que ce soit au niveau conceptuel au travers du caractère relatif des états quantiques ou au niveau des applications comme le calcul quantique.

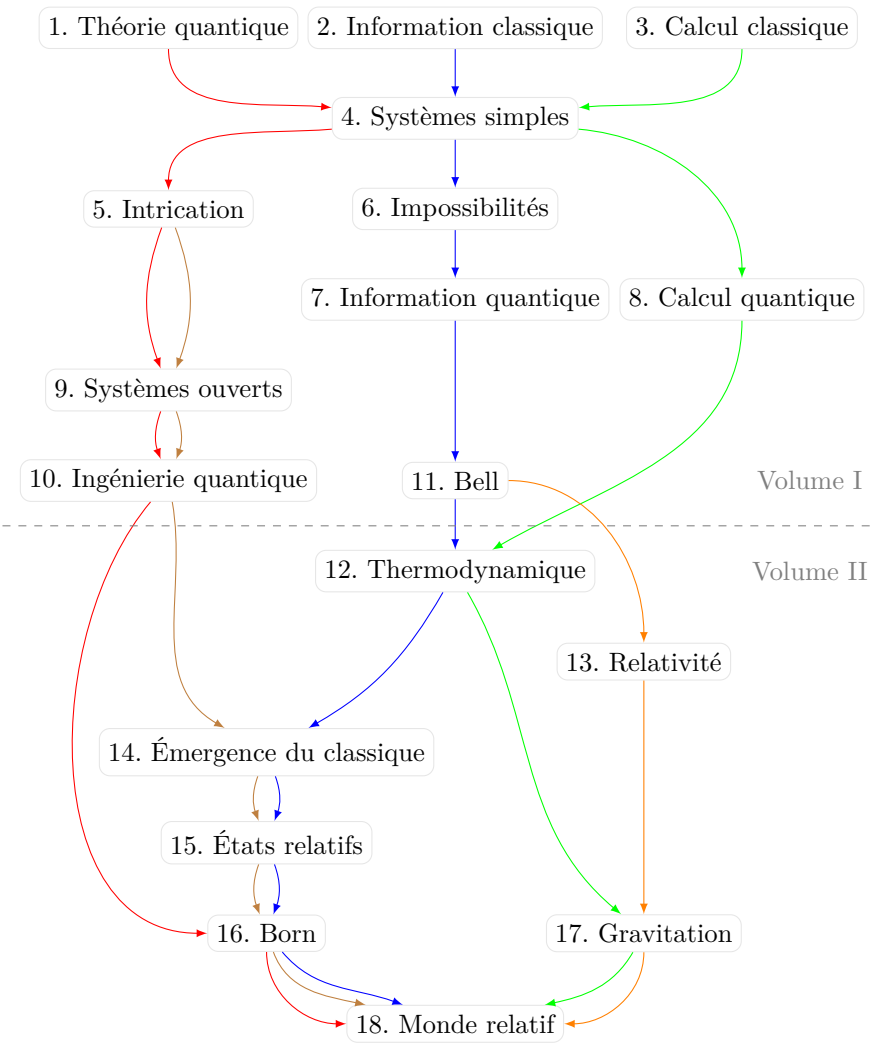
À partir de l'intrication, notre étude de la théorie quantique se poursuivra en développant deux histoires parallèles. La première est encore celle de la théorie

quantique que nous avons développée jusqu'alors. La seconde sera centrée sur l'intrication quantique comme source d'un changement de paradigme majeur. Bien que pleinement développée dans le second tome, cette histoire est bien présente dans la physique des systèmes quantiques ouverts tant au niveau général (chapitre 9) que dans leurs réalisations concrètes au sein des laboratoires (chapitre 10).

En parallèle de ces histoires viendront s'en dérouler deux autres issues des sciences de l'informatique et de l'information. Elle permettront à nouveau d'approfondir notre compréhension de la théorie quantique.

L'interface entre informatique théorique et physique est très récente et trouve son origine dans l'idée, développée entre autres par Landauer et Deutsch, que l'univers physique peut être compris en termes de processus de calcul ou de traitement d'information. Pour aborder ce point de vue et les perspectives qu'il ouvre, les théories de la calculabilité et de la complexité seront exposées dès le début de l'ouvrage (chapitre 3). Nous verrons dans le chapitre 8 en quoi disposer de ressources quantiques modifie la donne (ou non) sur ce qui est calculable (ou non) et avec quelle efficacité. Cette dernière question est évidemment ce qui motive l'effort entrepris pour fabriquer ce qu'on appelle un ordinateur quantique.

De son côté, la théorie de l'information ne s'est pas développée indépendamment de la physique. Les travaux de Shannon, que nous présenterons au chapitre 2, ont permis de quantifier nos capacités à transmettre, récupérer et modifier de l'information et ont popularisé la vision selon laquelle les systèmes physiques peuvent être vus comme des sources d'information dont les propriétés peuvent être quantifiées dans un cadre théorique universel. Là encore, cela amène à se demander comment la théorie quantique vient contraindre nos capacités d'action, d'abord par des résultats généraux d'impossibilités comme le théorème de non-clonage (chapitre 6), puis en revisitant et en approfondissant les problématiques de la théorie de l'information avec cette fois des ressources quantiques (chapitre 7). Enfin, qui dit information dit corrélations et un long chemin que nous relaterons dans le chapitre 11 fut nécessaire pour mettre en évidence la différence essentielle entre corrélations classiques et corrélations quantiques. Le théorème de Bell et les expériences visant à tester les corrélations quantiques forment une avancée majeure de la physique du XX^e siècle. Confirmant de manière magistrale la théorie quantique, ils ont conduit les physiciens à s'interroger sur des notions supposées évidentes comme celles de localité, de réalité et finalement à remettre en question leur vision du monde.



Niveaux de difficulté

Les sections marquées d’un symbole ☒ Certaines sections et paragraphes commencent par le symbole ☒. Celui-ci dénote des parties du livre plus ardues ou des compléments qui pourront être passés en première lecture. Ils demeureront intéressants pour approfondir le sujet en seconde lecture et contiennent de nombreuses références.

Les chapitres La difficulté relative de chaque chapitre est indiquée par une notation sur trois points : ●○○ signifie que le chapitre est lisible sans connaissance préalable, ●●○ que le chapitre nécessite d’avoir lu un chapitre précédent de l’ouvrage, et enfin les chapitres ●●● sont plus avancés. Le tableau qui suit permet donc d’avoir une vision des interdépendances entre les différents chapitres mais, bien entendu, chaque chapitre signalé comme difficile comprend des sections accessibles à tous les lecteurs ce qui permet une lecture en plusieurs passages de l’ouvrage.

	Quantique	Information	Calcul
1. Théorie quantique	● ● ○	○ ○ ○	○ ○ ○
2. Information classique	○ ○ ○	● ○ ○	○ ○ ○
3. Calcul classique	○ ○ ○	○ ○ ○	● ○ ○
4. Systèmes simples	● ○ ○	● ○ ○	○ ○ ○
5. Intrication	● ○ ○	● ○ ○	○ ○ ○
6. Impossibilités	● ○ ○	● ○ ○	○ ○ ○
7. Information quantique	● ● ●	● ● ●	○ ○ ○
8. Calcul quantique	● ● ●	○ ○ ○	● ● ●
9. Systèmes ouverts	● ● ●	● ○ ○	○ ○ ○
10. Ingénierie quantique	● ● ○	● ○ ○	● ● ○
11. Inégalités de Bell	● ● ○	● ○ ○	○ ○ ○

Chapitre 1

Théorie quantique

Qu'est-ce que la théorie quantique ? Quel que soit le sens que l'on donne à cette question, il est toujours difficile d'y répondre, même après un siècle de développement. Tout d'abord, la théorie quantique n'est pas à mettre sur le même plan que l'électromagnétisme, le modèle standard ou la théorie des cordes. C'est une théorie au sens propre du terme, c'est-à-dire un cadre conceptuel commun à tous les modèles microscopiques visant à décrire la Nature (et que l'on appelle improprement des « théories »). Pour reprendre une comparaison due à Scott Aaronson, la théorie quantique est le système d'exploitation qui nous permet de faire tourner des programmes qui sont des tentatives de description de la Nature appelées modèles. L'électrodynamique quantique, la chromodynamique quantique, le modèle standard de la physique des particules, les modèles de supergravités, la théorie des cordes ou la gravité quantique à boucles ne sont que des modèles parfaitement intégrés dans le cadre de la théorie quantique. Tous les résultats expérimentaux à ce jour tendent à confirmer que la Nature est fondamentalement quantique.

Sachant cela, on peut donc naturellement se demander quels sont les postulats fondamentaux de la théorie et quelle vision elle nous donne sur le monde. Traditionnellement, l'enseignement de la physique quantique revisite les découvertes expérimentales du début du xx^e siècle. Cette approche parfaitement légitime a cependant contribué à une image abstraite, hautement mathématique et contre-intuitive de la physique quantique. Or, ceci ne reflète pas nécessairement correctement le statut de cette théorie qui, de nos jours, fait partie intégrante de la base technologique de notre société et de son évolution (même si cela ne saute pas aux yeux au premier abord).

Pour véritablement nous préparer à rentrer dans le vif du sujet, le mieux est d'avoir l'attitude de l'ingénieur. La physique quantique est une science qui s'est illustrée par sa capacité à expliquer qualitativement et quantitativement un vaste ensemble de phénomènes, du tableau périodique de Mendeleïev aux phases exotiques de la matière comme les superfluides et les supraconducteurs (et bien d'autres choses encore). Elle est indispensable pour comprendre pourquoi les étoiles brillent, pour comprendre et concevoir les composants de nos

ordinateurs ou encore pour avoir une description quantitative du comportement des constituants élémentaires de l'Univers.

Dans cette perspective, parce qu'elle va directement à l'essentiel de la physique quantique, à savoir les interférences quantiques, l'approche de Feynman [153, 154] est particulièrement adaptée. Dans ce qui va suivre, nous allons donc présenter cette approche qui s'avère équivalente aux formulations habituellement présentées aux étudiants mais que nous trouvons bien plus intuitive. De plus, elle ne fait pas vraiment appel à une démarche inductive où l'on part d'un cadre classique, comme la dynamique décrite en formalisme hamiltonien, pour induire la dynamique quantique par un processus de quantification *ad hoc*.

L'approche de Feynman ne prétend ni être une axiomatisation de la théorie quantique, ni constituer une réflexion sur ses fondements. Elle vise avant tout à énoncer des règles opératoires pour calculer des quantités qui pourront être comparées à des données expérimentales. Grâce à ces règles de base ou à celles habituellement utilisées, les physiciens manipulent la physique quantique de la même manière que les ingénieurs mobilisent tout un corpus scientifique pour construire des ponts, des robots, des avions sans avoir à refaire le chemin qui a conduit à cette connaissance. Le but de ce chapitre sera donc d'énoncer clairement les règles de base du formalisme, sans chercher à les déduire d'un ensemble de postulats plus fondamentaux, ni à aller plus loin que leur simple énoncé. Pour ceux et celles qui découvrent cette physique, elle peut paraître totalement bizarre et ne pas vous plaire. Peut-être ne voyez-vous pas pourquoi l'Univers obéit à ces règles plutôt qu'à d'autres. Ces questions sont légitimes mais, avant de les aborder, il faut en premier lieu accepter la Nature telle qu'elle semble être et non pas chercher à imposer ce qu'on aimerait qu'elle soit. Connaître ces règles et être familier avec la physique qu'elles décrivent constituent les premiers pas vers une bonne compréhension de notre monde.

1.1 La formulation de Feynman

1.1.1 Statut de la théorie quantique

La théorie quantique est un édifice mathématique cohérent permettant de construire des modèles de phénomène physique aux plus petites échelles expérimentales accessibles. Elle permet de prédire et d'expliquer avec la plus grande précision tous les résultats des expériences faites à ce jour. Or on constate, de manière générale et contre-intuitive, qu'une prédiction exacte peut être faite non pas sur une réalisation unique d'une expérience mais sur la statistique des résultats de mesure. Ceci n'est en rien le reflet d'une limitation de l'expérimentateur. L'aléatoire fait partie intégrante de la fabrique du monde physique.

Un tel problème est parfaitement traité par la théorie mathématique des probabilités. Intuitivement, la théorie des probabilités se confronte aux données expérimentales au travers de la notion de fréquence statistique : dans la limite

d'un grand nombre de répétitions d'une expérience, la fréquence d'observation d'un phénomène donné tend vers un nombre qui est la probabilité d'occurrence de ce phénomène. C'est l'approche fréquentiste des probabilités. La physique quantique nous fournit un cadre qui, quand on l'applique à un modèle donné, permet de prédire la probabilité de transition entre un état initial et un état final pour n'importe quel système physique censé être décrit par ce modèle. Ces probabilités de transition peuvent alors être reliées à des données expérimentales, ce qui permet la comparaison théorie/expérience.

La physique quantique pourrait paraître proche de la physique statistique, car on sait que cette dernière décrit également le monde à travers des probabilités. Néanmoins, il y a une différence fondamentale entre les deux cadres : les probabilités de Boltzmann reflètent une ignorance de l'observateur sur le micro-état du système physique considéré alors que les probabilités en théorie quantique sont véritablement intrinsèques et ne peuvent être substituées par un approfondissement de la connaissance que l'on a de l'état du système. En réalité, le fait que ces probabilités soient intrinsèques et ne découlent pas d'une ignorance sur le système a été au cœur de nombreux débats sur les fondements de la théorie (voir chapitre 11) et leur statut apparaît en fait plus subtil suivant l'interprétation de la théorie que l'on adopte (chapitre 16, volume II). Un bon point de départ sera donc de penser la théorie quantique comme une sorte d'extension de la théorie des probabilités.

1.1.2 Les règles de Feynman

Comparer une théorie à l'expérience

Le point de vue de Feynman sur la physique quantique [156] est avant tout un point de vue pragmatique qui vise à formuler des règles simples permettant de produire des prédictions que l'on puisse confronter aux données expérimentales.

Dans cette approche, les choses se passent dans un laboratoire où un observateur essaye de caractériser le comportement d'un système. Il peut en fabriquer autant de réalisations physiques qu'il veut et on supposera qu'il peut en contrôler parfaitement la préparation.

Pour caractériser l'état d'une réalisation du système physique qu'il étudie, l'observateur va, dans un premier temps, s'appuyer sur son intuition classique et mettre en place une interrogation du système physique, c'est-à-dire un protocole expérimental de mesure bien conçu. Par exemple, pour un photon unique, on peut chercher à mesurer le vecteur d'onde, ou la réponse à la question « *Est-ce que ce photon est polarisé horizontalement ou verticalement ?* ». Un filtre polarisant ou un cube séparateur de polarisations permettent de répondre sans ambiguïté à cette question, une réponse excluant l'autre. L'important est donc de disposer de propriétés mutuellement exclusives ou, de manière synonyme, parfaitement discernables, comme ici les polarisations du photon.

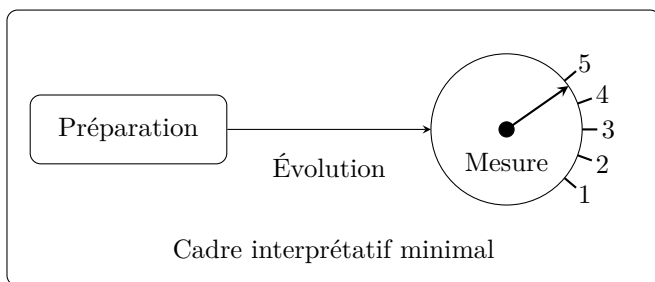


FIGURE 1.1 – Toute expérience de physique quantique peut être interprétée comme l'histoire d'un processus de préparation, d'évolution et de mesure. Toute information extraite de la théorie est relative à l'ensemble de tous ces étages.

Les quantités qui vont nous intéresser sont alors les probabilités de mesurer le système dans un état f sachant qu'on est parti d'un système préparé dans un état i . On cherche donc ce que l'on appelle la probabilité de transition entre l'état initial i et l'état final f . Notons qu'une probabilité de transition est une probabilité conditionnelle : elle nous donne la probabilité qu'une réalisation soit observée à l'issue de l'expérience dans l'état f sachant qu'elle a été initialement préparée dans i . La figure 1.1 résume ce cadre interprétatif minimal derrière toute analyse d'un modèle quantique d'un système physique. Préciser le contenu de cette affirmation est l'un des objets principaux de nos discussions.

D'un point de vue pratique, le point important est qu'on peut au moins en principe sonder si un système est dans un état donné et obtenir une réponse binaire à cette question : soit le système est détecté dans l'état considéré (réponse 1), soit il ne l'est pas (réponse 0). L'expérimentateur qui prépare un ensemble de réalisations du système physique dans l'état initial i et les détecte ensuite dans l'état final f va donc typiquement obtenir comme données une succession de 0 et de 1 qui décrivent si la transition souhaitée s'est produite ou non. La fréquence statistique des 1 doit converger, dans la limite d'un grand nombre de réalisations de l'expérience, vers la probabilité de transition $p_{i \rightarrow f}$ prédite par la physique quantique.

Règle 0 : Les amplitudes de probabilité

Le point de départ de la physique quantique est qu'à chaque transition entre un état initial i et un état final f pour un système isolé, on associe une *amplitude de probabilité de transition*¹, qui est un nombre complexe $A_{i \rightarrow f}$.

Les règles de base de la physique quantique doivent donc en premier lieu nous donner le sens physique de ces amplitudes, c'est-à-dire les relier à la notion

1. On utilisera rapidement la dénomination amplitude de probabilité en omettant le terme transition, le contexte étant suffisant pour savoir de quoi on parle.

de probabilité. Elles doivent ensuite nous dire comment calculer ces amplitudes : plus précisément, elles doivent nous dire comment calculer l'amplitude totale associée à une transition entre un état initial i et un état final f à partir d'un principe dynamique.

Insistons bien ici sur le fait que le point de vue adopté pour le moment est un point de vue complètement dynamique. Les amplitudes de transition caractérisent en effet un processus de préparation i , une évolution pendant un intervalle de temps donné (« transition ») et enfin un processus de mesure f . Nous verrons plus loin que le point de vue standard sur la théorie quantique permet et adopte en premier lieu un point de vue cinématique.

Règle 1 : La règle de Born

La première règle, appelée *règle de Born*, relie les amplitudes de transition aux probabilités de transition $p_{i \rightarrow f}$ des états i à f par

$$p_{i \rightarrow f} = |A_{i \rightarrow f}|^2. \quad (1.1)$$

Naturellement, il est possible de montrer que cette définition des probabilités à partir des amplitudes respecte bien les axiomes standard qu'une théorie des probabilités doit satisfaire [27].

Rappelons-nous que, pour accéder à cette probabilité expérimentalement, on doit effectuer un très grand nombre d'expériences dans lesquelles on prépare le système dans l'état initial et où l'on cherche à le détecter dans l'état final. Les données expérimentales permettent alors de remonter à la fréquence statistique de cette transition qui, dans la limite d'un très grand nombre de réalisations, converge vers la probabilité de transition.

Ainsi la règle de Born nous fournit une connaissance du comportement attendu du système dans l'expérience à partir de la donnée des amplitudes de probabilité. Une des grandes questions des fondements de la physique quantique consiste à déduire cette règle d'une approche plus fondamentale. En effet, cette apparition des probabilités et du hasard semble rentrer en conflit avec la règle suivante, complètement déterministe, sur le calcul de l'évolution temporelle des amplitudes.

Règle 2 : Sur tous les chemins tu sommeras

La deuxième règle de la physique quantique permet le calcul des amplitudes de transition pour un système isolé à partir des amplitudes associées à chaque trajectoire reliant un état initial noté i à un état final f ². Pour toute trajectoire γ reliant i et f , il existe une amplitude de probabilité $A[\gamma] \in \mathbb{C}$ associée à ce chemin. L'amplitude totale associée à la transition de i à f est la somme des

2. Attention, ce que l'on appelle ici une trajectoire prend place dans l'espace des configurations, et n'est pas nécessairement un déplacement spatial. Ce point sera éclairci avec les exemples donnés plus loin.

amplitudes associées à tous les chemins ou trajectoires possibles reliant i et f , appelée *intégrale de chemin*,

$$A_{i \rightarrow f} = \sum_{\gamma: i \rightarrow f} A[\gamma]. \quad (1.2)$$

Cette expression, additive par construction, incorpore le *principe de superposition*, principe fondamental de la théorie quantique que l'on exprimera mieux lors de l'exposition de la formulation hamiltonienne.

Grâce à cette équation, nous pouvons théoriquement évaluer les amplitudes de transition pour une dynamique donnée. La probabilité de transition s'écrit alors par la règle de Born :

$$p_{i \rightarrow f} = \left| \sum_{\gamma: i \rightarrow f} A[\gamma] \right|^2 \quad (1.3a)$$

$$= \sum_{\gamma: i \rightarrow f} |A[\gamma]|^2 + \sum_{\gamma_+ \neq \gamma_-: i \rightarrow f} A[\gamma_+] A^*[\gamma_-]. \quad (1.3b)$$

Le premier terme de la seconde ligne peut s'interpréter comme une somme de probabilités classiques associées à chaque chemin. En revanche, le second terme représente les *interférences* entre des chemins différents.

Ces termes d'interférences constituent le cœur de la spécificité quantique et il a fallu parcourir un long chemin pour en arriver à cette conclusion. Côté lumière, l'observation du phénomène d'interférence remonte au XIX^e siècle et aux expériences faites par Thomas Young qui montra le caractère ondulatoire de la lumière. En 1909, la même expérience fut faite mais cette fois à faible intensité et là, surprise, les effets d'interférences se produisaient même avec un seul photon [357]. Côté matière, de Broglie, en 1924, émet l'hypothèse de la dualité onde/corpuscule [74], hypothèse justifiée en 1924 par Davisson et Germer qui réalisèrent la diffraction des électrons par un cristal [121]. L'expérience d'interférence des trous d'Young avec des électrons uniques dans le vide ne fut réalisée qu'en 1961 [221], c'est-à-dire bien après la découverte du formalisme de la théorie quantique au milieu des années 1920. Les effets d'interférences quantiques ne sont pas uniquement visibles avec des particules élémentaires comme l'électron ou le photon. Récemment, des expériences de type trous d'Young, et l'observation des franges d'interférences, ont été réalisées notamment avec des atomes de néon en 1992 [334], des molécules de C₆₀ ou C₇₀ [16, 71] et enfin en 2011 avec des molécules organiques de 430 atomes [174].

À ce stade, on n'a rien dit sur la forme précise de ces amplitudes, ni même sur leurs propriétés. Pourtant, il doit bien y en avoir car si les amplitudes donnent accès aux probabilités de transition par la règle de Born, elles sont nécessairement contraintes. Quel que soit le système, il existe au moins un ensemble d'états que l'on peut considérer comme « complet et non redondant » au sens où ces états s'excluent mutuellement (les amplitudes de transition entre deux états de ce type sont nulles s'ils sont distincts) et décrivent l'ensemble

des possibilités finales lors d'une évolution physique. En conséquence, pour un système isolé, la somme des probabilités sur toutes les transitions possibles entre un état initial donné et les états d'un tel ensemble « complet et non redondant » d'états finals doit être égale à l'unité

$$\sum_f |A_{i \rightarrow f}|^2 = 1. \quad (1.4)$$

Qu'en est-il maintenant si l'on mesure par quel chemin passe le système lors de son évolution ? Ici le système n'est plus isolé mais en contact avec un appareil de mesure. La règle est alors fondamentalement différente : nous devons sommer directement les probabilités de transition $P[\gamma]$ pour chaque chemin γ empruntable. La probabilité de transition totale est alors

$$p_{i \rightarrow f} = \sum_{\gamma: i \rightarrow f} |A[\gamma]|^2. \quad (1.5)$$

À la différence de l'expression (1.3b), aucun terme d'interférence n'est présent. La détermination du chemin parcouru par un processus de mesure détruit les termes d'interférence quantiques. Nous verrons dans la suite du livre que ce versant de la règle peut être compris en modélisant proprement la dynamique des systèmes quantiques ouverts. Pour le moment, nous la prenons comme un fait expérimental simple conséquence du processus de mesure.

Règle 3 : La règle de localité

Il existe aussi une seconde contrainte qui reflète ce qu'on appelle la *localité* des interactions en physique. Considérons un état initial i et un chemin γ_1 qui amène le système de cet état initial à un état intermédiaire que l'on notera k , puis un second chemin γ_2 qui mène de k à l'état final f . La concaténation des deux chemins $\gamma_2 \# \gamma_1$ fournit un chemin qui va de l'état initial à l'état final. Si k parcourt l'ensemble des états physiques intermédiaires possibles, et que pour un état intermédiaire, on considère bien tous les chemins γ_1 et γ_2 possibles, on reconstruit ainsi bien l'ensemble des chemins de i à f . La troisième règle de Feynman nous dit comment calculer l'amplitude associée à une trajectoire par tranches :

$$A[\gamma_2 \# \gamma_1] = A[\gamma_2] \times A[\gamma_1]. \quad (1.6)$$

Par conséquent, cela nous permet de considérer formellement l'amplitude associée à une trajectoire comme un produit d'une infinité d'amplitudes associées à chaque tronçon infinitésimal.

Dans sa thèse de doctorat [154], en s'appuyant sur le formalisme quantique traditionnel, Feynman a montré qu'il existait une relation entre la phase de l'amplitude de probabilité $A[\gamma]$ et une quantité qui apparaît en mécanique classique, permettant de décrire entièrement la dynamique d'un système physique : l'*action*. La phase de l'amplitude de probabilité est, à une constante

additive près, l'action de la trajectoire $S[\gamma]$ en unités de *constante de Planck* réduite $\hbar$ (qui a bien la dimension d'une action, une énergie fois un temps). En conséquence, l'amplitude associée à une trajectoire γ est de la forme :

$$A[\gamma] = N_\gamma e^{iS[\gamma]/\hbar} \quad (1.7)$$

où $S[\gamma]$ est l'action associée à cette trajectoire et N_γ est un *facteur topologique* qui ne change pas si on déforme continûment le chemin γ . La phase $\exp(iS[\gamma]/\hbar)$ variant lorsque l'on déforme continûment le chemin, elle est appelée la phase dynamique. Pour le moment, notez que la règle (1.6) est équivalente à, d'une part, dire que $S[\gamma_2 \# \gamma_1] = S[\gamma_2] + S[\gamma_1]$, et d'autre part, que le facteur N_γ obéit également à (1.6).

✂ L'action en physique classique

Le *formalisme lagrangien* est une formulation de la physique classique permettant de retrouver les équations du mouvement par un *principe variationnel*, dit de moindre action : les trajectoires classiques sont celles qui minimisent l'action. Cette dernière est écrite comme l'intégrale au cours du temps d'une quantité appelée lagrangien. Le lagrangien est une fonction $L(q_i, \dot{q}_i, t)$, correspondant dans le cas typique de la mécanique newtonienne à la différence des énergies cinétique et potentielle $L(q, \dot{q}, t) = m\frac{\dot{q}^2}{2} - V(q)$. Dans le cas simple d'une particule se déplaçant dans l'espace, les q_i sont les coordonnées de cette particule et les $\dot{q}_i$ sont les dérivées temporelles de ces coordonnées. De manière complètement générale, les q_i vont désigner l'ensemble des coordonnées généralisées de la mécanique lagrangienne qui permettent de caractériser de manière non ambiguë une configuration du système : cela peut être des positions mais aussi des angles ou, pour des systèmes étendus, des configurations spatiales d'un champ. Les coordonnées q_i paramètrent un espace $\mathcal{C}$ appelé *espace des configurations*.

On définit alors l'action associée à une trajectoire dans l'espace des configurations de la façon suivante :

$$S[q_i(t)] = \int_{t_i}^{t_f} L(q_i(t), \dot{q}_i(t), t) dt.$$

La trajectoire physique suivie par le système dans l'espace des configurations est un extremum de l'action à conditions initiales et finales fixées. Ce principe peut être lui-même justifié à partir des règles de Feynman dans la limite du régime classique (section 1.2.3).

Regardons alors les conséquences de ce postulat. On définit deux trajectoires proches dans l'espace des configurations reliant toutes les deux la condition initiale A à la condition finale B et on étudie l'écart de l'action entre ces deux trajectoires (on prend ici une seule coordonnée pour simplifier l'écriture, la généralisation étant sans ambiguïté et les arguments des fonctions seront

parfois implicites) :

$$\delta S[q(t)] = \int_A^B L(q + \delta q, \dot{q} + \delta \dot{q}, t) dt - \int_A^B L(q, \dot{q}, t) dt.$$

On a alors

$$\delta S[q(t)] = \int_A^B \left(\frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q} \right) dt.$$

Sachant que $\delta \dot{q} = \frac{\partial \delta q}{\partial t}$, on effectue une intégration par parties sur le deuxième terme de l'intégrale :

$$\delta S[q(t)] = \left[\frac{\partial L}{\partial \dot{q}} \delta q \right]_A^B + \int_A^B \left(\frac{\partial L}{\partial q} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}} \right) \delta q dt.$$

Le premier des deux termes, qui est un terme dit de bord, est nul car nos deux trajectoires relient les mêmes points A et B , c'est-à-dire que $\delta q(A) = \delta q(B) = 0$. La variation de l'action à conditions initiales et finales fixées s'écrit donc :

$$\delta S[q(t)] = \int_A^B \left(\frac{\partial L}{\partial q} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}} \right) \delta q dt.$$

La trajectoire effectivement suivie extrémalisant l'action, elle vérifie donc $\delta S = 0$ au premier ordre en δq . L'intégrande doit donc être nul. Nous obtenons alors les équations du mouvement, sous la forme des *équations d'Euler-Lagrange*

$$\frac{\delta S[q(t)]}{\delta q(t)} = \frac{\partial L}{\partial q} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}} = 0.$$

La notation $\delta/\delta q(t)$ signifie que l'on fait une dérivée par rapport à la fonction $q(t)$, opération plus générale que la dérivation standard. On parle dans cette situation de dérivée fonctionnelle.

Dans le cas d'une particule dans un potentiel, on trouve

$$0 = \frac{\partial L}{\partial q} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}} = -\frac{\partial V(q)}{\partial q} - m\ddot{q}. \quad (1.8)$$

On obtient bien l'équation du mouvement newtonienne, à savoir $m\ddot{q} = \mathbf{F}_{\text{ext}}$, les forces dérivant ici d'une énergie potentielle $V(q)$. Ceci montre bien que l'action contient toute l'information nécessaire pour connaître la trajectoire réelle suivie par le système.

Ce qui est absolument remarquable, c'est ce que cette méthode est très générale. Elle permet de trouver les équations du mouvement d'un système mécanique quelconque ou d'une théorie des champs comme l'électromagnétisme, de manière automatique. C'est d'ailleurs pour cela qu'elle fut inventée à la fin du XVIII^e siècle par Joseph Louis Lagrange.

1.2 Applications et conséquences physiques

L'objectif de cette section est de montrer au travers de quelques discussions brèves la richesse qu'apporte l'approche de Feynman. Le lecteur sera peut-être frustré que l'on ne fasse qu'effleurer des sujets passionnants, mais le message à retenir ici est que des règles aussi simples que celles que nous venons d'énoncer permettent l'émergence de vastes pans de physique et fournissent un outil extrêmement puissant pour les appréhender de manière intuitive.

Pour commencer, nous allons étudier un système n'ayant que deux niveaux, qui est d'une importance fondamentale en information quantique. Ce système nous intéressera tout au long de ce livre et sera discuté en détail au chapitre 4. Nous allons ensuite étudier des systèmes qui ont une « limite classique » familière. L'exemple fondamental est celui d'une particule se déplaçant dans l'espace. Cependant, nous verrons que même dans un tel exemple, la physique quantique autorise des effets inédits qui sont liés à la topologie de l'espace des trajectoires possibles. On s'intéressera ainsi à l'effet Aharonov-Bohm, ainsi qu'au comportement de particules indiscernables.

1.2.1 Le système à deux niveaux

Le système le plus élémentaire que l'on puisse considérer en théorie quantique est le *système à deux niveaux*, souvent appelé *qubit* : c'est en effet le système fondamental de l'information quantique.

Ce système est extrêmement simple : d'un point de vue de théoricien, il s'agit de tout système qui ne comporte que deux « configurations ». Cela peut correspondre dans une expérience à deux niveaux d'énergie d'un atome par exemple ou à deux polarisations. Dans ce cas, on va noter une des configurations possibles $|0\rangle$ et l'autre $|1\rangle$.

Puisqu'il n'y a que deux états, le système n'a que deux choix possibles à chaque instant : soit il reste dans l'état dans lequel il se trouve, soit il change d'état. Par la règle zéro, chaque transition est modélisée par une amplitude de probabilité. On va supposer que l'amplitude associée à chaque saut d'un état à un autre s'écrit $-iA$ où $A \in \mathbb{R}$. La symétrie entre les sauts $|0\rangle \rightarrow |1\rangle$ (excitation) et $|1\rangle \rightarrow |0\rangle$ (désexcitation) tout comme la forme de cette amplitude se justifient facilement par la normalisation des probabilités.

Regardons maintenant ce que vaut l'amplitude de probabilité $A_{00}(t)$ ³ pour que le système soit dans l'état $|0\rangle$ à l'instant t sachant qu'il était dans l'état $|0\rangle$ à l'instant initial. Pour l'obtenir, on utilise la seconde règle sur la sommation sur tous les chemins possibles connectant ces deux états. Pendant la durée t , rien n'empêche le système de faire des sauts d'un état à l'autre, la seule chose qui est imposée est que les états final et initial soient $|0\rangle$. Soit alors $0 \leq t_1 < \dots < t_{2N} \leq t$ les coordonnées temporelles des sauts permettant de faire N aller-retours entre $|0\rangle$ et $|1\rangle$. Un schéma représentant deux trajectoires

3. Avec les notations que l'on a utilisées jusqu'ici, on pourrait l'appeler $A_{(0,0) \rightarrow (0,t)}$.

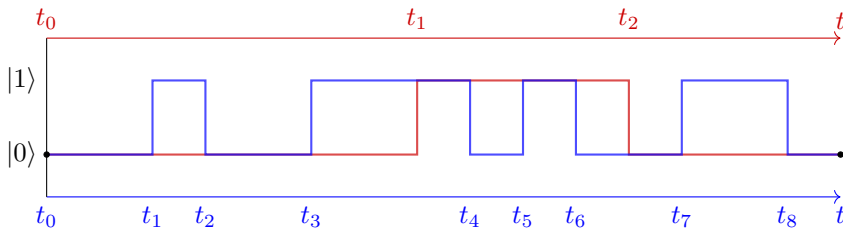


FIGURE 1.2 – Deux exemples de chemins reliant l'état $|0\rangle$ à l'instant t_0 à l'état $|0\rangle$ à l'instant t , dans le cas du système à deux niveaux. On voit qu'un tel chemin correspond à la donnée d'un certain nombre de sauts d'un état à l'autre, qui ont lieu à des instants quelconques. À chacun de ces sauts est associée une amplitude $-iA$. Pour calculer l'amplitude totale $A_{00}(t)$, il va falloir sommer sur toutes les trajectoires.

possibles, avec des nombres de saut et des instants de saut aléatoires, est donné par la figure 1.2. L'application de la troisième règle sur chaque trajectoire permet d'écrire l'amplitude $A_{00}(t)$ comme

$$A_{00}(t) = 1 + \sum_{n=1}^{\infty} \int_{0 \leq t_1 \leq \dots \leq t_{2n} \leq t} (-iA)^{2n} dt_1 \cdots dt_{2n}. \quad (1.9)$$

Le premier terme du développement correspond à l'absence de saut tandis les autres termes de la somme représentent n sauts ayant lieu à une suite d'instants aléatoires. En utilisant $\int_{0 \leq t_1 \leq \dots \leq t_{2n} \leq t} dt_1 \cdots dt_{2n} = t^{2n}/(2n)!$, on obtient donc :

$$A_{00}(t) = \sum_{n=0}^{\infty} \frac{(-iAt)^{2n}}{(2n)!} = \cos(At). \quad (1.10)$$

De la même manière, l'amplitude de transition $A_{01}(t)$ entre l'état initial $|0\rangle$ et l'état $|1\rangle$ à l'instant t s'écrit

$$A_{01}(t) = -i \sin(At). \quad (1.11)$$

Les expressions de ces deux amplitudes sont en fait valables pour une amplitude de saut complexe quelconque. Il est alors direct de vérifier que la normalisation des probabilités de transition impose à cette amplitude d'être imaginaire pure comme indiqué dès le début du raisonnement.

Les amplitudes de transition sur un temps t présentent des oscillations qui sont la conséquence des interférences entre toutes les trajectoires possibles pour le système à deux niveaux. Ces oscillations sont appelées *oscillations de Rabi* du nom de Isaac Rabi, physicien américain connu pour avoir inventé la résonance magnétique nucléaire dans les années 1930. Il s'agit clairement d'interférences mais qui, à la différence des franges d'Young, se révèlent dans une évolution temporelle. La physique de ce problème sera détaillée dans les prochains chapitres.

1.2.2 La particule libre

Le deuxième exemple que nous pouvons traiter est un système avec lequel nous devrions être familiers, à savoir celui d'une particule libre, ponctuelle et non relativiste. Dans ce cas, la somme sur tous les chemins est en fait une sommation sur un espace de fonctions appelée intégrale de chemin. Or cette notion est mathématiquement délicate et pose de sérieux problèmes d'analyse fonctionnelle étant donné qu'il n'est pas évident de définir une mesure (au sens de la théorie de la mesure) sur des espaces de fonctions. Mais ce n'est pas ce qui nous intéresse ici. Il s'agit plutôt d'explorer concrètement un exemple et d'inciter le lecteur à lire ou à relire l'article original de Feynman sur le sujet [153] ainsi que sa thèse de doctorat [154].

Considérons le cas d'une particule confinée sur un axe pour simplifier la discussion (cela ne restreint pas la généralité du problème). Un chemin γ reliant les points x_i et x_f est alors la donnée d'une fonction continue $\gamma : t \mapsto x(t)$ telle que $\gamma(t_i) = x_i$ et $\gamma(t_f) = x_f$. Dans sa thèse de doctorat, Feynman présente une construction intuitive de la somme sur tous les chemins qu'il est intéressant de suivre.

Pour construire l'intégrale de chemin, Feynman procède par analogie avec la construction de l'intégrale de Riemann. On se donne ainsi une subdivision $(t_k)_{k \in \{0, \dots, N\}}$ de l'intervalle de temps $[t_i, t_f]$, que l'on supposera régulière, de pas $\epsilon = \frac{t_f - t_i}{N}$. Un chemin γ va donc être la donnée, pour chaque instant t_k , $k \in \{1, \dots, N-1\}$, d'un point de l'espace x_k , sachant que $t_0 = t_i$, $t_N = t_f$, $x_0 = x_i$ et $x_N = x_f$. On construit finalement un chemin reliant deux points consécutifs par un segment, ou encore par le chemin classique reliant ces deux points. La sommation sur tous les chemins revient donc à prendre la limite quand ϵ tend vers 0 et à intégrer sur tous les paramètres x_k , $k \in \{1, \dots, N-1\}$. L'amplitude de probabilité $A_{i \rightarrow f}$ est l'intégrale de chemin définie comme limite :

$$A_{i \rightarrow f} = \int e^{iS[x(t)]/\hbar} \mathcal{D}x(t) = \lim_{\epsilon \rightarrow 0} \left(\int \mathcal{N} e^{i/\hbar \sum_{k=0}^{N-1} S_c(x_k, x_{k+1})} dx_1 \cdots dx_{N-1} \right) \quad (1.12)$$

où $\mathcal{N}$ est un facteur de normalisation (infini) s'écrivant $\mathcal{N} = \frac{1}{A^N}$, A ayant la dimension d'une longueur et où $S_c(x_k, x_{k+1})$ est la fonction de Hamilton (action évaluée sur la trajectoire classique physique) évaluée entre les points x_k et x_{k+1} . Pour une particule de masse m libre, ponctuelle et non relativiste, le lagrangien s'écrit :

$$L(x, \dot{x}) = \frac{m\dot{x}^2}{2}. \quad (1.13)$$

L'action classique vaut alors $S_c = \frac{m(x_f - x_i)^2}{2(t_f - t_i)}$, puisque le chemin classique d'une particule libre est la ligne droite parcourue à vitesse constante $\dot{x} = \frac{x_f - x_i}{t_f - t_i}$. On

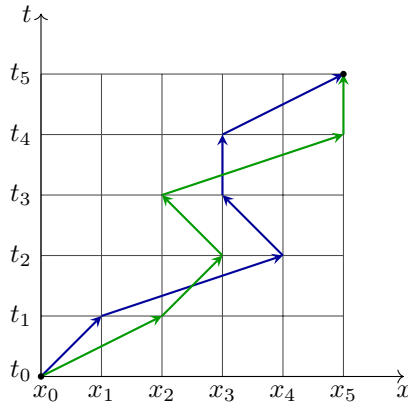


FIGURE 1.3 – Deux exemples de chemins. Afin de calculer la probabilité de transition entre l'état initial (x_0, t_0) et l'état final (x_5, t_5) , il faut sommer sur tous les chemins possibles.

a donc par définition de l'intégrale de chemin :

$$A_{i \rightarrow f} = \lim_{\epsilon \rightarrow 0} \int \frac{1}{A^N} \exp \left(\frac{im}{2\hbar\epsilon} \sum_{k=0}^{N-1} (x_{k+1} - x_k)^2 \right) dx_1 \cdots dx_{N-1}. \quad (1.14)$$

On a un ensemble d'intégrales gaussiennes que l'on peut calculer et qui nous donnent le résultat suivant :

$$A_{i \rightarrow f} = \lim_{\epsilon \rightarrow 0} \left(\frac{1}{A} \sqrt{\frac{2i\pi\hbar\epsilon}{m}} \right)^N \sqrt{\frac{m}{2i\pi\hbar(t_f - t_i)}} \exp \left(\frac{im(x_f - x_i)^2}{2\hbar(t_f - t_i)} \right). \quad (1.15)$$

Reste à déterminer la constante numérique A . Le choix naturel est tout simplement $A = \sqrt{\frac{2i\pi\hbar\epsilon}{m}}$ qui assure que la limite existe. On obtient au final l'amplitude de transition :

$$A_{i \rightarrow f} = A(x_f, t_f | x_i, t_i) = \sqrt{\frac{m}{2i\pi\hbar(t_f - t_i)}} \exp \left(\frac{im(x_f - x_i)^2}{2\hbar(t_f - t_i)} \right). \quad (1.16)$$

La première remarque que l'on peut faire est que cette expression est une limite un peu singulière puisque $|A(x_f, t_f | x_i, t_i)|^2$ ne dépend pas de x_f . Ceci signifie que la particule s'est complètement délocalisée pour tout instant $t_f > t_i$. Cette délocalisation parfaite semble assez étonnante intuitivement puisque la probabilité de présence de la particule est identique quel que soit l'endroit de l'Univers. Nous verrons un peu plus loin après avoir introduit le formalisme habituel que ceci est l'avatar d'un résultat majeur de physique quantique connu comme le principe d'indétermination de Heisenberg.

1.2.3 La limite classique

La formulation de Feynman permet de commencer à voir le lien entre la physique classique et la physique quantique. L'émergence d'un monde classique à partir d'un monde quantique est une chose très subtile qui demande de comprendre entre autres le phénomène de décohérence. Ce qui suit n'est donc qu'un premier pas vers une étude de la transition quantique/classique.

La *limite classique* revient à considérer un ensemble de trajectoires tel que les variations de $S[\gamma]$ soient bien plus grandes que $\hbar$. Considérons par exemple les trajectoires circulaires classiques autour du noyau d'un atome d'hydrogène : l'action associée au mouvement circulaire durant une période peut être évaluée en fonction du rayon de la trajectoire. Un calcul simple donne

$$S_r = \frac{h}{2} \sqrt{\frac{\alpha_{\text{qed}} m c r}{\hbar}} = \frac{h}{2} \sqrt{\frac{r}{a_B}}. \quad (1.17)$$

où $\alpha_{\text{qed}} = e^2/4\pi\epsilon_0\hbar c$ est la constante de structure fine. On voit donc apparaître le rayon de Bohr $a_B = \alpha_{\text{qed}} l_c$ où $l_c = \hbar/mc$ est la longueur de Compton comme échelle de longueur qui délimite la frontière entre un domaine quantique (trajectoires d'action caractéristique $\sim \hbar$) et un domaine semi-classique (trajectoires d'action caractéristique grande devant $\hbar$). De manière générale, les situations macroscopiques que nous considérons habituellement mettent en jeu des variations d'actions considérables devant $\hbar$.

Une bonne limite classique doit nous permettre de retrouver les équations classiques du mouvement. Grâce à la formulation de Feynman, il est possible de déduire et justifier le principe de moindre action de la physique classique. En effet, en considérant cette limite, les chemins dont l'action va contribuer à l'amplitude sont ceux pour lesquels les contributions ne se détruisent pas par interférences, c'est-à-dire les chemins pour lesquels la phase de l'amplitude de probabilité varie le moins possible. Comme la phase est proportionnelle à $S[\gamma]/\hbar$, les chemins qui contribuent à l'amplitude totale sont ceux qui sont très proches de ceux rendant stationnaire l'action, en d'autres termes les chemins classiques. C'est ce qui est représenté sur la figure 1.4. Nous venons ici de qualitativement justifier le principe de moindre action de la physique classique.

Dans cette approximation, on peut montrer que la contribution de ces chemins voisins d'un chemin classique γ_c est de la forme

$$\sum_{\gamma \sim \gamma_c} A[\gamma] \propto e^{iS[\gamma_c]/\hbar}. \quad (1.18)$$

Le facteur de proportionnalité est en général une fonction des conditions aux limites de la trajectoire classique que l'on calcule en faisant un développement autour du chemin classique, en général jusqu'à l'ordre deux. On constate que la phase n'est autre que l'action évaluée sur la trajectoire classique solution des équations du mouvement. Cette fonction $S[\gamma_c]$, qui est bien sûr différente de la fonctionnelle action, est appelée en général *fonction de Hamilton* et est fonction des données initiales, ici $S[\gamma_c] \equiv S(t_f, q_f, t_i, q_i)$.

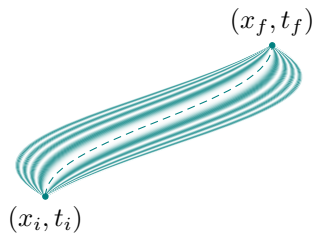


FIGURE 1.4 – Représentation des chemins contribuant à l’amplitude de probabilité associée à différents chemins. La trajectoire classique, qui extrémalise l’action, est représentée en tirets. Les trajectoires proches du chemin classique ont une phase similaire et interfèrent constructivement dans l’intégrale de chemin. Au contraire, les trajectoires éloignées interfèrent destructivement. La contribution dominante se résume alors à l’amplitude du chemin classique.

Nous n’avons fait là qu’effleurer une technique particulière de calcul des intégrales de chemin. L’imagination des physiciens donna naissance à un éventail considérable de méthodes pour calculer des intégrales de chemin, ce qui a joué un rôle crucial dans le développement de la théorie quantique des champs avec des applications en physique des hautes énergies aussi bien qu’en physique statistique. D’excellentes introductions aux méthodes d’intégrales de chemin [390] ainsi qu’à la théorie quantique des champs [251] sont accessibles et constituent le prolongement naturel des concepts introduits ici.

1.2.4 L’effet Aharonov-Bohm

L’expérience que nous allons décrire maintenant illustre le rôle subtil des phases en théorie quantique et comment elles viennent questionner le sens communément attribué à certains objets d’une théorie comme ici les potentiels électromagnétiques.

Lorsqu’une particule chargée se déplace dans un champ électromagnétique donné, celui-ci contribue par une phase à l’amplitude $A[\gamma] = e^{iS_{EM}/\hbar}$ où S_{EM} est l’interaction charge-champ. De façon remarquable, cette phase ne s’exprime pas en fonction des champs électrique et magnétique mais en fonction des potentiels $(\mathbf{A}, V)$ dont ils dérivent *via* les formules $\mathbf{E} = -\partial_t \mathbf{A} - \mathbf{grad} V$ et $\mathbf{B} = \mathbf{rot} \mathbf{A}$. La phase de $A[\gamma]$ est proportionnelle à l’intégrale spatiale de $\mathbf{j} \cdot \mathbf{A} - \rho V$ qui décrit le terme de couplage entre le champ électromagnétique et la matière décrite par sa densité de courant $\mathbf{j}$ et sa charge ρ . Pour une particule ponctuelle, nous avons $\rho(\mathbf{x}, t) = q\delta(\mathbf{x} - \mathbf{r}(t))$, où $\mathbf{r}(t)$ désigne la position de la particule au temps t , et $\mathbf{j}(\mathbf{x}, t) = q\dot{\mathbf{r}}(t)\delta(\mathbf{x} - \mathbf{r}(t))$. Le lagrangien s’écrit alors

$$L(\mathbf{r}, \dot{\mathbf{r}}) = q\dot{\mathbf{r}}(t) \cdot \mathbf{A}(\mathbf{r}(t), t) - qV(\mathbf{r}(t), t). \quad (1.19)$$

Avant de discuter plus précisément les effets de cette phase électromagnétique, analysons l'effet d'un changement des potentiels qui n'altère pas les champs électriques et magnétiques. Un tel changement, appelé une *transformation de jauge*, est paramétré par une fonction de jauge $\chi(\mathbf{x}, t)$ et est défini par :

$$\mathbf{A}'(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}, t) + \mathbf{grad} \chi(\mathbf{x}, t) \quad (1.20a)$$

$$V'(\mathbf{x}, t) = V(\mathbf{x}, t) - (\partial_t \chi)(\mathbf{x}, t) \quad (1.20b)$$

de sorte que la contribution à la phase est l'exponentielle d'un terme de dérivée totale que l'on peut intégrer et qui ne dépend alors plus que des conditions initiales $(\mathbf{x}_i, t_i)$ et finales $(\mathbf{x}_f, t_f)$ dans la somme sur tous les chemins (1.2) :

$$A'[\gamma] = A[\gamma] \exp \left(\frac{iq}{\hbar} (\chi(\mathbf{x}_f, t_f) - \chi(\mathbf{x}_i, t_i)) \right). \quad (1.21)$$

Cette phase additionnelle se factorise dans la somme sur tous les chemins (1.2) de sorte que les prédictions physiques comme les probabilités de transition (1.1) sont identiques, que l'on emploie les potentiels $(\mathbf{A}, V)$ ou les potentiels transformés de jauge $(\mathbf{A}', V')$.

Imaginons maintenant une expérience dans laquelle on utilise un faisceau de particules chargées qui se propagent en présence d'un solénoïde dans lequel circule un courant électrique comme sur la figure 1.5. Un potentiel V repousse les particules hors du solénoïde de sorte que, de manière effective, les particules se propagent librement en dehors du solénoïde sans pouvoir y pénétrer. Ce dernier génère un champ non nul en son intérieur et est donc traversé par un flux magnétique Φ_B . Cela se traduit par un potentiel vecteur $\mathbf{A}$ non nul hors du solénoïde, précisément dans la région où la particule est présente : ce champ orthoradial décroît en $1/r$ où r est la distance à l'axe du solénoïde. Nous disposons donc de tous les éléments pour comprendre l'effet du solénoïde sur les interférences entre trajectoires de la particule.

Considérons γ_+ et γ_- deux trajectoires partant toutes les deux d'un point S à l'instant initial t_i et arrivant au point D à l'instant final t_f (voir figure 1.5). La contribution électromagnétique à la phase de chacune d'entre elles dépend *a priori* du choix de jauge et des potentiels mais, comme nous le montre l'équation (1.21), les contributions de la jauge et du potentiel scalaire disparaissent pour le produit $A[\gamma_+] A[\gamma_-]^*$. En revanche, à travers la contribution du potentiel vecteur, ce produit dépend du flux magnétique. En effet, l'amplitude $A[\gamma_+] A[\gamma_-]^*$ est égale à celle d'une trajectoire obtenue en parcourant d'abord γ_+ dans le sens direct puis γ_- dans le sens inverse $\gamma_-^{(r)} \# \gamma_+$. Or cette trajectoire s'enroule un certain nombre de fois autour du solénoïde. En notant n ce nombre de tours, la contribution électromagnétique à $A[\gamma_+] A[\gamma_-]^*$ est donc donnée par $\exp(iqn\Phi_B/\hbar)$ que l'on réécrit $\exp(2\pi in\Phi_B/\Phi_0)$ où $\Phi_0 = h/q$ est appelé le quantum de flux pour la charge q . L'ensemble des couples de trajectoires (γ_+, γ_-) contenant les conditions initiales et finales se décompose donc en une

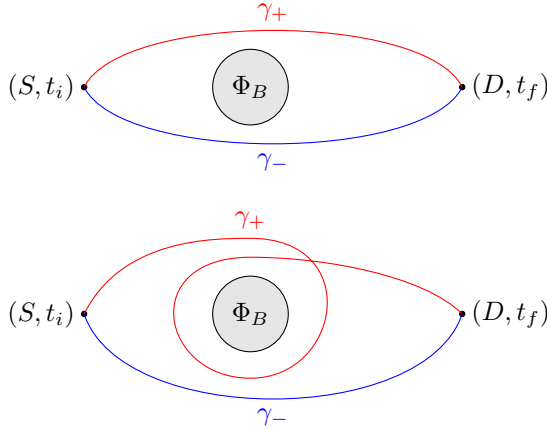


FIGURE 1.5 – L’effet Aharonov-Bohm est l’effet d’interférences quantiques modulées par un flux magnétique (partie grisée) qui apparaît lors de la propagation d’une particule chargée dans une région comme l’extérieur d’un solénoïde (vue en coupe) où le champ magnétique est pourtant nul. La paire de trajectoires du haut encerre un flux magnétique Φ_B et celle du bas, un flux $2\Phi_B$.

réunion de secteurs regroupant tous les (γ_+, γ_-) tels que $\gamma_-^{(r)} \# \gamma_+$ s’enroule n fois autour du solénoïde. La probabilité de transition de (S, t_i) à (D, t_f) est donc

$$p_{i \rightarrow f} = \sum_{n \in \mathbb{Z}} e^{2\pi i n \Phi_B / \Phi_0} \sum_{N[\gamma_-^{(r)} \# \gamma_+] = n} A_0[\gamma_+] A_0[\gamma_-]^*, \quad (1.22)$$

où $N[\gamma]$ est le nombre de tours qu’effectue un chemin fermé autour du solénoïde. On voit donc apparaître des oscillations de période Φ_0 dans l’amplitude de transition en fonction du flux magnétique Φ_B . Les sommes sur les couples de trajectoires correspondant à des nombres d’enroulements donnés donnent la série de Fourier correspondant à la dépendance en Φ_B de la probabilité $p_{i \rightarrow f}$. Ces oscillations en fonction du flux magnétique constituent l’*effet Aharonov-Bohm*.

Plusieurs remarques peuvent être faites. Un premier point important pour notre discussion est que cet effet, dans la géométrie discutée ici, provient de phases topologiques qui reflètent l’enroulement d’une paire de trajectoires dans l’expression de la probabilité de transition. Il est en effet impossible de déformer continûment deux trajectoires fermées l’une dans l’autre s’enroulant un nombre différent de fois autour du solénoïde. Un deuxième point, finalement similaire au précédent, est que cet effet est non local. En effet, quel que soit la trajectoire qu’emprunte la particule, elle ne subit aucune influence du champ physique $\mathbf{B}$ qui est confiné dans la région inaccessible interne au

solénoïde. Pourtant, la figure d'interférence est modulée par ce flux magnétique. Cet effet d'interférence quantique montre la pertinence d'utiliser les potentiels électromagnétiques et non les champs pour décrire le couplage matière-lumière.

Bien que des expériences, mentionnées par Feynman dans ses cours, aient été tentées dans les années 1960 [88], les expériences confirmant pour la première fois de manière indiscutable l'effet Aharonov-Bohm dans le contexte discuté ici remontent aux années 1980 [358, 292]. La principale difficulté consistait à exclure toute influence du champ magnétique créé par le solénoïde employé dans l'expérience qui, étant de longueur finie, ne satisfaisait pas $\mathbf{B} = \mathbf{0}$ dans la région de déplacement de la particule. Cette difficulté fut résolue par l'emploi d'un blindage supraconducteur.

✂ Pour aller plus loin

À la même période, une autre approche a permis de mettre en évidence l'effet Aharonov-Bohm, non pas en utilisant des particules chargées se propageant dans le vide mais des électrons au sein d'anneaux métalliques obtenus par micro-lithographie et suffisamment petits et froids pour espérer voir l'effet Aharonov-Bohm lié à la cohérence quantique des électrons sur l'ensemble de l'anneau. Cet exemple est intéressant car il montre comment les effets liés à la présence du potentiel électrique et du champ magnétique peuvent se combiner pour donner des résultats inattendus.

En 1985, aux États-Unis, une expérience portant sur un anneau unique fut réalisée dans les laboratoires de recherche d'IBM [372] et montra une figure d'interférences pour la conductance de cet anneau (connecté à deux électrodes) avec la périodicité $\Phi_0 = h/e$. Elle suivait de quelques années la première mise en évidence d'oscillations cohérentes dans une structure métallique par deux physiciens soviétiques, Sharvin et Sharvin, au début des années 1980 [331]. De manière surprenante, cette expérience qui utilisait de petits cylindres au lieu d'anneaux plans, montra des figures d'interférences avec une période égale à $\Phi_0/2$ et non pas Φ_0 .

La raison de ce changement de période est liée au fait que l'anneau étant remplacé par un cylindre, tout se passe comme si on explorait un grand nombre de petits anneaux empilés. Or les matériaux utilisés ne sont pas des cristaux idéaux : ils comportent un grand nombre de défauts et d'impuretés statiques. Les électrons dans le métal voient donc un potentiel compliqué. Même dans le cas de l'anneau unique d'IBM, changer le champ magnétique modifie les phases accumulées et engendre, par-dessus la périodicité en Φ_0 une variation sur des échelles de flux d'environ une dizaine de Φ_0 qui est reproductible pour un anneau donné mais qui dépend de l'anneau. On appelle cela l'empreinte magnétique. Dans le cas du cylindre, la probabilité pour un électron de traverser l'anneau, dont on peut montrer qu'elle est reliée à la conductance, est une moyenne des probabilités associées à chaque anneau élémentaire. Or la contribution Aharonov-Bohm à la période Φ_0 provient de paires de trajectoires qui, mises bout à bout comme discuté précédemment, entourent

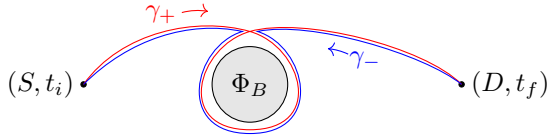


FIGURE 1.6 – Paire de trajectoires qui contribuent à la période $\Phi_0/2$ mais qui subsistent dans la moyenne sur le désordre. On voit que les trajectoires γ_+ et γ_- sont pratiquement les renversées dans le temps l’une de l’autre sauf au voisinage du point de croisement.

l’anneau exactement une fois. Cependant, dans le produit $A[\gamma_+]A[\gamma_-]^*$, le potentiel électrique vu par les électrons dans l’anneau intervient. Lorsque l’on considère un ensemble d’anneaux, cette contribution due au potentiel électrique donne une phase qui est aléatoire et se moyennera à zéro. Cela explique la disparition des termes d’interférences associées à la paire (γ_+, γ_-) de nombre d’enroulement ± 1 ou même ± 2 comme celles représentées sur la figure 1.5. Mais y en a-t-il qui survivent à cet effet de moyenne ? Dès l’enroulement ± 2 , nous pouvons considérer des paires de trajectoires telles que celles représentées sur la figure 1.6, c’est-à-dire qui, presque partout, sont le renversé temporel l’une de l’autre. Le point important est que comme ces trajectoires passent par les mêmes endroits, la phase due au potentiel électrique se compense entre γ_+ et γ_- et l’effet de moyenne pour cette paire de trajectoires et toutes celles qui sont similaires disparaît. Ces paires de trajectoires qui enserrant donc un flux magnétique pair en unité de Φ_B subsistent donc. Cela explique que l’on observe bien les oscillations de période $\Phi_0/2$ mais pas celles de période Φ_0 . Cet effet est une conséquence de la diffusion des ondes de probabilité électroniques sur le potentiel désordonné au sein du conducteur. Cette théorie de la diffusion en milieu désordonné, qui joue un rôle fondamental dans l’étude du transport électronique quantique, est décrite de manière pédagogique et complète dans le livre d’Akkermans et Montambaux [273].

1.2.5 Particules indiscernables

L’analyse de la forme des amplitudes de probabilité a fait ressortir la présence d’un facteur topologique ne dépendant que de la classe d’équivalence des chemins sous déformation continue. Il s’avère que les propriétés de ces phases permettent d’éclairer les conséquences d’une symétrie fondamentale, celle de l’*indiscernabilité* de particules identiques. Comme elles sont indiscernables, la donnée d’un état pour N de ces particules est la donnée de leur N positions à une permutation entre les particules près. En effet, si on pouvait distinguer entre particule 1 en $\mathbf{r}_1$ et particule 2 en $\mathbf{r}_2$, et particule 1 en $\mathbf{r}_2$ et particule 2 en $\mathbf{r}_1$, les particules seraient discernables.

Imaginons que l'on regarde deux états i et f pour les N particules indiscernables. Si on considère des particules sans interactions, l'action pour les N particules est tout simplement la somme des actions pour chacune des particules. La phase en $e^{iS[\gamma]/\hbar}$ se décompose donc en un produit des phases associées à chaque particule. Mais on a vu qu'il existe également un facteur de phase topologique, qui dépend *a priori* de manière non triviale du chemin total à N particules.

Considérons une configuration caractérisée par un ensemble de N positions $P = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$. Quel que soit le chemin γ que l'on considère, on peut le déformer continûment en un chemin qui part de P et retourne à P , à une permutation entre les particules près. Un tel chemin reliant N positions à elles-mêmes modulo une permutation est appelé une *tresse*. Le facteur de phase topologique que l'on considère ne peut donc dépendre que de la topologie de cet espace des tresses.

De plus, il est facile de voir que les permutations se composent lorsque l'on concatène les chemins : $\sigma_{\gamma_1 \# \gamma_2} = \sigma_{\gamma_2} \circ \sigma_{\gamma_1}$. Une manière simple d'obtenir un facteur de phase topologique consiste alors à utiliser les représentations du groupe des permutations à valeurs dans le groupe des nombres complexes de module unité que l'on note $U(1)$.

Puisque le groupe des permutations est engendré par les transpositions, qui correspondent à des chemins échangeant deux particules, une représentation du type recherché est entièrement définie par la valeur associée à une transposition. Le seul paramètre dont on a besoin pour connaître notre représentation est donc le nombre de $U(1)$ qui est associé à une transposition. Comme le carré d'une transposition est l'identité, cela veut dire que le facteur de phase associé à une transposition est de carré un⁴. Le facteur de phase recherché vaut donc 1 ou -1 : le premier cas correspond à ce qu'on appelle les *bosons* alors que le second correspond au cas des *fermions*. En fait, le facteur lié à l'échange détermine ce qu'on appelle la *statistique quantique* des particules.

Mais sont-ce là toutes les possibilités ? La figure 1.7 présente une trajectoire à deux particules qui envoie $\mathbf{r}_1$ sur $\mathbf{r}_1$ et $\mathbf{r}_2$ sur $\mathbf{r}_2$. Peut-on lui associer un facteur de phase topologique non trivial qui ne pourrait donc provenir de la permutation σ_γ ? Pour le savoir, il suffit de déterminer si le chemin représenté sur la figure 1.7 peut être déformé en la tresse triviale. Lorsque l'espace de base est de dimension $d \geq 3$, il est évident que l'on peut dénouer la tresse de gauche dans la tresse de droite : on peut ramener les deux brins l'un sur l'autre en déplaçant les particules dans un plan, comme représenté sur la figure 1.7 puis les défaire en utilisant la troisième dimension. Cet argument montre que pour des particules indiscernables vivant dans un espace de dimension $d \geq 3$, les seules statistiques quantiques permises sont celle de *Fermi-Dirac* et celle de *Bose-Einstein*.

4. Ceci provient du fait que pour une représentation du groupe des permutations dans $U(1)$, $R : \mathfrak{S}_N \rightarrow U(1)$, on a $R(\sigma_1 \circ \sigma_2) = R(\sigma_1) \cdot R(\sigma_2)$.

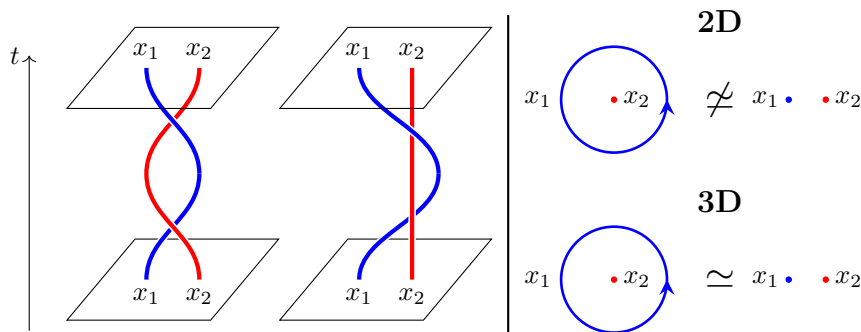


FIGURE 1.7 – Une tresse à deux brins non triviale en dimension deux, triviale en dimension trois, obtenue en faisant tourner une des particules autour de l'autre.

Mais clairement cela ne fonctionne pas lorsque les particules sont confinées dans un plan. Dans ce cas, on peut assigner un facteur de phase arbitraire à l'échange élémentaire de deux particules : rien ne le contraint. Le facteur de phase topologique peut prendre n'importe quelle valeur. Pour cette raison, on ne parle plus de bosons ni de fermions mais d'*anyons*.

✂ Statistiques anyoniques

Cette discussion peut sembler très académique mais il existe des réalisations expérimentales des statistiques anyoniques. Comme l'espace dans lequel nous vivons est tridimensionnel, les constituants élémentaires sont soit bosoniques, soit fermioniques. Cependant, depuis les années 1980, on sait fabriquer des matériaux dans lesquels un gaz électronique est confiné dans un plan. Sous certaines conditions, les interactions coulombiennes entre électrons conspirent pour créer une nouvelle phase du fluide électronique appelé *effet Hall quantique fractionnaire*. Dans cette phase, dont la découverte expérimentale en 1982 par Tsui et Störmer [361] fit l'effet d'une bombe, les excitations élémentaires ne sont plus électroniques mais anyoniques, ce qui est possible car elles sont confinées dans un plan. Ces excitations sont collectives et Laughlin, qui les a prédites théoriquement en 1983, a montré qu'elles portaient une charge fractionnaire [249]. Ce n'est qu'en 1996 qu'une expérience a permis de mettre en évidence la charge fractionnaire [317] et, en 2005, une expérience semble avoir mis en évidence la statistique fractionnaire au travers d'une expérience d'interférométrie [80].

Il y a trois idées à retenir : la première, c'est que comme l'a dit un jour A. Polyakov, un des grands physiciens théoriciens russes de la fin du XX^e siècle, *Whatever you may think about in Physics, it might be realized somewhere*. La deuxième c'est que derrière tout cela se cache un vaste domaine qui est celui

de la matière topologique actuellement très étudiée tant au niveau théorique qu'expérimental et qui fit l'objet du prix Nobel de Physique de Haldane, Kosterlitz et Thouless en 2016. La troisième idée, c'est que certains espèrent pouvoir utiliser ces phases topologiques de la matière pour faire du calcul quantique. Mais l'ordinateur quantique topologique n'existe pas encore. Pour compléter Polyakov, il conviendrait d'ajouter que ce qui germe dans le cerveau des théoriciens n'est pas forcément facile à observer au laboratoire.

1.2.6 Physique quantique et topologie

On va ici aller un peu plus loin que ce qui est nécessaire dans la compréhension de ce livre, et faire une courte digression sur le rôle joué par le facteur de phase topologique dans l'amplitude de probabilité. En fait, les facteurs de phase topologiques ont pris de l'importance en physique quantique bien avant que l'on ne découvre le petit argument ci-dessus pour les statistiques quantiques.

Pour ceux que les mathématiques intéressent, essayons d'aller plus loin que les exemples que nous venons juste de discuter. À un instant donné, l'état du système est constitué par une configuration x d'un espace X et les trajectoires sont des chemins dans cet espace de configurations. Lorsque l'on considère les chemins reliant i à f à une déformation près, on obtient une structure appelée *groupoïde fondamental* de l'espace X qui consiste à associer à une configuration initiale et une configuration finale l'ensemble $\pi_1(i, f)$ des classes de chemins les reliant à une déformation près. On peut alors composer ces classes de chemins car la notion d'équivalence à déformation près est compatible avec la composition $(\gamma_1, \gamma_2) \rightarrow \gamma_1 \# \gamma_2$. Le facteur topologique est tout simplement une représentation de ce groupoïde fondamental sur les nombres complexes. Habituellement, on se limite aux chemins qui partent d'un point et qui y reviennent. Dans ce cas, $\pi_1(x, x)$ est un groupe que l'on appelle le groupe fondamental basé en x . Pour un espace connexe par arcs, les groupes basés en divers points sont tous isomorphes et on parle de groupe fondamental de l'espace X (noté $\pi_1(X)$).

Grâce à la propriété (1.6), on voit que le facteur topologique fournit une représentation du groupe fondamental en chaque point. Les représentations associées aux groupes fondamentaux basés en différents points sont en fait équivalentes. Considérons x et y deux points de l'espace de base X et choisissons un chemin η qui relie x à y , on peut alors associer à tout chemin $\gamma \in \text{Traj}(x, x)$ un chemin $\eta\gamma\eta^{-1} \in \text{Traj}(y, y)$ et bien sûr, grâce à l'équation (1.6), nous avons

$$N_{\gamma'} = N_\eta N_\gamma N_{-\eta} = N_\eta N_\gamma N_\eta^{-1} = N_\gamma. \quad (1.23)$$

On voit donc que l'information contenue dans le facteur de phase topologique pour les différents points de base est indépendante de ces différents points de base. Mathématiquement, le facteur de phase topologique fournit une

représentation du groupe fondamental à valeurs dans le groupe multiplicatif des nombres complexes de module un.

De plus, le même argument montre que l'information sur les différentes valeurs que peut prendre ce facteur topologique pour les chemins reliant deux points différents est contenue dans la représentation du groupe fondamental. Prenons pour cela deux chemins γ et γ' reliant x à y , le chemin $\eta = (\gamma'^{(r)})\sharp\gamma$ obtenu en concaténant γ avec le chemin γ' parcouru en sens inverse est dans $\text{Traj}(x, x)$ et donc nous avons :

$$N_{\gamma'} = N_{\eta}^{-1} N_{\gamma}. \quad (1.24)$$

Tout l'enjeu est donc de comprendre quel est le groupe fondamental de l'espace des positions que l'on considère pour pouvoir classer tous les facteurs topologiques autorisés par la physique quantique.

Ce que nous avons fait dans le paragraphe précédent, c'est en fait classer les représentations d'un groupe que l'on appelle fort naturellement le groupe des tresses à N brins. Dans le cas d'une dimension $d \geq 3$, ce groupe des tresses s'identifie avec le groupe des permutations, ce qui n'autorise que les bosons et les fermions. Dans le cas d'une dimension $d = 2$, le groupe des tresses est bien plus vaste : il est en fait infini. Ces groupes des tresses à n brins sont en fait extrêmement intéressants pour les mathématiques [55].

✂ Calcul et topologie

Le lecteur averti aura remarqué que le fait que les amplitudes soient des nombres qui commutent joue un rôle crucial dans les raisonnements ci-dessus. Il peut être intéressant de considérer une situation un peu plus générale dans laquelle les configurations du système ne sont pas uniquement données par un point dans l'espace X . Dans le cas des particules indiscernables, on pourrait imaginer que les particules portent des degrés de liberté quantiques internes, analogues à une espèce de spin « collectif ». Dans ce cas, on peut faire une sorte d'approche mixte de l'intégrale de chemins pour N particules indiscernables dans laquelle on traite quantiquement ces degrés de liberté internes. L'amplitude topologique associée à chaque chemin devient alors une matrice qui transporte l'état quantique au point de départ vers l'état quantique au point d'arrivée. La structure mathématique associée devient alors une représentation du groupoïde fondamental de l'espace des configurations que nous avons discuté plus haut.

Vers la fin des années 1980, début des années 1990, J. Fröhlich [165] et indépendamment G. Moore et N. Read [277] ont développé cette idée. Pendant plusieurs années, ces anyons non abéliens ou ces para-statistiques comme on les appelait, ont été considérés comme des curiosités. Ce sont les travaux du russe A. Kitaev en 1997 qui ont relancé l'intérêt pour ces idées : il a en effet montré qu'il était possible de mettre en œuvre des algorithmes quantiques en déplaçant de manière astucieuse de telles excitations [230].

1.3 La formulation hamiltonienne

La formulation de Feynman de la théorie quantique est celle qui repose le plus sur l'intuition classique. Bien qu'étant la plus adaptée pour écrire les modèles les plus fondamentaux des interactions, elle ne l'est cependant pas pour l'étude de systèmes physiques simples.

Jusqu'à maintenant, nous considérons les amplitudes de probabilité entre des configurations initiale et finale bien déterminées. En d'autres termes, nous avons une caractérisation complète de l'état du système au niveau de la préparation, de son évolution et de la mesure.

Cependant une telle caractérisation n'est pas forcément la plus adéquate pour l'expérimentateur. En effet, on gagnerait en souplesse à travailler avec des objets caractérisant de manière relativement indépendante les trois étapes de préparation, d'évolution et de mesure. La formulation hamiltonienne standard, écrite en termes d'états, de produit scalaire, d'espace de Hilbert et d'équation de Schrödinger, permet de démêler ces étapes et de répondre à cette problématique.

1.3.1 Notion d'état quantique et linéarité

Un prémisses de la formulation de Feynman était la donnée des états initial i et final f . Nous allons ré-écrire cela en disant que l'étape de mesure est caractérisée par un ensemble complet d'états exclusifs que nous noterons $|m\rangle$ et l'étape de préparation par un état quelconque noté $|\psi\rangle$ ⁵. Chaque valeur de m va donc correspondre à un état parfaitement discernable des autres qui peuvent être des configurations classiques différentes comme des positions ou des nombres quantiques comme la valeur du moment angulaire le long d'un axe.

Maintenant, peut-on construire deux objets associés respectivement à la préparation et à l'évolution du système ? Pour répondre à cela, il est nécessaire d'écrire les choses plus précisément. Pour un état initial quelconque $|\psi\rangle$, on définit un objet noté $|\psi(t)\rangle$ que l'on nomme état quantique du système évolué au temps t comme l'application

$$|\psi(t)\rangle : |m\rangle \mapsto A_{\psi \rightarrow m}, \quad (1.25)$$

schématisée sur la figure 1.8. Défini de la sorte, cet objet contient l'ensemble de toutes les amplitudes de probabilité de transition associées à chaque résultat de mesure sachant que le système était dans l'état $|\psi\rangle$. Pour le moment, rien de complètement nouveau n'a été introduit dans le formalisme. Nous venons simplement de réorganiser différemment les amplitudes de transition.

Avant d'aller plus loin, étudions les propriétés de $|\psi(t)\rangle$. Tout d'abord, puisque l'ensemble des états $|m\rangle$ est complet et non redondant, ces amplitudes

5. L'origine de cette notation sera donnée plus loin.

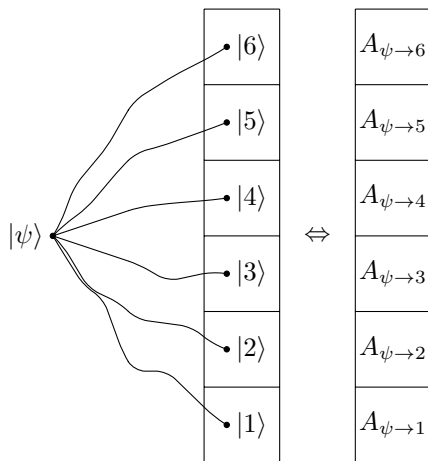


FIGURE 1.8 – Un état quantique du système $|\psi(t)\rangle$ est la donnée d’amplitudes de transition $A_{\psi \rightarrow m}$ pour chacun des états de l’ensemble $|m\rangle$ associé à un processus de mesure sachant que le système était préparé dans un état quelconque $|\psi\rangle$.

doivent obéir à la condition de normalisation sur la somme des probabilités :

$$\sum_m |A_{\psi \rightarrow m}|^2 = 1. \quad (1.26)$$

Ensuite, on peut dès maintenant noter qu’il y a un arbitraire dans la correspondance entre les états physiques et ces amplitudes : nous n’avons défini nulle part de référence pour la phase des nombres complexes dans la théorie. Le jeu d’amplitudes obtenu en multipliant toutes les amplitudes $A_{\psi \rightarrow m}$ par une même phase $e^{i\theta}$ représente donc le même état physique que l’original. Remarquons que les termes d’interférences quantiques que nous avons déjà vus dans l’équation (1.3b) dépendent des différences de phase entre ces amplitudes. Ce sont elles qui contiennent l’information sur l’état physique du système. En conséquence, une modification de la seule phase globale ne change pas l’état physique.

Enfin, un point fondamental est que, pour un système physique donné, il existe plusieurs ensembles d’états complets et mutuellement exclusifs permettant de le décrire : ces différents choix correspondent à différents choix d’appareils de mesures qui permettent de répondre sans ambiguïté à la question « Dans quelle configuration le système se trouve-t-il ? ». Dans le cas du photon unique, on peut considérer deux polarisations linéaires orthogonales et on voit bien qu’il y a autant de choix possibles que de couples d’axes orthogonaux. De même, on pourrait considérer les polarisations circulaires droite et gauche qui forment un couple parfaitement discernable. Un état quantique générique est donc la donnée d’une collection d’amplitudes relativement à un ensemble donné.

Un principe fondamental de la physique quantique est le principe de superposition : la combinaison linéaire de deux états physiques est aussi un état physique. C'est ici une conséquence de la règle de sommation sur tous les chemins mais reste un point hautement non trivial : la théorie quantique permet toutes les superpositions quantiques et n'ostracise donc aucun jeu d'amplitudes (autre que le jeu d'amplitudes nulles)⁶. Regardons plus précisément comment se fait la correspondance : soit $|\psi_1(t)\rangle$ et $|\psi_2(t)\rangle$ deux préparations, décrites relativement à l'ensemble à la mesure $|m\rangle$ par la donnée des correspondances $m \mapsto A_{\psi_1 \rightarrow m}$ et $m \mapsto A_{\psi_2 \rightarrow m}$. Dans ce cas, on peut définir simplement un nouvel état $|\psi(t)\rangle$ par

$$m \mapsto (A_{\psi \rightarrow m} = \alpha_1 A_{\psi_1 \rightarrow m} + \alpha_2 A_{\psi_2 \rightarrow m}) \quad (1.27)$$

avec α_1 et α_2 des nombres complexes. Cela définit un nouvel état quantique après l'avoir normalisé par le facteur $\sum_m |A_{\psi \rightarrow m}|^2$. Notons que cela ne change pas les différences de phase entre les amplitudes de trouver le système dans les états m et m' . *A priori*, la Nature autorise tous les états quantiques donnés par un jeu d'amplitudes non toutes nulles, ce qui ne veut pas dire qu'ils sont tous facilement observables dans une expérience.

D'un point de vue mathématique, l'ensemble des états quantiques à l'instant t muni de l'opération de superposition est un espace vectoriel complexe, et les ensembles complets et mutuellement exclusifs qui nous servent à décrire les états définissent des bases de cet espace vectoriel. Cet ensemble associé à un système est appelé l'*espace des états* et est traditionnellement noté $\mathcal{H}$ ⁷. Les amplitudes $A_{\psi \rightarrow m}$ sont alors les coordonnées du vecteur $|\psi(t)\rangle$ dans la base $|m\rangle$, soit

$$|\psi(t)\rangle = \sum_m A_{\psi \rightarrow j} |m\rangle . \quad (1.28)$$

L'avantage majeur de ce point de vue « géométrique » est que l'on peut s'abstraire du choix de la décomposition $|m\rangle$ utilisée lors de la définition de $|\psi(t)\rangle$. Du vecteur abstrait $|\psi(t)\rangle \in \mathcal{H}$, qui contient toute l'information sur l'étape de préparation et d'évolution, nous pouvons extraire toutes les amplitudes de transition à l'instant t associées à n'importe quel choix de mesure. Nous pouvons même aller plus loin et répondre complètement au problème posé en permettant de séparer l'étape de préparation et d'évolution. En effet, nous pouvons adopter le point de vue que la préparation est caractérisée par le choix d'un état quelconque de l'espace des états $\mathcal{H}$, sans aucune référence à une évolution temporelle. La dynamique ne joue plus de rôle particulier.

6. Certains modèles possèdent des règles de supersélection qui contraignent le principe de superposition.

7. Pour être plus rigoureux, nous devrions le noter $\mathcal{H}_t$ puisque *a priori*, compte tenu de notre construction, un espace des états doit être associé à chaque instant t . Il s'avère que, pour un système de mécanique quantique, ces espaces sont isomorphes d'où notre liberté notationnelle. Ceci n'est cependant plus vrai en théorie quantique des champs.

Finalement, nous parlerons de l'*état quantique* d'un système tout élément de l'espace des états $\mathcal{H}$ ⁸. Un état quelconque $|\psi\rangle$, associé à l'étape de préparation, peut être décomposé dans n'importe quelle base $|m\rangle$, associée à l'étape de mesure, comme

$$|\psi\rangle = \sum_m a_{\psi,m} |m\rangle, \quad (1.29)$$

où $a_{\psi,m}$ est un nombre complexe aussi nommé amplitude de probabilité. Son module carré est interprété comme « la probabilité de trouver le système dans l'état quantique $|m\rangle$ sachant qu'il se trouve dans l'état quantique $|\psi\rangle$ ». Bien entendu, si ces deux états sont parfaitement discernables, cette probabilité est nulle. De même, il est complètement naturel de supposer que $a_{\psi,\psi} = 1$.

La construction de l'espace des états nous permet donc de nettement séparer préparation et mesure de l'évolution. Cependant, avant de détailler comment nous pouvons modéliser la dynamique dans ce nouveau cadre, il est nécessaire de mieux comprendre la structure de l'espace des états lui-même.

1.3.2 Produit scalaire et espace de Hilbert

Nous allons montrer maintenant que $\mathcal{H}$ est muni d'un produit scalaire naturel, ce qui permettra de dégager la structure mathématique générale de l'espace des états.

On peut sans risque affirmer qu'en choisissant une base $|j\rangle$ de $\mathcal{H}$, nous avons

$$a_{\psi,\phi} = \sum_j a_{\psi,j} a_{j,\phi}. \quad (1.30)$$

On a de plus les deux conditions de normalisation

$$\sum_j |a_{\psi,j}|^2 = 1 \text{ et } a_{\psi,\psi} = 1. \quad (1.31)$$

En utilisant les trois équations ci-dessus, on arrive à la conclusion suivante, valable quel que soit l'état quantique $|\psi\rangle$ et la base utilisée

$$\sum_j a_{\psi,j} (a_{\psi,j}^* - a_{j,\psi}) = 0. \quad (1.32)$$

En particulier, on peut prendre pour $|\psi\rangle$ un état $|i\rangle$ d'une autre base. Par permutation, nous obtenons les équations suivantes

$$\sum_j a_{i,j} (a_{i,j}^* - a_{j,i}) = 0, \quad (1.33a)$$

$$\sum_i a_{j,i} (a_{j,i}^* - a_{i,j}) = 0. \quad (1.33b)$$

8. Plus précisément, il s'agit des rayons de cet espace puisque la multiplication par une phase ne change pas l'état quantique.

L'astuce consiste à regarder la quantité $\sum_i (1.33a) + \sum_j (1.33b)^*$, ce qui donne

$$\sum_{j,i} |a_{i,j} - a_{j,i}^*|^2 = 0. \quad (1.34)$$

Il faut alors que chacun des termes de la somme soit nul. On vient donc de montrer que pour tout couple d'états appartenant chacun à une base, on a la relation

$$a_{i,j} = a_{j,i}^*. \quad (1.35)$$

Cette relation est en fait bien plus générale, puisque pour tout état $|\psi\rangle$, on peut construire une base dont le premier état est $|\psi\rangle$. On a donc, pour tout couple d'états quantiques $|\psi\rangle, |\phi\rangle$:

$$a_{\psi,\phi} = a_{\phi,\psi}^*. \quad (1.36)$$

On peut alors remarquer que la forme développée (1.30) de l'amplitude se réécrit finalement

$$a_{\psi,\phi} = \sum_j a_{\psi,j} a_{\phi,j}^* \quad (1.37)$$

ce qui est exactement la forme naturelle d'un produit scalaire sur notre espace vectoriel complexe. Ceci nous permet également d'affirmer que les amplitudes $a_{j,i}$ et $a_{i,j}$ vérifient

$$\sum_i a_{j,i} a_{k,i}^* = \delta_{j,k} \quad (1.38)$$

ce qui montre que les bases $|j\rangle$ et $|i\rangle$ sont toutes deux orthonormées et que les amplitudes de probabilité définissent des matrices de passages unitaires d'une base à l'autre.

L'espace des états $\mathcal{H}$ est donc finalement un espace vectoriel complexe muni d'un produit scalaire, c'est-à-dire un *espace de Hilbert*. C'est habituellement comme cela qu'on présente la théorie quantique, en postulant directement l'existence de l'espace des états (et l'équation d'évolution). La notion de produit scalaire est également la raison pour laquelle on note les états quantiques sous la forme $|\phi\rangle$. En effet, en notant le produit scalaire sur notre espace des états sous la forme habituelle $\langle \cdot | \cdot \rangle$, on voit tout de suite que

$$a_{\psi,\phi} = \langle \phi | \psi \rangle. \quad (1.39)$$

Nous venons d'achever l'analyse du cadre cinématique de la théorie quantique exprimée en termes d'état $|\psi\rangle$ appartenant à un espace de Hilbert $\mathcal{H}$. Sachant comment tenir compte de l'étape de préparation et de mesure dans ce cadre, il ne nous reste plus qu'à décrire comment modéliser la dynamique de l'état quantique.

1.3.3 Évolution temporelle

On souhaite maintenant comprendre comment un système préparé à un instant initial t_i dans un état $|\psi_i\rangle$ évolue à un instant ultérieur t_f . L'état final sera noté $|\psi_f\rangle$. Un état quantique est décrit par la donnée d'amplitudes de probabilité exprimées dans une base donnée. Le choix de la base pouvant être différent à différents instants, l'état initial $|\psi_i\rangle$ sera caractérisé par la donnée des amplitudes $\langle j|\psi_i\rangle$ exprimées dans une base $|j\rangle$, tandis que l'état final $|\psi_f\rangle$ sera caractérisé par $\langle k|\psi_f\rangle$ dans une base $|k\rangle$.

Compte tenu de la définition de l'état quantique, de la sommation sur tous les chemins et de la règle de localité, nous pouvons écrire l'équation d'évolution des amplitudes de probabilité

$$\langle k|\psi_f\rangle = \sum_j \langle k, t_f | j, t_i \rangle \langle j|\psi_i\rangle. \quad (1.40)$$

Cette nouvelle écriture de l'intégrale de chemin permet de trouver la forme de l'évolution de l'état à partir de celle des amplitudes. Pour cela, on définit un opérateur sur l'espace de Hilbert appelé *opérateur d'évolution* entre les instants t_i et t_f et noté $U[t_i, t_f]$ comme l'opérateur dont les éléments de matrice sont

$$\langle k|U[t_i, t_f]|j\rangle = \langle k, t_f | j, t_i \rangle. \quad (1.41)$$

L'équation (1.40) réécrite avec cet objet $\langle k|\psi_f\rangle = \sum_j \langle k|U[t_i, t_f]|j\rangle \langle j|\psi_i\rangle$ permet de déduire la forme générale de l'évolution d'un état quantique comme

$$|\psi_f\rangle = U[t_i, t_f] |\psi_i\rangle. \quad (1.42)$$

Une représentation schématique de la démarche que nous venons d'adopter est donnée sur la figure 1.9.

Essayons d'approfondir un peu plus cette définition. Tout d'abord, on voit par construction que l'opérateur d'évolution est linéaire, et donc que l'état évolué à partir d'une superposition est bien égal à la superposition des états évolués dans le temps. Ensuite, par normalisation, $\sum_k |\langle k, t_f | \psi_i \rangle|^2 = 1$. Cela implique, par construction de l'état $|\psi_f\rangle$, que

$$\sum_k |\langle k|\psi_f\rangle|^2 = \sum_k |\langle k|U[t_i, t_f]|\psi_i\rangle|^2 = 1. \quad (1.43)$$

On peut alors utiliser ces deux propriétés ensemble, en regardant l'évolution d'un état superposé de la forme $|\Psi\rangle = (|\phi\rangle + \lambda|\psi\rangle)/\mathcal{N}$ avec $\lambda \in \mathbb{C}$, le coefficient $\mathcal{N}$ étant choisi pour que cet état soit de norme 1, soit

$$1 = \sum_j |\langle j|\Psi\rangle|^2 = \frac{\lambda\langle\phi|\psi\rangle + \lambda^*\langle\psi|\phi\rangle + 1 + |\lambda|^2}{\mathcal{N}^2}. \quad (1.44)$$

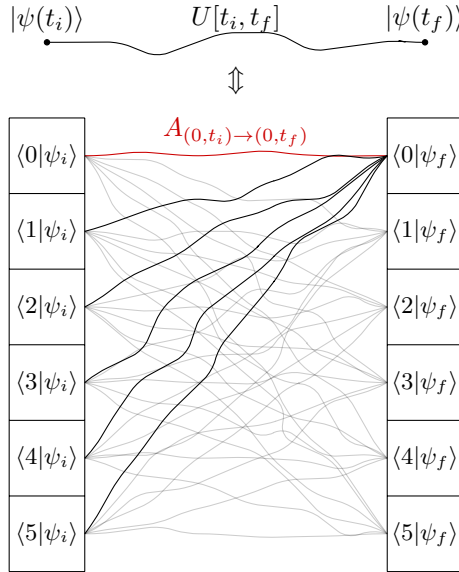


FIGURE 1.9 – Étude de l'évolution temporelle d'un état quantique. À l'instant t_i , l'état $|\psi_i\rangle$ peut être représenté par une collection d'amplitudes $\langle j|\psi_i\rangle$ dans une base donnée $|j\rangle$. En calculant l'ensemble des évolutions possibles pour les états de la base, on définit l'état évolué $|\psi_f\rangle$ dont les amplitudes sont données par $\langle k|\psi_f\rangle = \sum_j \langle k, t_f|j, t_i\rangle \langle j|\psi_i\rangle$. Cela permet de définir un opérateur d'évolution temporelle vu comme une matrice dont les éléments dans la base $|j\rangle$ sont $\langle k|U[t_i, t_f]|j\rangle = \langle k, t_f|j, t_i\rangle$ et qui agit directement sur les états quantiques : $|\psi_f\rangle = U[t_i, t_f] |\psi_i\rangle$.

Si on regarde l'état noté $|U\Psi\rangle$, qui correspond à l'évolution $|\Psi\rangle$ sous l'action d'un opérateur U linéaire (on oublie ici le temps pour ne pas alourdir les notations), on a

$$1 = \sum_j |\langle j|U\Psi\rangle|^2 = \frac{\lambda \langle U\phi|U\psi\rangle + \lambda^* \langle U\psi|U\phi\rangle + 1 + |\lambda|^2}{\mathcal{N}^2}. \quad (1.45)$$

On a alors égalité des équations (1.44) et (1.45), quel que soit $\lambda \in \mathbb{C}$. On voit que cela impose que l'opérateur d'évolution conserve le produit scalaire, ce qui signifie que l'évolution se fait de manière *unitaire* :

$$\langle U\phi|U\psi\rangle = \langle \phi|\psi\rangle. \quad (1.46)$$

L'unitarité de l'opérateur $U[t_i, t_f]$ apparaît ici comme une conséquence de la linéarité et de la conservation des probabilités de la théorie quantique, c'est-à-dire la compatibilité entre l'évolution d'un système quantique isolé et le principe de superposition.

Maintenant que l'on a un peu mieux défini cet opérateur évolution, on peut essayer de trouver une équation qui régirait l'évolution des états quantiques de manière locale. Pour faire cela, regardons $U[t, t + \delta t]$. Tout d'abord, comme nous l'avons vu précédemment, l'état quantique n'est pas modifié dans la limite $\delta t \rightarrow 0$. L'opérateur d'évolution dans cette limite est donc l'identité :

$$U[t, t] = \mathbf{1}. \quad (1.47)$$

En remarquant que la relation ci-dessus impose que $A_{j \rightarrow k} = 0$, on retrouve que les états d'un ensemble complet et non redondant doivent s'exclure mutuellement. Si on regarde l'ordre suivant en δt , on peut écrire

$$U[t, t + \delta t] = \mathbf{1} - \frac{i}{\hbar} H \delta t + O(\delta t^2) \quad (1.48)$$

où l'on a introduit le facteur $-i/\hbar$ pour des raisons qui apparaîtront sous peu, et où l'opérateur H , appelé opérateur hamiltonien du système, peut être retrouvé à partir de la forme exacte de l'opérateur évolution en termes d'intégrale de chemin. Comme $|\psi(t + \delta t)\rangle = (\mathbf{1} - iH\delta t/\hbar)|\psi(t)\rangle$, nous voyons que, dans la limite $\delta t \rightarrow 0$

$$i\hbar \frac{d|\psi(t)\rangle}{dt} = H|\psi(t)\rangle. \quad (1.49)$$

Cette équation est l'*équation de Schrödinger*. Elle est au cœur de la formulation opératorielle de la théorie quantique que nous utiliserons tout au long de ce livre. Cette équation est la traduction dans le langage des états de la dynamique donnée par l'intégrale de chemin. Avec la définition des états, elle complète la formulation hamiltonienne, équivalente à celle lagrangienne de Feynman, de la théorie quantique.

☞ Le hamiltonien en physique classique

La formulation opératorielle exposée précédemment introduit un objet mathématique appelé opérateur hamiltonien. Pour mieux le comprendre, nous allons de nouveau faire un détour par la mécanique classique.

Lors de l'énoncé des règles de Feynman, nous avons vu que la phase des amplitudes était proportionnelle à l'action. Entre autres, nous savons que cette fonctionnelle $S[\gamma(t)]$ permet d'écrire l'ensemble de la physique classique à l'aide d'un principe variationnel, principe lui-même justifié par la théorie quantique (voir section 1.2.3). En étudiant plus précisément la limite classique, nous avons ensuite vu qu'une amplitude de transition semi-classique pouvait être approchée grâce à la fonction de Hamilton $S[\gamma_c] \equiv S_c(t_i, \gamma_i, t_f, \gamma_f)$, évaluation de l'action sur la trajectoire physique suivie. Cette quantité contient toute la physique d'un système à la limite classique. Pour mieux comprendre cet objet et éclairer le statut du hamiltonien, nous allons regarder plus précisément le contenu de cette fonction en regardant ses variations.

Faisons varier les conditions finales t_f et q_f que l'on renommera t et q . Soit $S_c(t_i, \gamma_i, t_f, \gamma_f) \equiv S_c(t, q)$. À temps fixé, une variation de la position q nous donne, grâce à l'équation (1.8) (terme intégral nul et variation au temps initial nulle), $\delta_q S_c(t, q) = \frac{\partial L}{\partial \dot{q}} \delta q$. La quantité $p = \frac{\partial L}{\partial \dot{q}}$ évaluée au temps t est appelée le *moment conjugué* ou *impulsion* p de la variable q . Cette variable est reliée à la vitesse $\dot{q}$, la plupart du temps de manière univoque⁹. Un calcul similaire à l'équation (1.8) permet d'obtenir la variation temporelle $\delta_t S_c(t, q) = -(p\dot{q} - L)\delta t$. La quantité $H = p\dot{q} - L$ est appelée le *hamiltonien* associé au lagrangien L . L'un est obtenu par transformée de Legendre de l'autre. Finalement, nous pouvons donc écrire la variation totale de la fonction de Hamilton comme $dS_c(t, q) = p dq - H dt$.

Nous pouvons comprendre deux choses de ces résultats. Dans le formalisme lagrangien, les équations du mouvement restent du second ordre. Elles sont déduites d'un formalisme variationnel dans lequel on fixe les positions initiales et finales. Or dans un problème concret de mécanique classique comme la mécanique céleste, on a pour condition initiale la donnée des vitesses et positions initiales. Il est donc légitime de se demander s'il existe une formulation variationnelle de la mécanique classique adaptée à de telles conditions initiales. La réponse fut trouvée par Hamilton en 1833. Pour cela, il suffit d'exprimer le problème non plus dans l'espace des configurations $\mathcal{C}$ mais dans l'*espace des phases* $\mathcal{P}$ paramétré par les variables (q, p) . Les équations du mouvement, qui sont deux fois plus nombreuses mais du premier ordre seulement dans cette formulation, sont déduites d'un principe variationnel provenant de l'action $S[q(t), p(t)] = \int_i^f (p dq - H dt)$. Bien que très similaire à l'action lagrangienne, ce n'est pourtant pas le même problème mathématique (on travaille sur l'espace des phases) puisqu'on extrémalise sur deux fonctions indépendantes $q(t)$ et $p(t)$. La grandeur centrale dans cette formulation est le hamiltonien $H(q, p, t)$ défini précédemment¹⁰.

Enfin, nous pouvons mieux comprendre la forme de l'opérateur hamiltonien apparaissant dans l'équation (1.49) de Schrödinger. En effet, nous avons vu que $\partial_t S_c(q, t) = -H(q, p, t)$ mais aussi que $p = \partial_q S_c(q, t)$. Ainsi, nous obtenons une équation différentielle de premier ordre non linéaire pour S_c appelée *équation de Hamilton-Jacobi* $\partial_t S_c = -H(q, \partial_q S_c, t)$. Elle est complètement équivalente aux formulations lagrangienne ou hamiltonienne de la mécanique classique et a l'avantage d'être très proche de l'équation (1.49) de Schrödinger. En effet, l'équation de Hamilton-Jacobi peut s'écrire comme la limite classique de celle

9. Les cas pathologiques amènent à la théorie des systèmes contraints.

10. Le formalisme hamiltonien a joué un rôle extrêmement important à trois titres : tout d'abord, ce sont les techniques hamiltoniennes qui, en permettant de développer des méthodes perturbatives classiques, ont donné lieu aux prédictions sur le mouvement des planètes (éphémérides astronomiques). Celles-ci ont permis la découverte de Neptune suite aux prédictions de Le Verrier et, *in fine*, la découverte de l'anomalie du mouvement orbital de Mercure qui ne put être expliquée que par la théorie de la relativité générale d'Einstein. Ensuite, le formalisme mathématique a fourni le cadre approprié à l'exploration de la notion de chaos dans les systèmes dynamiques à partir du problème de la stabilité du système solaire qui occupa les esprits dès le XIX^e siècle [248].

de Schrödinger. Le procédé est analogue à la façon d'obtenir les équations eikonalles de l'optique géométrique à partir des équations de Maxwell. On constate donc que l'opérateur hamiltonien de la théorie quantique est relié au hamiltonien de la mécanique classique par un changement du type $p \rightarrow \partial_q$ comme nous le suggère l'équation de Hamilton-Jacobi. Ceci fut, au cours du développement de la théorie quantique, érigé en un principe inductif (plus communément, une recette de cuisine) dit de *quantification canonique* qui permet, à partir d'un système classique donné, d'écrire sa description quantique hamiltonienne. Le principe s'énonce de différentes façons. L'une, proche de ce que l'on vient de voir, fait correspondre à chaque quantité conjuguée p un opérateur différentiel $-i\hbar\partial_q$. D'une manière plus générale et plus formelle, à chaque couple (q_i, p_i) paramétrant l'espace des phases $\mathcal{P}$ de la théorie classique, on fait correspondre des opérateurs q_i et p_i sur l'espace des états vérifiant les relations de commutation de Heisenberg (qui peuvent encore une fois être reliées à des structures classiques sur l'espace des phases) :

$$[q_i, p_i] = i\hbar \delta_{ij} . \quad (1.50)$$

L'ensemble de ces discussions à l'interface entre la théorie classique et la théorie quantique n'ont en aucun cas pour but de « déduire » la théorie quantique à partir du classique. Elles permettent simplement d'illustrer et motiver la forme des objets mathématiques comme le lagrangien ou le hamiltonien que l'on utilise.

1.3.4 Retour sur le système à deux niveaux

Pour comprendre un peu mieux les concepts de la théorie quantique que nous venons d'exposer, il est bon de revenir à l'exemple du système à deux niveaux (voir section 1.2.1).

Un système à deux niveaux possède deux états $|0\rangle$ et $|1\rangle$ formant une base de l'espace des états et tout état quantique du système peut s'écrire $|\psi\rangle = a_0|0\rangle + a_1|1\rangle$, avec $|a_0|^2 + |a_1|^2 = 1$. Si on veut reprendre les notations que nous venons de voir, on aurait $a_0 = a_{\psi,0} = \langle 0|\psi\rangle$, $a_1 = a_{\psi,1} = \langle 1|\psi\rangle$, et la condition de normalisation donnée est bien $\sum_j |a_{\psi,j}|^2 = 1$.

Lors de l'étude de ce système, nous avons calculé les amplitudes de transition $A_{ij}(t) = A_{(i,0) \rightarrow (j,t)}$ de passer de l'état i à l'instant 0 vers l'état j à l'instant t . Ce faisant, nous avons construit complètement l'opérateur d'évolution pour ce système. En effet, nous avons construit exactement la matrice :

$$U[0, t] = \begin{pmatrix} A_{00}(t) & A_{01}(t) \\ A_{10}(t) & A_{11}(t) \end{pmatrix} = \begin{pmatrix} \cos(At) & -i \sin(At) \\ -i \sin(At) & \cos(At) \end{pmatrix} . \quad (1.51)$$

Comme les colonnes de la matrice représentant $U[0, t]$ sont orthogonales, U est bien unitaire. Si A n'était pas réel, l'unitarité ne serait pas vérifiée et c'est

précisément ce qui justifie de nouveau la forme de l'amplitude de transition infinitésimale $-iA$ entre les deux états.

L'évolution au temps t de l'état $|\psi\rangle$, que l'on note $|\psi(t)\rangle = a_0(t)|0\rangle + a_1(t)|1\rangle$, est alors donnée par

$$\begin{pmatrix} a_0(t) \\ a_1(t) \end{pmatrix} = U[0, t] \begin{pmatrix} a_0(0) \\ a_1(0) \end{pmatrix}. \quad (1.52)$$

La physique de ce système, fondamentale en information quantique, sera étudiée très en détail au chapitre 4.

1.4 La mesure en physique quantique

Les règles de Feynman que nous avons énoncées permettent de calculer les amplitudes de transition $A_{i \rightarrow f}$ d'un système préparé dans un état i et mesuré dans un état f après une évolution pendant un temps donné. Le système est complètement isolé pendant cette phase. En revanche, les phases de préparation et de mesure correspondent à une interaction entre le système et des appareils extérieurs. Les règles de Feynman ne disent rien *a priori* sur ces processus de préparation et de mesure et furent utilisées de façon intuitive dans tous les énoncés. La formulation hamiltonienne a permis d'éclaircir et de démêler le rôle de chacune de ces étapes. Cependant, des subtilités essentielles se cachent derrière la présentation intuitive de l'étape de mesure que nous avons donnée et doivent être mises au jour.

Idéalement, si la théorie quantique est une description complète de la Nature, on aimerait pouvoir décrire le processus de mesure lui-même. Pour que l'observateur détermine si le système est trouvé dans un état final f , il doit interagir avec lui. *A priori*, modéliser le processus de mesure est une tâche ardue car l'appareil de mesure est aussi un système quantique complexe. Or on aimerait pouvoir en décrire l'influence facilement sans avoir à entrer dans les détails de son fonctionnement. Fort heureusement, les fondateurs de la théorie quantique ont proposé une règle, appelée postulat de la mesure, qui permet de décrire un processus de mesure de façon générique.

Il faudra bien se rappeler que si ce postulat nous offre une manière simple de calculer les probabilités d'obtenir un résultat donné, il ne s'agit en aucun cas d'une description microscopique de ce qui se passe pour le système couplé à un appareil de mesure. C'est un outil mathématique simple, empirique et non une explication physique.

1.4.1 Mesures projectives

Le *postulat de la mesure*, tel qu'il est énoncé habituellement en physique quantique, est le suivant :

La mesure d'une grandeur physique ne peut produire qu'un certain nombre de résultats et à chacun de ces résultats est associé un état.

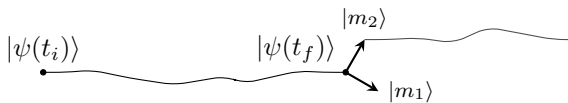


FIGURE 1.10 – Évolution d'un vecteur d'état : l'évolution est unitaire tant qu'il n'y a pas de mesure, entre t_i et t_f (ligne courbe fine). Lors d'une mesure, l'état saute aléatoirement sur un état associé au résultat de la mesure (ici $|m_2\rangle$) avant de reprendre son cours unitaire. L'amplitude de probabilité pour que le saut ait lieu de $|\psi(t_f)\rangle$ vers $|m_i\rangle$ est donnée par le produit scalaire $\langle m_i|\psi(t_f)\rangle$ entre ces deux états.

Les états correspondant aux divers résultats possibles sont deux à deux orthogonaux. L'amplitude de transition correspondante est le produit scalaire entre l'état avant la mesure et l'état associé au résultat effectivement mesuré. Après la mesure, l'état du système est l'état correspondant au résultat trouvé.

Une telle mesure est appelée mesure projective. Hormis la partie correspondant à l'état post-mesure, nous avons déjà implicitement pris en compte dans nos discussions de l'étape de mesure ce qui est explicitement énoncé ici.

De manière plus précise, la mesure modélisée au travers de ce postulat introduit un saut de l'état du système vers l'état associé au résultat obtenu et ce saut est aléatoire, comme cela est représenté schématiquement sur la figure 1.10. Si on considère plusieurs réalisations du système physique, préparées de manière identique et soumises en parallèle au même processus de mesure, on verra se produire différents sauts quantiques dont les fréquences statistiques correspondent aux probabilités données par la règle de Born. Ainsi, si les différents résultats possibles sont les états $|m\rangle$ (qui doivent nécessairement former une base de l'espace des états), alors la probabilité d'observer un saut vers l'état $|m\rangle$ est donnée par

$$p_{\psi,m} = |\langle m|\psi\rangle|^2. \quad (1.53)$$

On voit réapparaître la notion d'amplitude de probabilité donnée par un produit scalaire discutée précédemment. Lorsque nous l'avions introduite, nous l'avions interprétée comme l'amplitude de probabilité que le système « soit dans l'état $|m\rangle$ sachant qu'il est dans l'état $|\psi\rangle$ ». Nous avons également remarqué que ce choix était approprié si l'on ne voulait pas prendre en compte le processus réel qui permettait de déterminer si l'état physique était ou non dans l'état $|m\rangle$. Toutes ces interprétations sont bien en accord avec l'idée que l'on se fait d'une mesure projective, c'est-à-dire d'un processus dont on ne veut pas modéliser les détails et qui répond à la question « l'état est-il dans $|j\rangle$? ».

Insistons sur les points importants à retenir pour la suite :

- l'évolution de l'état quantique lors du processus de mesure n'est plus unitaire, ni linéaire ;

- la mesure engendre un ensemble d'états quantiques dès lors que l'on considère plusieurs réalisations du système physique sur lesquelles ont fait la mesure. Elle introduit donc un indéterminisme du fait du caractère aléatoire du « choix » du résultat ;
- l'orthogonalité des états associés à deux résultats différents est naturelle. deux résultats différents correspondent à des états après mesure parfaitement discernables. L'amplitude de transition entre eux, c'est-à-dire leur produit scalaire, doit donc être nulle ;
- si on refait immédiatement une seconde mesure de la même quantité A après avoir trouvé un résultat de mesure a et que l'on est donc dans l'état $|A_a\rangle$, on va retrouver le même résultat et on va rester dans le même état. En effet, la probabilité d'obtenir un résultat $a' \neq a$ est $|\langle A_{a'}|A_a\rangle|^2 = 0$, du fait de l'orthogonalité des états associés à des résultats de mesure différents.

Cette dernière remarque montre immédiatement que le processus de mesure décrit ici n'est pas forcément celui que nous rencontrons dans nos expériences. L'exemple le plus clair est la détection des photons par un photodétecteur ou même par notre rétine : le photodétecteur ou la cellule photosensible absorbent le photon lors de sa détection. Cela ne correspond pas au processus de mesure projective puisque ce dernier devrait laisser le photon détecté toujours présent pour une nouvelle détection. Ceci montre bien que la photo-détection habituelle est une mesure destructive qui détruit le quanta qu'elle est censée détecter. Une mesure de détection non destructive du photon nécessite en fait un système transparent qui est capable de signaler sa présence voire de compter combien de photons l'ont traversé. De telles mesures ont été réalisées en optique puis, plus récemment, dans le domaine micro-ondes pour des photons uniques piégés dans une cavité électromagnétique, notamment par l'équipe de S. Haroche. Cette nouvelle manière de voir a permis de grands progrès dans la manipulation et la caractérisation des états quantiques individuels. Nous discuterons ces expériences dans le chapitre 10.

1.4.2 Observables et indétermination de Heisenberg

Formalisons un peu plus les choses exposées précédemment. Les états associés aux résultats d'une mesure d'une grandeur physique sont deux à deux orthogonaux et forment une base de l'espace des états. Dans cette base, on peut donc construire un opérateur A représentant cette grandeur physique comme étant une somme de projecteurs sur les différents résultats possibles. En notation de Dirac, $A = \sum_a m_a |A_a\rangle\langle A_a|$ où m_a est un résultat de mesure possible de la grandeur physique en question, et $|A_a\rangle$ est un état tel que la grandeur physique prenne la valeur m_a . De plus, les résultats d'une mesure étant des nombres réels, cet opérateur A est nécessairement hermitien. Une grandeur physique mesurable, ou de manière équivalente l'opérateur hermitien qui y est associé, est appelée une *observable*.

Lorsque nous faisons de la physique quantique pour particule, les deux observables naturelles que l'on utilise sont les opérateurs position x et impulsion p . Suivant ce que l'on vient de décrire, leurs états propres décrivent respectivement les états de position donnée $|x\rangle$ ou d'impulsion donnée $|p\rangle$. Un autre exemple habituel est donné par le système à deux niveaux, pour lequel on donnera une famille d'observables et les états associés à leurs résultats de mesure dans le chapitre 4.

Quelles informations peut-on extraire d'une observable ? La physique quantique nous permet de prédire la statistique des résultats de mesure, par exemple la valeur moyenne et l'écart-type de la grandeur A . Préparons le système dans un état $|\psi\rangle$ quelconque. Dans cet état, la valeur moyenne de la grandeur physique considérée est donnée par $\sum_a m_a p_a$, où p_a est la probabilité de transition entre $|\psi\rangle$ et $|A_a\rangle$, c'est-à-dire la probabilité d'obtenir le résultat m_a . Celle-ci étant donnée par la règle de Born (1.1), la valeur moyenne se réécrit donc comme $\langle\psi|A|\psi\rangle \equiv \langle A\rangle$. Pour caractériser plus finement la distribution statistique au-delà de la valeur moyenne, il est intéressant de regarder l'écart-type de la distribution. Noté ΔA , il est défini comme $\Delta A = \sqrt{\langle A^2\rangle - \langle A\rangle^2}$.

Un résultat majeur de la physique quantique, obtenu entre autres par Heisenberg, et connu sous le nom de *principe d'indétermination*¹¹, exprime que les dispersions des résultats de mesure pour certains couples d'observables ne peuvent être réduites simultanément à zéro. Formellement, pour deux observables A et B ,

$$\Delta A \Delta B \geq \frac{1}{2} |\langle [A, B] \rangle| \quad (1.54)$$

où $[A, B] = AB - BA$ est le commutateur entre A et B . Cette inégalité nous dit que, si les observables A et B ne commutent pas, le produit des dispersions est borné par une valeur minimale. Un tel couple de variables est dit *incompatible*. Cette inégalité a des conséquences inattendues pour notre intuition classique. En effet, elle interdit de minimiser conjointement et de façon arbitrairement faible la dispersion des résultats de mesure de deux observables incompatibles. Une mesure exacte d'une grandeur implique une dispersion infinie d'une grandeur incompatible.

L'exemple standard en mécanique quantique est de nouveau le couple des opérateurs position x et impulsion p , qui ont un commutateur $[x, p] = i\hbar$. L'inégalité de Heisenberg prend alors la forme $\Delta x \Delta p \geq \frac{\hbar}{2}$. Ainsi, une mesure exacte de la position implique une ignorance totale de son impulsion (dispersion infinie). Ceci permet de commenter à nouveau la forme de l'évolution d'une particule libre. En effet, l'amplitude de transition, donnée par l'équation (1.16), de trouver la particule en x_f au temps t_f sachant qu'elle était parfaitement localisée en x_i au temps t_i , ne dépend de x_f qu'à travers une phase. Par conséquent, la probabilité de transition ne dépend pas de cette

11. Il s'agit en réalité d'un théorème mathématiquement démontré, mais l'habitude historique est de l'appeler principe.

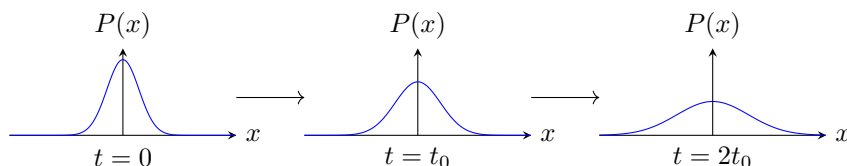


FIGURE 1.11 – Évolution d'un paquet d'ondes gaussien au cours du temps. Le principe d'indétermination de Heisenberg est ici à l'œuvre : en partant d'un paquet bien résolu en position, on a nécessairement une forte dispersion sur les impulsions ce qui entraîne un étalement du paquet d'ondes au cours du temps.

position finale. Ceci signifie que la particule s'est complètement délocalisée pour tout instant $t_f > t_i$. Cela n'est plus totalement surprenant en regard des relations d'indétermination pour la position et l'impulsion. En effet, une particule parfaitement localisée possède une dispersion en impulsion infinie et peut donc, dans un modèle non relativiste, se propager infiniment vite et ainsi se délocaliser instantanément. Pour résoudre cela, on pourrait considérer un problème un peu moins singulier : si on considère initialement une fonction d'onde gaussienne, on trouve que celle-ci reste gaussienne mais s'étale au cours du temps, comme cela est visible sur la figure 1.11. Sur cet exemple moins pathologique, nous voyons que la somme d'amplitudes de module unité donne des amplitudes de transition dont le module n'est pas l'unité : ce sont les interférences quantiques destructives entre les différents chemins qui sont responsables de cela.

1.5 Perspectives

Ce que nous venons de développer sur la théorie quantique permet en principe de comprendre tous les phénomènes physiques connus à ce jour. Grâce aux techniques de la théorie de l'information, du calcul et de la complexité, nous serons amenés à développer notre compréhension des phénomènes quantiques à travers la physique de systèmes simples comme le système à deux niveaux et l'oscillateur harmonique. Ceci permettra d'une part d'avoir une meilleure compréhension des fondements de la théorie quantique et des différences fondamentales qu'elle a avec les théories classiques, mais aussi de discuter des avancées expérimentales modernes ayant pour but, soit de tester les concepts fondamentaux de la théorie, soit de construire un ordinateur quantique tirant parti des spécificités du monde quantique.

Cependant, bien que nous ayons tous les outils nécessaires pour construire des modèles et faire des prédictions, de nombreuses questions autant techniques que conceptuelles peuvent être posées sur les fondements théoriques eux-mêmes.

La mesure projective montre déjà l'une des spécificités de la mesure dans le cas quantique : il est impossible d'acquérir de l'information sur l'état d'un

système physique sans une interaction non négligeable entre le système et l'appareil de mesure. On dit de façon imagée que l'état quantique est « altéré », mis à jour, par la mesure. C'est une différence majeure par rapport au cadre classique où la possibilité de mesurer en minimisant à volonté l'interaction entre le système et l'appareil de mesure est tout à fait envisageable. Pour aller encore plus loin, le postulat de la mesure n'est qu'une règle empirique permettant de passer outre la modélisation de l'appareil de mesure. Or celui-ci est un système physique comme les autres et devrait être décrit aussi par la théorie quantique. Ceci pose donc la question du statut de la mesure, de savoir comment la modéliser proprement ainsi que de la transition quantique/classique. Comprendre tout cela en détail nécessitera l'introduction de nouveaux outils pour caractériser les systèmes quantiques ouverts mais aussi de nouveaux concepts comme la décohérence.

En continuant à pousser cette logique, la notion d'état quantique elle-même et le sens qu'on lui attribue sont ambigus. Donner un sens à la notion d'état quantique est toujours une question sujette à débat. À la suite des explorations théoriques et expérimentales que nous ferons, nous verrons qu'il est possible de dégager certains éléments interprétatifs pouvant permettre d'éclairer le sens de la notion d'état quantique. De la même façon, l'apparition des probabilités et de l'aléatoire dans l'évolution d'un état quantique peut être questionnée au regard de l'évolution purement déterministe de Schrödinger. Comprendre s'il est possible de faire émerger la règle de Born à partir des autres postulats est aussi une question toujours sujette à débat mais qui peut à nouveau être éclairée à l'aide des outils de théorie de l'information et des systèmes ouverts.

Finalement, la perspective de cet ouvrage va être d'utiliser les règles de la théorie quantique exposées dans ce chapitre, d'analyser de nombreux systèmes physiques et leurs mises en œuvre expérimentales pour, en retour, refaire une analyse et une synthèse des fondements de la théorie quantique.

Chapitre 2

Théorie de l'information

Un système physique préparé dans un état donné, qu'il soit classique ou quantique, peut être vu comme une source d'information constituée par les différents résultats des mesures que l'on effectue sur lui. Les propriétés de cette source dépendent évidemment du processus de préparation mais aussi de l'appareillage de mesure accessible à l'observateur (et donc de ses limitations). L'objectif de la théorie de l'information est de quantifier les ressources nécessaires au stockage et au traitement de ces données.

À l'origine, la théorie de l'information a été formalisée en 1949 par Shannon et Weaver [330] qui travaillaient sur les télécommunications à la sortie de la seconde guerre mondiale. À cette époque, le développement massif des télécommunications rendait la question des ressources nécessaires pour communiquer particulièrement importante. Ils cherchaient donc à quantifier les performances des systèmes de communications, de codage et de décodage pour ensuite les optimiser. Les questions qui se posaient étaient très concrètes : quel est le taux de compression maximal que l'on peut appliquer à une source de données sans qu'il y ait de pertes ? Quelle est la capacité de transmission d'un canal bruité ?

Le point de départ de la théorie de Shannon consiste à remarquer que la notion d'information d'une source de données peut être définie de manière indépendante du contenu en mesurant l'aléatoire de la source de données. En d'autres termes : une source qui émettrait toujours la même donnée n'apporte rien de nouveau, et on lui associe donc une quantité d'information nulle. Toute l'approche de Shannon consiste donc à quantifier la quantité d'information associée à une source de données classiques aléatoires. Ces données sont portées par des objets classiques comme des séquences binaires, des mots composés de lettres d'un alphabet, etc. L'information au sens de Shannon quantifie notre ignorance relative à un ensemble de « mots classiques » que l'on peut appeler des « états classiques ».

Ces deux questions sont au cœur de la théorie de l'information quantique qui s'est considérablement développée ces 25 dernières années. Ce ne sont pas des questions vides de sens car travailler avec des objets quantiques introduit une nouveauté fondamentale par rapport au contexte classique. En effet,

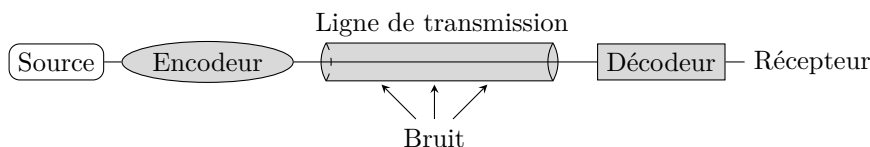


FIGURE 2.1 — Un système de communication est constitué de différents étages modélisés par la théorie des probabilités. Le premier, la source, émet des symboles qui sont manipulés par différents éléments de communication (en gris). Dans l'étage d'encodage, on les transforme pour pouvoir ensuite les transmettre. Cette transmission se fait à travers un canal soumis dans le cas général à un environnement venant bruite le signal, ce qui va entraîner des erreurs. Enfin, ce message est décodé puis réceptionné par l'utilisateur.

toute observation effectuée sur un système quantique le perturbe. De même, il est impossible d'accéder à l'état d'un système quantique en effectuant une seule mesure sur lui. Si l'état quantique d'un système véhicule bien toute la connaissance de toutes les prédictions que nous pourrions faire sur lui, cette connaissance n'est pas aussi accessible que son analogue classique. Enfin, nous savons qu'il est possible de dupliquer une information classique. Nous verrons néanmoins dans le chapitre 6 qu'il n'est pas possible de fabriquer un « photocopieur quantique » qui permette de dupliquer un état quantique quelconque. Cela montre la différence fondamentale entre les ressources classiques pouvant véhiculer de l'information classique et les ressources quantiques.

Il apparaît donc clairement que la théorie de l'information n'est pas simplement un corpus d'outils mathématiques mais est bien une façon originale de poser des questions concrètes sur le monde. Dans ce chapitre, nous allons présenter les bases de la théorie classique de l'information, avec une attention toute particulière donnée à la notion d'entropie de Shannon mais aussi à la notion de probabilité. Cet angle de vue va nous apporter un éclairage nouveau sur de nombreuses notions de physique tout au long de cet ouvrage.

2.1 Les systèmes de communication

L'objectif de la *théorie de l'information* est de quantifier la notion d'information associée à un système de communication. Dans notre vie quotidienne, nous sommes habitués à parler d'information de manière subjective, c'est-à-dire en faisant référence au fond et à la forme de ce que nous recevons. Mais pour évaluer les performances d'un système de communication, il faut s'abstraire du sens mais aussi des détails physiques des messages transmis pour définir une notion d'information qui soit objective. Pour cela, Shannon a adopté une approche statistique et probabiliste. Un réseau de communication est découpé

en différents étages, représentés sur la figure 2.1, que nous allons analyser successivement. Nous avons [330] :

- des *sources d'information* (Ω, p_Ω) qui émettent des symboles x . Une source d'information est caractérisée par un ensemble Ω , souvent appelé alphabet, de symboles qu'elle peut émettre et par une distribution de probabilité p_Ω sur cet ensemble ;
- des éléments de communication : ce sont les *lignes de transmission* bruitées ou non, *encodeurs* et *décodeurs*, *mélangeurs de signaux*, etc., qui permettent à partir d'un ensemble de sources en entrée $(\Omega_e, p_e)_e$ de fabriquer un ensemble de sources en sortie $(\Omega_s, p_s)_s$.

C'est la statistique des symboles qui caractérise la source. Il n'y a aucune référence à leur contenu ni même à leur forme spécifique (cela peut être une lettre de l'alphabet, un nombre, une couleur, une proposition mathématique complexe ou un livre). On dira d'une source qu'elle est sans mémoire si les symboles sont émis de façon indépendante. L'ensemble Ω peut être discret, infini dénombrable, ou même continu.

Dans la pratique, la distribution de probabilité des symboles n'est pas nécessairement bien connue. Caractériser la source nécessite donc d'échantillonner cette distribution par des mesures statistiques. Cela se fait à l'aide d'une suite de symboles appelée *message*. Lorsque le message est suffisamment long, la fréquence d'apparition des symboles nous permet d'accéder à leurs probabilités d'émission avec une erreur contrôlée.

Connaissant notre source (Ω, p_Ω) , le premier résultat important de Shannon permet de connaître les ressources minimales nécessaires pour bien représenter la source. C'est le problème de la compression d'une source auquel la réponse quantitative est fournie par la notion d'entropie S de Shannon.

La description des canaux d'un réseau de communication ne fera référence qu'à leurs fonctionnalités et non à leur mise en œuvre physique. Les propriétés statistiques du canal caractérisent les erreurs éventuelles introduites par la présence d'un environnement et du bruit qu'il engendre. De même que pour les sources, nous pouvons avoir des canaux discrets ou continus.

Le second résultat de Shannon sur les canaux de communication nous fournit une borne sur la quantité d'information pouvant être transmise sans perte. La quantité optimale est appelée capacité C du canal de transmission. Le résultat remarquable obtenu par Shannon [330] est que si l'on transmet un taux d'information plus petit que C , alors il existe une procédure de codage/décodage avec un taux d'erreur arbitrairement faible alors que ce n'est pas possible pour un taux de transmission plus grand que la capacité. Dans ce cas, il subsiste un taux d'erreur résiduel fini. Une conséquence surprenante de ce résultat est que, même en présence d'erreurs assez importantes, il est possible de transmettre de l'information sans erreur mais au prix d'un débit faible. C'est ce point précis qui permet le développement massif de systèmes de communications y compris dans des situations fortement génératrices d'erreurs, qu'il s'agisse d'environnements dégradés ou de très longues distances.

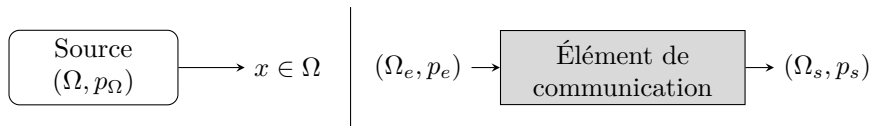


FIGURE 2.2 – Les deux parties principales de la théorie de l’information : l’étude des sources de symboles modélisées par un ensemble Ω et une distribution de probabilité sur cet ensemble, et les éléments de communication qui transforment et transportent l’information modélisés aussi grâce à la théorie des probabilités.

Avant d’entrer dans les détails de la théorie, discutons qualitativement la notion d’information. Comme résumé par la figure 2.2, les différents objets d’étude sont modélisés à l’aide de la théorie des probabilités permettant de quantifier notre méconnaissance des systèmes. Intuitivement, on comprend que l’on acquiert de l’information lorsqu’on lève une partie de l’incertitude sur l’issue des résultats expérimentaux. L’unité de mesure de l’information est le bit. Un bit est la quantité d’information fournie par le choix entre deux alternatives équiprobables. Par exemple, l’annonce « face » lors d’un tir de pile ou face bien équilibré correspond à la transmission d’un bit d’information (avant la mesure, le tir est représenté par un élément binaire $0 = \text{pile}$ et $1 = \text{face}$). Nous allons voir maintenant comment quantifier tout cela précisément.

2.2 L’entropie de Shannon

2.2.1 La mesure de l’information

Le point crucial consiste à trouver une bonne quantité mesurant l’information issue d’une source. Pour cela, partons d’un problème bien concret qui est celui de la compression d’une source de messages. Imaginons que l’on considère une source (Ω, p_Ω) et que l’on cherche à compresser ce qu’elle émet. Cela revient à se demander comment représenter de manière minimale un ensemble de messages émis par cette source.

L’exemple le plus simple est fourni par un jeu de pile ou face. On utilise comme générateur de messages une pièce ayant la probabilité p de tomber sur pile et la probabilité $1 - p$ de tomber sur face. Notre source est caractérisée par deux états $\Omega = \{\text{pile}, \text{face}\}$ et par la distribution de probabilité $(p, 1 - p)$. Pour générer un message m de taille N , on effectue N lancers avec cette pièce.

Un ensemble de N lancers est décrit par une succession aléatoire de N chiffres binaires 0 (pile) et 1 (face). Un message est donc un mot de N chiffres binaires et tous ces messages forment un ensemble Ω^N de cardinal $|\Omega^N| = 2^N$. Lorsque $p = 1/2$, tous les messages m sont équiprobables avec la probabilité $p_m = 1/|\Omega^N| = 1/2^N$. Dans le langage de la théorie de l’information, nous disons que nous avons besoin de N bits d’information pour pouvoir

encoder nos messages. Un bit est la quantité minimale d'information nécessaire pour encoder un symbole.

Lorsque la pièce est pipée ($p \neq 1/2$), tous les tirages ne sont plus équiprobables. Le cas extrême est celui où la pièce tombe toujours sur la même face. Un ensemble de N lancers donnera alors toujours le même mot binaire, constitué uniquement de 0 ou de 1 selon la face en question : il n'y a qu'un résultat possible. Dans la situation intermédiaire $0 < p < 1$, on s'attend à ce que dans la limite $N \gg 1$, un ensemble de N lancers comporte Np fois le résultat pile et $(1-p)N$ face. Les différents tirages typiques diffèrent uniquement par l'ordre des sorties pile ou face. Cela montre qu'il y a $\mathcal{N}_N \approx \binom{N}{Np}$ tirages typiques¹. À la limite $N \gg 1$, nous pouvons utiliser la formule de Stirling pour évaluer ce nombre, ce qui nous donne² :

$$\log \mathcal{N}_N \simeq -N(p \ln p + (1-p) \ln(1-p)). \quad (2.1)$$

On peut réécrire ce nombre comme³

$$\mathcal{N}_N \asymp 2^{-N(p \log_2 p + (1-p) \log_2(1-p))}, \quad (2.2)$$

où le symbole $\asymp$ signifie que les logarithmes des deux membres situés de part et d'autre sont asymptotiquement équivalents à la limite $N \gg 1$. En faisant l'analogie avec le lancer non pipé où la distribution de probabilité sur les messages était uniforme, l'ensemble des tirages typiques à N lancers pourra donc être indexé par des messages comportant $N(-p \log_2 p - (1-p) \log_2(1-p))$ chiffres binaires. Cela veut dire que pour chaque lancer typique, on doit mobiliser $-p \log_2 p - (1-p) \log_2(1-p) < 1$ bit d'information et non plus un bit comme pour le cas uniforme.

Bien sûr, cette discussion n'est pas très rigoureuse. L'évaluation obtenue ici se base sur une approximation que l'on peut qualifier de thermodynamique : nous n'avons considéré que les séquences typiques de résultats comportant Np fois le résultat pile et $(1-p)N$ fois le résultat face. On s'attend intuitivement à ce que les tirages non typiques soient extrêmement rares et qu'on puisse, en première approximation, les négliger. Mais il faudrait s'assurer qu'à la limite $N \rightarrow \infty$, la contribution des tirages « atypiques » est bien de plus en plus faible. En admettant que c'est le cas, essayons de comprendre les implications de cette analyse.

Tout d'abord, on voit que les ressources nécessaires sont proportionnelles à la taille N des messages émis par la source. Ceci permet de définir un nombre de bits par symbole nécessaires pour caractériser la source. Dans le cas présent,

1. Ici $\binom{N}{K}$ désigne le coefficient binomial $N!/K!(N-K)!$ qui décrit le nombre d'ensembles à $0 \leq K \leq N$ éléments parmi N . L'écriture $\binom{N}{Np}$ suppose donc implicitement que Np est entier, ce qui n'est évidemment pas le cas en général.

2. Comme on s'intéresse à la dépendance dominante à $N \gg 1$ et $p \neq 0$, on oublie les préfacteurs constants dans la formule de Stirling.

3. Ce calcul est un calcul dit de grandes déviations [359].

ce nombre est égal à

$$H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p) \quad (2.3)$$

et est une moyenne sur l'ensemble des symboles (ici deux) d'une fonction de la probabilité p de chaque symbole

$$I(p) = -\log_2 p. \quad (2.4)$$

C'est cette fonction qui caractérise le contenu informationnel d'un symbole. Elle vérifie les propriétés suivantes [330] :

- Elle ne dépend que de la probabilité p_x du symbole x considéré $I(p_x)$.
- Plus un symbole est improbable, plus il contient d'information. Dit autrement, les évidences n'apportent rien. La fonction $I(p) = -\log_2 p$ est bien une fonction décroissante de p .
- Enfin, l'information obtenue en groupant deux symboles indépendants est la somme des informations de chacun des symboles :

$$I(p_1 p_2) = I(p_1) + I(p_2). \quad (2.5)$$

Réciproquement, on peut montrer que si ces propriétés sont vérifiées par une fonction I donnée, alors :

$$I(p) = -\mathcal{K} \ln p, \quad (2.6)$$

où $\mathcal{K}$ est une constante qui, pour Shannon, est sans dimension. Elle spécifie juste la base dans laquelle on prend le logarithme et, parce que nous avons pris l'habitude de travailler en binaire, on utilise habituellement la base 2 ($\mathcal{K} = 1/\ln 2$). Dans la suite, on utilisera le $\log_2$ pour interpréter les formules en termes de bits d'information.

2.2.2 L'entropie de Shannon

Nous pouvons maintenant donner une caractéristique de la source dans son ensemble. L'*entropie de Shannon* d'une source $A = (\Omega, p)$ est simplement la moyenne sur tous les symboles Ω de l'information associée à chaque symbole [330] :

$$S[A] = - \sum_{x \in \Omega} p_x \log_2 p_x. \quad (2.7)$$

Cette notion d'information mesure en fait le déficit de connaissance sur la source étudiée. Ainsi, une source qui émet un symbole unique de manière certaine a une entropie nulle. Prenons par exemple une chaîne de TV X et regardons les mots de la langue diffusés avec leur probabilité. Cela donne un ensemble Ω (le dictionnaire) et p (une mesure de probabilité sur le dictionnaire, fonction de la chaîne). On peut alors calculer $S[X]$. Plus ce nombre sera grand, plus cela voudra dire qu'il y a une richesse sémantique dans la chaîne TV⁴,

4. Ou encore que les hybrides de la série Battlestar Galactica ont un meilleur contenu en information que les autres personnages.

ce qui semble surprenant au premier abord. Il est intéressant de traduire cela en termes de déficit de connaissance : notre capacité à prédire correctement le prochain mot diffusé va être plus faible quand l'information de la source augmente.

Cherchons, parmi les sources dont les symboles sont dans l'ensemble Ω , lesquelles contiennent le plus d'information. Nous devons donc maximiser la fonction (2.7) sous la contrainte $\sum_{x \in \Omega} p_x = 1$. Le théorème des extrema liés nous dit qu'à un point extrémal, il existe un multiplicateur de Lagrange μ tel que $\partial_{p_x} S = \mu$. Ceci montre que tous les p_x doivent être égaux. Dans le cas où l'ensemble Ω contient $|\Omega|$ symboles, et compte tenu de la contrainte, l'extremum est atteint pour $p_x = 1/|\Omega|$. L'entropie est alors maximale et vaut

$$S_{\max} = \log_2 |\Omega|. \quad (2.8)$$

Comme pour le jeu de pile ou face, l'entropie, et donc notre ignorance, est maximale lorsque l'incertitude est maximale : c'est le cas où tous les symboles sont équiprobables. Notons que cette information correspond au nombre de chiffres binaires qu'il faut pour décrire tous les symboles émis par la source. Dans l'exemple du pile ou face, on avait $|\Omega| = 2$ résultats équiprobables possibles, et la fonction S nous indique ici que, si l'on encode l'information en binaire, on aura besoin de $\log_2 2 = 1$ bit pour bien décrire l'ensemble des symboles et donc, par conséquent de N bits pour écrire un message binaire de taille N . On retombe bien sur le résultat énoncé précédemment.

Ainsi apparaît la connexion possible avec la physique statistique classique : l'état d'un système physique peut être vu comme un message qui, dans un grand nombre de cas, n'est pas connu avec certitude. Pour décrire les résultats d'un grand nombre d'expériences effectuées sur le système étudié⁵, la physique statistique utilise alors la notion d'ensemble. Un ensemble statistique est décrit par une distribution de probabilités sur l'ensemble des états possibles du système physique considéré, qui peut donc être vue comme une source de messages. L'information au sens de Shannon est alors proportionnelle à l'entropie du système définie par Boltzmann. Cependant, l'entropie n'est pas adimensionnée : elle est en joules par kelvin. Cela provient du fait que nous comptons les températures en kelvin et non dans une unité d'énergie. En physique statistique, l'entropie mesure la méconnaissance que nous avons de l'état du système, c'est-à-dire notre ignorance sur cet état. C'est précisément ce que Shannon appelle l'entropie de la source de messages que constitue l'ensemble statistique utilisé pour décrire l'état du système.

2.2.3 Premier théorème de Shannon sur le codage

Pour donner un angle de vue général sur la notion d'entropie, le problème que l'on va étudier ici est celui de la compression maximale d'une source. C'est

5. Il peut s'agir d'expériences effectuées sur un grand nombre de réalisations expérimentales distinctes ou bien, moyennant l'hypothèse ergodique, d'un grand nombre d'expériences effectuées les unes après les autres sur une même réalisation physique.

précisément ce que l'on a analysé sur l'exemple du jeu de pile ou face. On va désormais reprendre l'analyse dans le cas général, ce qui donnera le *premier théorème de Shannon sur le codage*.

Donnons au préalable quelques définitions des objets que nous utilisons. Nous avons vu par définition qu'une source A est définie par (Ω, p_Ω) où Ω est l'ensemble des symboles (alphabet) pouvant être émis par la source avec la distribution de probabilité p_Ω . La notion de suite de symboles, autrement appelée mot ou message, est primordiale. Les mots de taille N forment un ensemble noté Ω^N qui, si la source est sans mémoire, peut être muni d'une distribution de probabilité p_{Ω^N} à partir de p_Ω . L'ensemble des messages de taille N peut donc être vu lui-même comme une source d'information. Enfin, nous serons amenés à considérer l'ensemble des messages de taille finie qui sera noté $\Omega^* = \bigcup_N \Omega^N$.

Échantillonnage : raisonnement thermodynamique

Avant de discuter à proprement parler du théorème de Shannon sur le codage, il est intéressant de s'arrêter sur la question pratique de comment caractériser la distribution de probabilité d'une source. Ceci se fait grâce à un échantillonnage et à une analyse statistique. Ce problème est important puisqu'il fait écho à des questions d'interprétation des probabilités à travers l'approche fréquentiste mais aussi à des questions de physique statistique, d'économie, de sondage, etc.

On considère donc une source sans mémoire. L'hypothèse sans mémoire, c'est-à-dire de l'émission indépendante des symboles, est ici cruciale. Nous cherchons à voir comment représenter en utilisant le moins de ressources possibles les émissions de $N \gg 1$ symboles issus de cette source. Une telle émission peut être représentée par la donnée du nombre de fois n_x qu'apparaît chaque symbole $x \in \Omega$. L'entier n_x est appelé la multiplicité du symbole x . Les symboles étant émis indépendamment, la probabilité d'une émission où le symbole x apparaît n_x fois est $p_x^{n_x}$. Pour obtenir la probabilité que dans une émission de N symboles, on observe les multiplicités n_x , il faut multiplier par le facteur combinatoire qui donne le nombre de manières d'arranger les N symboles étant donné ces multiplicités. Cela nous donne :

$$p(n_x) = \frac{N!}{\prod_{x \in \Omega} n_x!} \prod_{x \in \Omega} p_x^{n_x}. \quad (2.9)$$

À la limite $N \rightarrow \infty$, on peut utiliser la formule de Stirling pour donner une forme approximative de cette distribution de probabilité en fonction des fréquences statistiques de chaque symbole : $f_x = n_x/N$. Cela nous donne :

$$p(n_x) \underset{N \rightarrow +\infty}{\sim} \exp \left(-N \sum_{x \in \Omega} f_x \ln (f_x/p_x) \right). \quad (2.10)$$

Le théorème des « extrema liés » nous dit comme précédemment que les fréquences statistiques qui maximisent la probabilité sont, sans surprise, données par $f_x^* = p_x$ pour tout $x \in \Omega$. Néanmoins, on peut aller plus loin dans ce raisonnement, car à partir de la formule (2.10) nous pouvons dire comment décroît la probabilité $p(n_x)$ lorsque les fréquences statistiques f_x s'écartent des valeurs optimales p_x . Pour cela, nous pouvons développer au second ordre en $f_x - p_x$ l'exposant du membre de droite de (2.10). Le terme du premier ordre est éliminé en utilisant la contrainte $\sum_x f_x = 1$ et nous obtenons finalement :

$$p(n_x) \asymp \exp \left(-\frac{N}{2} \sum_{x \in \Omega} \frac{(f_x - p_x)^2}{p_x} \right). \quad (2.11)$$

Nous voyons donc que la fluctuation des fréquences statistiques est en $N^{-1/2}$. En conséquence, lorsque N tend vers l'infini, la probabilité que les fréquences statistiques f_x s'écartent d'une quantité donnée des probabilités p_x tend exponentiellement vers zéro. Pratiquement toutes les séquences auront donc des fréquences statistiques égales à p_x .

Pour donner un exemple commun, la plupart des instituts de sondage utilise des échantillons de l'ordre de $N = 1000$ personnes. Dans l'hypothèse où l'échantillon est représentatif et que l'opinion est partagée, ceci nous donne des fluctuations statistiques en $1/\sqrt{1000} \simeq 0,03$, c'est-à-dire de l'ordre de 3 % sur les résultats du sondage. L'échantillonnage et l'analyse statistique sont en pratique plus subtils qu'il n'y paraît du fait du problème de la représentativité : il faut que l'échantillon de taille finie représente correctement la source que l'on souhaite analyser. Sinon la statistique n'a aucun sens. En physique statistique, ce problème est connu sous le nom de la typicalité. Nous n'entrerons pas dans les détails ici et on suppose que l'on sait toujours bien échantillonner les éléments probabilistes des modèles.

Pour avancer, procédons maintenant comme dans le cas du jeu de pile ou face : on fait l'hypothèse que les tirages de N symboles ont des fréquences statistiques d'apparition des symboles exactement données par les probabilités p_x . Le nombre de séries possibles de N messages est alors facile à calculer par la formule de Stirling :

$$\mathcal{N}_N(\Omega, p) = \frac{N!}{\prod_{x \in \Omega} (N p_x)!} \asymp \exp \left(-N \sum_{x \in \Omega} p_x \ln p_x \right). \quad (2.12)$$

Nous voyons donc apparaître l'entropie de Shannon (2.7) comme dans le cas du jeu de pile ou face. Cette formule montre clairement que l'interprétation que nous avons donnée à la section 2.2.1 est correcte. Puisque $\mathcal{N}_N \asymp 2^{NS(\Omega, p)}$, l'entropie de Shannon quantifie bien le nombre de chiffres binaires qui sont nécessaires pour numéroter les symboles issus d'une source.



FIGURE 2.3 – Une source de messages est connectée à un encodeur qui associe à chaque émission de N messages $(x_1, \dots, x_N)$ un codage $C(x_1, \dots, x_N)$.

Le codage d'une source

Ce résultat peut être rendu plus précis en posant la question en termes de codage. Imaginons que l'on connecte la source à un encodeur. À chaque émission $X = (x_1, \dots, x_N)$ de N symboles par la source, l'encodeur associe un codage $C(X)$ comportant N_C chiffres binaires, comme cela est représenté sur la figure 2.3. Un codage d'une source associe donc à chaque mot un mot binaire : c'est une application $\Omega^* \rightarrow \{0, 1\}^*$. Un codage est dit régulier s'il est injectif, c'est-à-dire que chaque mot de la source est bien représenté par un mot binaire sans ambiguïté. Un code quant à lui associe à chaque symbole de la source un mot binaire : c'est une application $\Omega \rightarrow \{0, 1\}^*$. Il est évident qu'à partir d'un code nous pouvons construire un codage. La réciproque n'est cependant pas vraie.

Une question naturelle que l'on peut se poser est celle de l'existence d'un codage optimal, avec ou sans perte. Dès lors que $N_C < N$, le codage, appelé dans ce cas compression, n'est pas rigoureusement parfait : comme dans un jeu de pile ou face, lorsque l'on cherche à étiqueter tous les tirages de N lancers par moins de N chiffres binaires, on doit associer le même mot codé à plusieurs tirages. Il ne peut pas être régulier. C'est ce que l'on appelle une erreur dans la compression. Mais la théorie de la compression ne s'arrête pas à cette limitation car elle s'intéresse à ce qui se passe sur un grand nombre de messages émis. On peut donc chercher à voir s'il est possible de rendre arbitrairement faible la probabilité d'erreur lorsque N tend vers l'infini.

Analysons d'abord la question de codage sans perte et des ressources minimales à mobiliser.

Regardons pour cela un grand nombre de tirages N . La probabilité d'une suite individuelle X de N symboles issus de la source est $p(X) = \prod_{x \in X} p_x$. On peut exprimer cette probabilité en fonction des fréquences statistiques d'apparition $f_x(X)$ de chacun des symboles $x \in \Omega$ de ce message spécifique X ,

$$\log_2 p(X) = N \sum_{x \in \Omega} f_x(X) \log_2 p_x. \quad (2.13)$$

Notons que lorsque les fréquences statistiques liées à cette séquence sont données par les probabilités, alors $-\log_2 p(X) = N S[A]$ où $S[A]$ est l'entropie de Shannon de la source $A = (\Omega, p)$. Du fait de la loi des grands nombres, on sait que l'on peut pour N assez grand se limiter aux séquences qui ont des fréquences arbitrairement proches des probabilités. De manière plus précise,

pour ϵ et η strictement positifs, il existe un N_0 tel que pour $N > N_0$ on ait :

$$N(S[A] - \eta) \leq -\log_2 p(X) \leq N(S[A] + \eta) \quad (2.14)$$

et que la probabilité totale des messages ne vérifiant pas la condition (2.14) soit inférieure à ϵ (loi faible des grands nombres). Les messages qui vérifient cette condition sont appelés les messages typiques (ϵ, η) . La condition (2.14) revient à donner des bornes inférieure et supérieure sur la probabilité des messages typiques :

$$2^{-NS[A]} 2^{-N\eta} \leq p(X) \leq 2^{-NS[A]} 2^{N\eta}. \quad (2.15)$$

La probabilité qu'un message quelconque soit typique (ϵ, η) est égale à la somme des probabilités des messages de la catégorie. C'est donc au moins le nombre total $N(\epsilon, \eta)$ de tels messages multiplié par un minorant de la probabilité de chaque message. Une borne inférieure de cette probabilité est donc $2^{-NS[A]} 2^{-N\eta} N(\epsilon, \eta)$. Mais nous savons que cela doit être supérieur à $1 - \epsilon$ et donc nous en déduisons un minorant de $N(\epsilon, \eta)$:

$$N(\epsilon, \eta) \geq (1 - \epsilon) 2^{NS[A]} 2^{-N\eta}. \quad (2.16)$$

Cette borne inférieure donne le nombre minimal de bits à utiliser pour bien étiqueter les messages typiques avec un taux d'erreurs arbitrairement faible : c'est $NS[A]$. De même, comme la somme des probabilités des messages typiques est inférieure à un, nous obtenons un majorant du nombre en utilisant la borne supérieure des $p(X)$:

$$N(\epsilon, \eta) \leq 2^{NS[A]} 2^{N\eta}. \quad (2.17)$$

Cette borne supérieure montre qu'en utilisant $N(S[A] + \eta)$ chiffres binaires, nous pouvons étiqueter tous les messages typiques. La probabilité de commettre une erreur, c'est-à-dire de donner une étiquette à un message non typique, est inférieure à ϵ .

Nous venons de démontrer le premier théorème de Shannon concernant le « codage sans bruit ». Il donne une borne sur les ressources à mobiliser pour comprimer les émissions venues de la source. Le théorème relatif au codage sans bruit nous dit que si on cherche à encoder des messages de taille N issus de la source A , nous avons besoin de mobiliser au moins $NS[A]$ chiffres binaires, ce qui représente un taux de $S[A]$ par symbole émis.

Remarquons que le théorème de Shannon ne donne pas de recette pour construire un schéma de compression optimal. Il donne juste une limite théorique à ce qu'il est possible de réaliser. Le problème de la construction d'un schéma de compression adapté est un vaste champ d'activité aux retombées concrètes évidentes.

Enfin, le nom de ce premier résultat important de Shannon peut paraître surprenant mais il fait référence au fait que le processus d'encodage n'introduit aucune erreur sur les messages originaux. Le mot bruit renvoie au fait que dans

une réalisation physique d'un système de traitement de l'information, c'est souvent le bruit électrique qui est à l'origine des erreurs. Le second problème discuté par Shannon est celui de la transmission à travers un canal bruité qui introduit des erreurs. Avant de discuter ce problème, nous allons plutôt chercher à nous familiariser avec quelques-unes des propriétés de l'entropie de Shannon.

Un exemple : le codage de Huffman

Nous allons ici présenter un exemple de schéma d'encodage adapté pour le cas d'une source très courante dans la vie de tous les jours, à savoir un texte composé de lettres. Dans les standards informatiques on associe un certain nombre de bits pour chaque symbole que l'on souhaite encoder. Ainsi, en ASCII, on va encoder chaque symbole sur 8 bits, quelle que soit la fréquence d'apparition de ce symbole. En UTF-8, qui est le standard le plus utilisé actuellement car permettant d'encoder un bien plus grand nombre de symboles que l'ASCII (alphabets latin avec signes diacritiques, grec et cyrillique, katakana japonais, etc.), on va utiliser selon les symboles des séries de 8, 16, 24 ou 32 bits respectant des règles bien particulières. Néanmoins, comme pour l'ASCII, ces règles sont totalement indépendantes du texte que l'on cherche à encoder.

Or, comme l'entropie de Shannon nous l'a fait comprendre, il est possible, lorsque la loi de probabilité d'apparition des messages est connue et différente d'une loi équiprobable, d'obtenir un codage nécessitant moins de ressources. Par exemple, en français, on s'attend à ce que si on encode le caractère « espace », le « e » ou le « s » avec moins de bits que les autres (quitte à devoir donner un nombre de bits élevé à des caractères comme le « k » ou le « w ») nous puissions arriver, pour un texte typique, à mobiliser moins de ressources qu'avec un nombre constant de bits par caractère. C'est l'idée derrière le *codage de Huffman*.

Pour construire ce codage, on pourrait partir de la loi de probabilité d'apparition des lettres en français, qui est assez bien connue. Pour simplifier, on va ici se donner comme fréquences statistiques d'apparition les fréquences d'apparition dans le texte que vous êtes en train de lire. L'idée du codage de Huffman va être de construire un arbre permettant d'associer à chaque caractère un code binaire qui ne soit le préfixe d'aucun autre. Ainsi, si le code 01 est attribué à la lettre « e », alors aucun caractère n'aura de code de la forme 01xxx. Pour cela, on va le construire de manière récursive en partant des plus petites fréquences d'apparition. En prenant les deux caractères qui apparaissent le moins, on indexe l'un par xxx0 et l'autre par xxx1, où xxx est une chaîne de caractères correspondant à un caractère fictif dont le poids est donné par la somme des poids des deux caractères de départ. Une telle construction est visible sur le schéma 2.4 pour un cas assez simple.

Quel est alors le nombre de bits mobilisés par caractère pour le message que l'on vient de coder par cette méthode ? Il est donné simplement par la somme sur tous les caractères des longueurs de chaîne associées à chaque caractère,

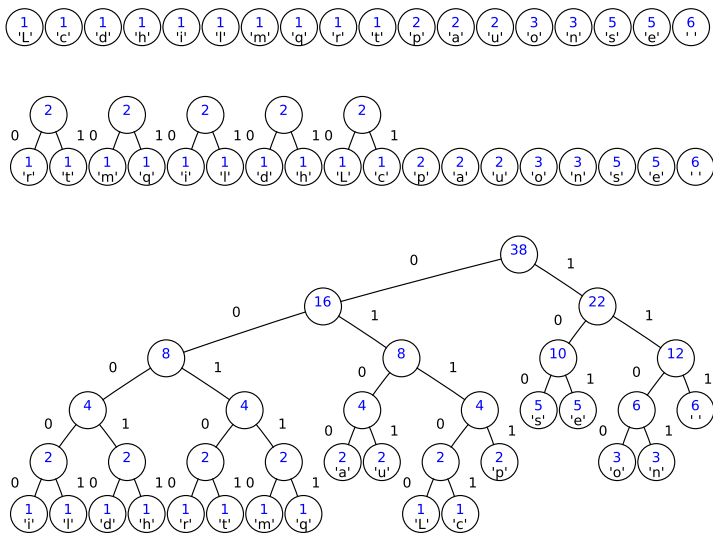


FIGURE 2.4 – Un exemple de codage de Huffman pour encoder la phrase « La phrase que nous encodons est simple ». Cette phrase contient 38 caractères au total, dont 18 différents. Les poids de chaque caractère sont donnés en haut de la figure. On construit ensuite l'arbre en associant à un couple de caractères dont les poids p_1 et p_2 sont faibles un caractère fictif dont le poids est $p_1 + p_2$, plus la donnée d'un nombre binaire 0 ou 1 renvoyant vers chacun des caractères initiaux. Puis on itère ce processus pour construire un arbre associant à chaque caractère utilisé dans la phrase un code binaire propre, qui n'est le préfixe d'aucun autre code (de sorte que si on obtient cette série de chiffres binaires dans un message codé, on sait avec certitude qu'il s'agit du caractère choisi, sans avoir besoin de regarder la suite du message). Ainsi, le code pour « e » est 101, celui pour « h » est 00011. On voit bien que l'on a suivi l'intuition naturelle qui consiste à associer les messages les plus courts aux caractères revenant le plus souvent. La visualisation utilisée ici est disponible de manière libre sur le site [166], et permet de visualiser le codage de Huffman d'un texte quelconque.

pondérées par la probabilité d'apparition de ce caractère. Pour la phrase très simple que l'on vient de coder sur la figure 2.4, il y a 18 caractères différents, 38 caractères au total, et les fréquences sont simplement les poids initiaux divisés par le nombre total de caractères ($f(e) = 5/38$ par exemple). L'entropie associée est alors $S = 3,82$ bit et le nombre moyen de bits par caractère donné par ce codage est 3,84 bit. On est extrêmement proche de la borne donnée par Shannon. On peut montrer que quel que soit le message regardé et tant qu'aucun caractère n'apparaît avec une probabilité supérieure à $1/2$, la méthode de Huffman va donner un nombre moyen de bits par caractère l tel que $S \leq l < S + 1$, l'égalité étant atteinte si toutes les fréquences d'apparition

sont de la forme $1/2^n$, $n > 1$. De plus, le codage de Huffman est optimal, dans le sens où c'est celui qui est le plus proche de la borne donnée par l'entropie de Shannon.

2.3 Propriétés de l'entropie de Shannon

L'information classique peut être manipulée de multiples manières. Nous allons ici analyser les relations vérifiées par l'entropie de Shannon dans diverses opérations : lors de l'extraction de l'information (analogue d'un processus de mesure), lors de la concaténation de messages issus de sources différentes où les corrélations jouent un rôle primordial et enfin lors d'un mélange aléatoire de sources.

2.3.1 Extraction d'information

Le premier processus que l'on va regarder est celui de l'extraction d'information qui consiste à classer les symboles émis par la source (Ω, p) en deux catégories A et B et à ne retenir que cette donnée. Nous noterons $A \cup B$ la source complète. Les physiciens désignent ce processus par la dénomination anglo-saxonne de *coarse graining*.

Cela revient à faire une partition de Ω en A et B disjoints. Notons p_A et p_B les probabilités d'avoir un symbole respectivement dans A et B , avec $p_A + p_B = 1$. Nous pouvons définir des distributions de probabilités associées à A et B , à savoir les probabilités conditionnelles définies par :

$$\begin{cases} p(k|A) = p(k)/p_A & k \in A, \\ p(k'|B) = p(k')/p_B & k' \in B. \end{cases} \quad (2.18)$$

Ces sources ont pour entropies associées $S(A)$ et $S(B)$. On note (p_A, p_B) la source à gros grains (*coarse grained*) qui donne l'information sur l'appartenance du symbole à A ou B . Son entropie est alors donnée par :

$$\begin{aligned} S[(p_A, p_B)] &= -p_A \log_2 p_A - p_B \log_2 p_B \\ &= -\left(\sum_{k \in A} p(k) \log_2 p_A + \sum_{k' \in B} p(k') \log_2 p_B \right) \end{aligned} \quad (2.19a)$$

$$\begin{aligned} &= -\left(\sum_{k \in A} p(k) [\log_2 p(k) - \log_2 p(k|A)] \right. \\ &\quad \left. + \sum_{k' \in B} p(k') [\log_2 p(k') - \log_2 p(k'|B)] \right) \\ &= S[A \cup B] - p_A S[A] - p_B S[B]. \end{aligned} \quad (2.19b)$$

Ce raisonnement se généralise au cas d'une partition quelconque de $\Omega = \cup_{k=1}^N \Omega_k$ en sous-ensembles disjoints. Notant Ω' la source à gros grains, l'information de la source Ω est donc obtenue en additionnant à l'information de Ω'

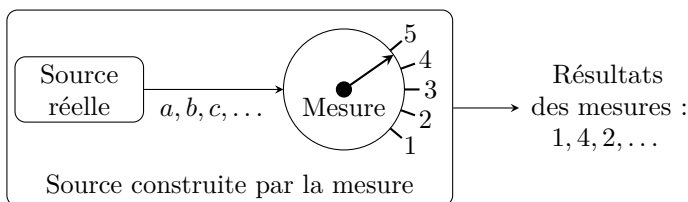


FIGURE 2.5 – Un processus de mesure revient à construire une nouvelle source d'information par extraction à partir de la source d'information que représente un système physique. Cette source émet comme messages les résultats des mesures. Plusieurs états physiques peuvent donner un même résultat de mesure.

la somme des informations des sources Ω_k pondérées par leurs poids respectifs p_k :

$$S[\Omega'] + \sum_{k=1}^N p_k S[\Omega_k] = S[\Omega]. \quad (2.20)$$

Ceci entraîne que

$$S[\Omega'] \leq S[\Omega]. \quad (2.21)$$

L'interprétation physique est simple : l'information d'une source obtenue par *coarse graining* est inférieure à celle de la source originelle. Un informaticien dirait qu'une source Ω' obtenue en appliquant une fonction à une source originelle Ω possède une information inférieure à $S[\Omega]$.

L'équation (2.20) montre que l'information gagnée correspond à l'information de la source définie par le processus d'extraction. Dans la théorie classique de la communication, il y a donc un bilan détaillé du transfert d'information lié à une extraction. Un physicien verra cela comme un processus de mesure classique (voir figure 2.5) : après une mesure, on obtient une nouvelle source de symboles qui renseigne uniquement sur l'information extraite de chaque symbole de la source originelle. Dans les travaux de Shannon [330], on part même de cette propriété pour établir la définition de l'entropie d'une source.

2.3.2 Cas d'une source composée

Maintenant que nous avons vu le comportement de l'entropie de Shannon lors d'un partitionnement (vu comme un processus d'extraction de l'information), regardons ce qui se passe lorsque l'on concatène différentes sources.

On regroupe deux sources A et B pour former une source jointe en concaténant les messages qu'elles émettent. On note $S[A, B]$ l'entropie de cette source composée. Le message typique sera alors un couple (i, j) avec $i \in A$ et $j \in B$. On note p_{ij} la probabilité sur $\Omega = A \times B$. Du fait des corrélations entre les sources, on a dans le cas général $p_{ij} \neq p_i p_j$.

- Sources indépendantes : supposons dans un premier temps les sources indépendantes. Explicitement $p_{ij} = p_i p_j$ et donc :

$$\begin{aligned} S[A, B] &= - \sum_{i,j} p_i p_j (\log_2 p_i + \log_2 p_j) \\ &= - \sum_j p_j \left(\sum_i p_i \log_2 p_i \right) - \sum_i p_i \left(\sum_j p_j \log_2 p_j \right), \end{aligned}$$

soit comme $\sum_i p_i = 1$:

$$S[A, B] = S[A] + S[B]. \quad (2.22)$$

- Sources corrélées à l'extrême : prenons le cas extrême où les deux sources sont totalement corrélées. Les messages de B se déduisent donc de ceux de A . On a alors

$$p(i, j) = p_i \cdot \delta_{j, f(i)} \quad (2.23)$$

et donc

$$\begin{aligned} S[A, B] &= - \sum_{(i,j) \in A \times B} p_i \cdot \delta_{j, f(i)} \log_2 p_i \cdot \delta_{j, f(i)} \\ &= - \sum_{i \in A} p_i \log_2 p_i. \end{aligned}$$

Finalement :

$$S[A, B] = S[A]. \quad (2.24)$$

Maintenant que nous avons une idée du comportement de l'entropie de Shannon pour quelques cas particuliers, revenons au cas plus général. On remarque que, en décomposant $A \times B = \cup_j (A \times \{j\})$, on obtient un minorant de $S[A, B]$. En effet, l'équation (2.21) nous dit que $S[A, B] \geq S[B]$. De même en décomposant $A \times B = \cup_i (\{i\} \times B)$, nous obtenons $S[A, B] \geq S[A]$. En conséquence :

$$\max(S[A], S[B]) \leq S[A, B]. \quad (2.25)$$

Comme on s'y attend intuitivement, la source obtenue en concaténant les messages des deux sources de départ contient plus d'information que chacune des deux sources de départ du fait des corrélations.

Après avoir obtenu un minorant de $S[A, B]$, cherchons en un majorant. Nous allons montrer que :

$$S[A, B] \leq S[A] + S[B]. \quad (2.26)$$

Ce résultat s'appuie sur un lemme simple : si (p_i) et (q_i) sont deux distributions de probabilité sur Ω , alors

$$- \sum_i p_i \log p_i \leq - \sum_i p_i \log q_i. \quad (2.27)$$

En effet, $\sum_i p_i \log q_i - \sum_i p_i \log p_i = \sum_i p_i \log (q_i/p_i)$. Or $\log (q_i/p_i) = \log \left(1 + \frac{q_i - p_i}{p_i}\right)$ et, par concavité du logarithme, $\log (1 + t) \leq t$, on en déduit la relation $\log (q_i/p_i) \leq \frac{q_i - p_i}{p_i}$. Par conséquent,

$$\sum_i p_i [\log q_i - \log p_i] \leq \sum_i (q_i - p_i) = 0, \quad (2.28)$$

d'où le résultat.

Armés de ce lemme, nous pouvons prouver l'équation (2.26). On dispose de la distribution de probabilité p_{ij} sur la source composée (A, B) . Les distributions de probabilité sur les sources A et B sont données par

$$\begin{cases} p_i = \sum_{j \in B} p_{i,j} & i \in A, \\ p_j = \sum_{i \in A} p_{i,j} & j \in B. \end{cases} \quad (2.29)$$

Comme $q_{i,j} = p_i p_j$ est une distribution de probabilité nous avons, avec l'équation (2.27), $-\sum_{i,j} p_{i,j} \log p_{i,j} \leq -\sum_{i,j} p_{i,j} \log q_{i,j}$. Soit

$$-\sum_{i,j} p_{i,j} \log p_{i,j} \leq -\sum_{i,j} p_{i,j} \log p_i - \sum_{i,j} p_{i,j} \log p_j. \quad (2.30)$$

Avec l'équation (2.29), on en déduit le résultat annoncé

$$S[A, B] \leq S[A] + S[B]. \quad (2.31)$$

L'information d'une source composée est donc inférieure à celle que l'on obtiendrait si les deux sources étaient indépendantes. La concavité du logarithme entraîne la propriété de sous-additivité de l'entropie. Le déficit d'information de la source composée provient des corrélations entre les deux sources. Nous verrons dans la section 2.4.3 comment quantifier plus précisément ces idées autour des corrélations.

2.3.3 Multiplexeur

Le processus précédant traitait les symboles en termes de couples et la distribution de probabilité de la nouvelle source était simplement la distribution de probabilité produit. Cependant, on peut très bien imaginer un processus combinant deux sources en les mélangeant. De façon imagée, on met deux collections de symboles d'un même ensemble Ω dans un même sac. Un tel élément s'appelle un multiplexeur. Il tire au hasard un symbole soit de la source A soit de la source B , la sortie agissant comme une nouvelle source de symboles pour l'utilisateur. Cette source multiplexée est modélisée par l'ensemble Ω et la distribution de probabilité $p_x = \lambda_A p_x^{(A)} + \lambda_B p_x^{(B)}$ avec $\lambda_A + \lambda_B = 1$. Ces coefficients caractérisent le mélange effectué par le multiplexeur.

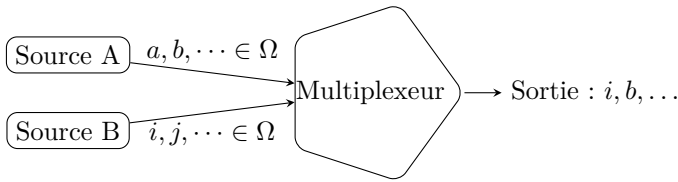


FIGURE 2.6 – Multiplexage de deux sources d'informations A et B piochant dans le même ensemble de messages : le multiplexeur choisit aléatoirement s'il diffuse un message issu de A ou issu de B .

Regardons comment un tel élément agit sur l'entropie de Shannon. L'entropie des deux sources est donnée par :

$$S[A] = - \sum_j p_j^{(A)} \log_2 p_j^{(A)} \text{ et } S[B] = - \sum_j p_j^{(B)} \log_2 p_j^{(B)}. \quad (2.32)$$

Dans la source multiplexée Ω , la probabilité du message j étant donnée par $p_j = \lambda_A p_j^{(A)} + \lambda_B p_j^{(B)}$ avec $\lambda_A + \lambda_B = 1$, on a

$$S[\Omega] = - \sum_{j \in \Omega} \left(\lambda_A p_j^{(A)} + \lambda_B p_j^{(B)} \right) \log_2 \left(\lambda_A p_j^{(A)} + \lambda_B p_j^{(B)} \right). \quad (2.33)$$

Mais $-x \log x$ est convexe, donc

$$S[\Omega] \geq - \sum_{j \in \Omega} \left\{ \lambda_A p_j^{(A)} \log_2 p_j^{(A)} + \lambda_B p_j^{(B)} \log_2 p_j^{(B)} \right\},$$

ce qui entraîne finalement

$$S[\Omega] \geq \lambda_A S[A] + \lambda_B S[B]. \quad (2.34)$$

L'entropie d'une source obtenue par mixage de deux sources est donc plus grande que celle des sources de départ.

Ce schéma peut être utilisé pour modéliser le bruitage d'une source d'information. Imaginons que A soit une source de 0 et de 1 d'entropie maximale (0 et 1 sont équiprobables). La source B par contre est totalement biaisée et n'émet que des 1. C'est notre source d'erreurs, car son action revient à effacer l'information véhiculée par A . Le multiplexeur laisse passer le bit émis par A avec probabilité $1 - p$ et celui émis par B avec la probabilité p . Nous pouvons calculer la perte d'information introduite par la source d'erreur très simplement : la probabilité d'obtenir 1 en sortie est $(1 - p)/2 + p = (1 + p)/2$. Donc en sortie nous avons un jeu de pile ou face biaisé dont l'entropie de Shannon est

$$S_{\text{bruit}} = H\left(\frac{p+1}{2}\right) = -\frac{1+p}{2} \log_2 \left(\frac{1+p}{2}\right) - \frac{1-p}{2} \log_2 \left(\frac{1-p}{2}\right). \quad (2.35)$$

La source d'erreurs que nous avons introduite agit sur une fraction p des bits émis par la source A . Naïvement, on pourrait donc penser que l'information de la source bruitée par le multiplexeur est $(1-p)S[A] = 1-p$. Mais en fait ce n'est pas le cas : $H((1+p)/2)$ est supérieur à $1-p$ pour $0 < p < 1$. Pour un faible taux d'erreurs $p \ll 1$: $H((1+p)/2) \simeq 1-p^2$. L'estimation naïve est incorrecte car elle ne prend pas en compte le fait que même si le bruit n'altère qu'une fraction p des bits émis par A , on ne sait pas lesquels sont altérés.

2.3.4 ☒ La divergence de Kullback-Leibler

Pour montrer que l'entropie d'une source composée est plus petite que la somme des entropies de chacune des sources qui la composent, nous avons montré une inégalité intermédiaire (2.28) que l'on peut interpréter en termes de la théorie de l'information. En effet, considérons une source de symboles A émettant le symbole i avec la probabilité p_i . Le théorème de Shannon nous dit qu'il faut $S[A]$ bits pour encoder les symboles typiques émis par cette source. Imaginons maintenant que nous ayons une estimation différente de la probabilité d'émission de chacun des symboles émis par A , à savoir q_i pour le symbole i . Dans ce cas, nous serions conduits à employer un code qui affecte $-\log_2 q_i$ bits pour le symbole i soit, en moyenne pour la source A un nombre de bits égal à $-\sum_i p_i \log_2 q_i$. L'inégalité (2.28) signifie donc que nous utiliserions plus de bits que ce qui est prescrit par le théorème de Shannon. La quantité

$$D[A\|A'] = \sum_i p_i \log_2(p_i/q_i) \geq 0, \quad (2.36)$$

où A' désigne un encodage dans lequel $-\log_2 q_i$ bits sont mobilisés pour le message i émis avec la probabilité p_i représente le nombre de bits en excès utilisés par l'encodage A' par rapport à l'encodage optimal nécessaire pour la source A .

Cette quantité est appelée la *divergence de Kullback-Leibler*. Sa propriété essentielle est que $D[A\|A'] = 0$ si et seulement si les distributions de probabilités $(p_i)_i$ et $(q_i)_i$ sont identiques. En effet, la concavité stricte du logarithme nous dit que

$$D[A\|A'] \geq \log_2 \left(\sum_i p_i \frac{q_i}{p_i} \right) = 0, \quad (2.37)$$

avec égalité si et seulement si chaque rapport q_i/p_i est constant. La condition de normalisation $\sum_i p_i = \sum_i q_i = 1$ implique alors que $p_i = q_i$ pour tout i . Il est également facile de vérifier que si A est obtenu en réunissant deux sources indépendantes A_1 et A_2 , alors

$$D[\{A_1, A_2\} \|\{A'_1, A'_2\}] = D[A_1\|A'_1] + D[A_2\|A'_2]. \quad (2.38)$$

La divergence de Kullback-Leibler est donc une quantité qui permet de comparer, dans le cadre de la théorie de l'information, deux distributions de probabilités sur un même ensemble d'événements.

2.3.5 Entropie et dynamique stochastique

Pour finir, nous allons considérer le comportement de l'entropie de Shannon dans un modèle simple de dynamique. Nous pouvons par exemple reprendre l'idée du chapitre 1 et considérer que les symboles sont des configurations parfaitement discernables qui sont transformées dans un processus d'évolution décrit par les probabilités de transition quantiques $p_{i \rightarrow f} = |\mathcal{A}_{i \rightarrow f}|^2$. Nous pouvons aussi réfléchir dans un cadre purement classique. Certes, une dynamique parfaitement connue envoie une configuration i sur une configuration $F(i)$ mais en général, la dynamique n'est pas parfaitement connue et nous devons, là encore, introduire des probabilités de transition $p_{i \rightarrow j}$ pour la décrire. La dynamique transforme la distribution de probabilité $(p_i)_i$ en

$$p'_j = \sum_i p_{i \rightarrow j} p_i, \quad (2.39)$$

où les probabilités de transition vérifient

$$\sum_j p_{i \rightarrow j} = 1 \quad (2.40)$$

pour chaque i , de manière à ce que $\sum_j p'_j = 1$. Le résultat central est que la divergence de Kullback-Leibler est contractée dans cette dynamique :

$$D[p' \| q'] \leq D[p \| q]. \quad (2.41)$$

Par rapport à une distribution de probabilité uniforme, nous remarquons que l'entropie de Shannon est reliée à la divergence de Kullback-Leibler par $S[p] = N - D[p \| 1/N]$ où N est le nombre total de configurations. De plus, si la distribution uniforme est stable sous la dynamique donnée par (2.39), alors l'entropie doit nécessairement croître $S[p'] \geq S[p]$. Or cette distribution est stable lorsque la matrice des probabilités de transition est bistochastique, c'est-à-dire qu'en plus de l'équation (2.40), sa transposée est également stochastique. Remarquons que par unitarité de la matrice d'évolution en théorie quantique, c'est le cas pour une dynamique obtenue en utilisant des probabilités de transition quantiques.

2.4 Bayes et la théorie de l'information

2.4.1 Une autre conception des probabilités

Habituellement, la théorie des probabilités est présentée sous l'angle fréquentiste, c'est-à-dire à partir d'un comptage ou d'une mesure sur un ensemble d'événements qui sont construits à partir d'un ensemble d'événements élémentaires. Une probabilité est tout simplement une manière d'associer un poids réel entre zéro et l'unité à chaque événement en vérifiant des axiomes

simples à savoir que la probabilité d'une union dénombrable d'événements disjoints est la somme des probabilités de chacun des événements et que la somme des probabilités de tous les événements élémentaires est l'unité. Dans cette approche, les probabilités apparaissent comme des quantités objectives, c'est-à-dire qui ne dépendent que du système considéré mais absolument pas d'un quelconque observateur.

Cependant, nous n'avons pas accès à une distribution de probabilité dans ce sens-là car il nous est impossible d'explorer l'ensemble des événements possibles. Un grand nombre d'événements restent contrefactuels, c'est-à-dire ne se produisent pas mais auraient pu se produire. La notion de probabilité du sens commun renvoie en fait à une appréciation de la vraisemblance d'événements futurs (probabilités causales des effets E de causes C) $p(E|C)$ ou bien passés (probabilités inductives) $p(C|E)$ qui permettent de remonter aux causes C à partir d'effets E observés. D'ores et déjà, nous avons introduit un ingrédient nouveau absent de l'approche traditionnelle qui est la causalité entre les événements.

Notons que les probabilités dont nous avons parlé dans le chapitre 1 sont des probabilités causales : elles décrivent les transitions entre une situation initiale et une situation finale et sont expérimentalement accessibles de manière asymptotique à travers un grand nombre d'expériences de laboratoire où la préparation de l'état initial est parfaitement contrôlée. On peut donc dire que la physique (quantique) nous fournit un outil pour calculer les probabilités causales.

Mais ce ne sont pas les seules avec lesquelles notre esprit jongle. Tout enquêteur traite avec un autre type de probabilités qui quantifient le degré de vraisemblance d'une cause sachant que l'on a observé un certain nombre d'effets. Ce sont précisément les probabilités inductives. Il est clair que l'on ne peut y accéder par un grand nombre d'expériences de laboratoire, puisqu'un seul événement s'est réellement produit. Ces probabilités sont par essence subjectives et représentent la confiance que l'on accorde à la reconstruction d'un historique passé.

Une personne ayant cherché à exprimer de manière quantitative le degré de confiance d'un historique passé reconstitué est le révérend britannique Thomas Bayes au XVIII^e siècle. Dans l'approche traditionnelle, le *théorème de Bayes* est souvent exprimé de la manière suivante :

$$p(A \cap B) = p(A|B) p(B) = p(B|A) p(A), \quad (2.42)$$

où $p(A \cap B)$ est la probabilité d'avoir A et B et où $p(A|B)$ est la probabilité conditionnelle d'avoir A sachant B . La relation $p(A|B) p(B) = p(A \cap B)$ peut être vue comme une manière de définir les probabilités conditionnelles. Le théorème est donc la seconde égalité. Dans l'approche fréquentiste, la première égalité apparaît comme une conséquence triviale de la définition

d'une probabilité comme un comptage d'occurrence d'événements :

$$p(A|B) = \frac{\#(A \text{ et } B)}{\#(B)} = \frac{\#(A \text{ et } B)}{\#(\Omega)} \frac{\#(\Omega)}{\#(B)} = \frac{p(A \cap B)}{p(B)}. \quad (2.43)$$

Mais dans cette approche, on ne tient pas compte du lien de causalité entre A et B . Or si A est causalement postérieur à B , B est forcément ⁶ causalement antérieur à A . Donc dans l'égalité

$$p(A|B)p(B) = p(B|A)p(A), \quad (2.44)$$

une des probabilités est causale alors que l'autre est inductive. Il est donc clair que le théorème de Bayes concerne des probabilités inductives. Plus précisément, si on considère que C est une cause et E un effet, alors la relation de Bayes nous permet d'obtenir une probabilité inductive $p(C|E)$ à partir de probabilités causales $p(E|C)$, de fréquences d'observation des effets $p(E)$ et d'une probabilité d'une cause $p(C)$ qui doit donc être comprise comme une probabilité *a priori* ou bien conditionnée à des observations antérieures.

2.4.2 Exemples d'utilisation de l'inférence bayésienne

Pour expliquer comment fonctionne l'*inférence bayésienne*, le mieux est encore de considérer quelques exemples en commençant par les plus simples, à savoir les tests biologiques. Nous présenterons ensuite un exemple historique de recherche bayésienne.

Les tests biologiques

Les tests biologiques fonctionnent tous sur le même principe : deux résultats possibles positif (noté P) et négatif (noté N). En principe, le résultat positif devrait révéler un état biologique noté $+$ et le résultat négatif un état biologique noté $-$. Sauf que le test peut donner un certain nombre de faux positifs avec une probabilité $p(P|-)$ et de faux négatifs $p(N|+)$. Ces probabilités peuvent en principe être obtenues par des études statistiques ou des expériences de laboratoire. Cependant, elles ne renseignent pas sur les conséquences en termes de santé publique.

Ce qui nous intéresse est la probabilité $p(-|P)$ qu'un individu dont le résultat du test est positif soit en fait $-$ ou encore que la probabilité $p(+|N)$ qu'un individu testé négatif soit en fait $+$. Il s'agit de probabilités inductives que nous allons calculer par une démarche bayésienne.

Pour cela, considérons une population dont on connaît par une étude statistique les probabilités fréquentistes d'être $+$ ou $-$. Le théorème de Bayes nous dit que les deux probabilités que nous cherchons peuvent être obtenues

6. En l'absence de machine à remonter le temps.

par :

$$p(+|N)p(N) = p(N|+)p(+), \quad (2.45a)$$

$$p(-|P)p(P) = p(P|-)p(-). \quad (2.45b)$$

Dans ces équations, nous connaissons les probabilités qui apparaissent dans les membres de droite. Il nous reste juste à calculer $p(N)$ et $p(P)$. On peut les obtenir en sommant des probabilités conditionnelles :

$$p(P) = p(P|+)p(+) + p(P|-)p(-), \quad (2.46a)$$

$$p(N) = p(N|+)p(+) + p(N|-)p(-). \quad (2.46b)$$

Nous obtenons ainsi les expressions finales des probabilités de mise en défaut du test :

$$p(+|N) = \frac{p(N|+)p(+)}{p(N|+)p(+) + p(N|-)p(-)}, \quad (2.47a)$$

$$p(-|P) = \frac{p(P|-)p(-)}{p(P|+)p(+) + p(P|-)p(-)}. \quad (2.47b)$$

Ce calcul illustre comment on peut évaluer la confiance que l'on peut accorder à un test biologique à partir de la démarche bayésienne. Notons que la relation de Bayes n'a pas suffi et qu'il nous a fallu calculer les probabilités totales des résultats du test $p(P)$ (resp. $p(N)$) en sommant les probabilités des événements disjoints $p(P, \pm)$ (resp. $p(N, \pm)$). Le point à retenir est que le résultat dépend de deux choses : d'une part des performances du test, contrôlées en laboratoire, qui sont décrites par les probabilités $p(N|\pm)$ et $p(P|\pm)$ et d'autre part, des probabilités fréquentistes des deux états biologiques + et -.

Mettons quelques chiffres dans cette discussion. On prend un test qui donne le bon résultat avec 95 % de chance, c'est-à-dire $p(N|-) = p(P|+) = 0,95$ et donc $p(N|+) = p(P|-) = 0,05$. Si on sait que les répartitions dans la population sont $p(+) = 0,1$ et $p(-) = 0,9$ on obtient par les équations (2.47a) une probabilité de faux négatif de 3,6 % et de faux positif de 3,2 % ($p(+|N) = 0,036$ et $p(-|P) = 0,032$).

À la recherche de l'*USS Scorpio*

L'*USS Scorpio* était un sous-marin nucléaire d'attaque américain disparu en mer en mai 1968. Une station d'écoute sonar des Açores enregistra un bruit qui fut identifié comme l'implosion de la coque lorsque le sous-marin atteignit sa profondeur limite. La Navy n'aimant pas perdre un de ses précieux navires, *a fortiori* lorsque celui-ci est à propulsion nucléaire et transporte des armes nucléaires, une opération de recherche fut lancée pour retrouver l'épave.

Mais détecter une épave posée sur les fonds marins par 3000 m de profondeur n'est pas chose aisée. On peut envoyer des sous-marins automatiques photographier le fond mais la détection optique ne marche pas à très grande

distance. On peut employer un sonar pour cartographier le fond mais là aussi, la portée est limitée. Enfin, on peut déplacer un magnétomètre au-dessus du fond afin de repérer une masse métallique. Quoi qu'il en soit la portée effective d'un système de détection ne dépassait pas, à l'époque, une centaine de mètres. La zone de naufrage étant de plusieurs dizaines de kilomètres carrés, on comprend qu'une recherche complète avait peu de chances d'aboutir en un temps raisonnable.

Les méthodes bayésiennes s'avèrent en fait très utiles dans ce cas. Pour les appliquer, l'idée consiste à partir d'un découpage de la zone de recherche en carrés auxquels on affecte une probabilité de trouver l'épave. On lance alors les recherches et on actualise la grille des probabilités en fonction des résultats des recherches. Même si on n'arrive à scanner qu'une petite partie de la zone totale, l'inférence bayésienne permet de modifier les probabilités de présence de l'épave dans les différents carrés et donc de focaliser les recherches. C'est ainsi que l'USS Scorpio fut retrouvé à l'automne 1968 (figure 2.7).

De manière concrète, supposons que l'on ait découpé la zone de recherches en cellules et que l'on ait affecté une probabilité de présence *a priori* de l'épave égale à $p(\alpha)$ à la cellule α . Pour une campagne de recherche, si l'épave est dans la cellule α , on a une efficacité de détection q_α . Si la campagne est fructueuse, la campagne s'arrête. Dans le cas où la campagne est infructueuse, alors on va mettre à jour les probabilités de présence de l'épave dans chaque cellule par la formule de Bayes

$$p(\alpha|\text{ND})p(\text{ND}) = (1 - q_\alpha) \times p(\alpha), \quad (2.48)$$

où la probabilité totale de non-détection n'est autre que

$$p(\text{ND}) = \sum_{\alpha'} p(\alpha') (1 - q_{\alpha'}). \quad (2.49)$$

Nous avons alors la table des probabilités mise à jour

$$p(\alpha|\text{ND}) = \frac{(1 - q_\alpha) \times p(\alpha)}{\sum_{\alpha'} (1 - q_{\alpha'}) p(\alpha')}. \quad (2.50)$$

Si on effectue les recherches dans une seule zone α , alors q_α est entre zéro et un et tous les q_β pour $\beta \neq \alpha$ sont nuls. Dans ce cas, il est facile de voir qu'en cas de non-détection, $p(\alpha|\text{ND}) < p(\alpha)$ et que en revanche $p(\beta \neq \alpha|\text{ND}) > p(\beta)$ comme le bon sens le laisse supposer.

Cette brève discussion ne donne qu'un très mince aperçu des problèmes pratiques posés par la mise en œuvre de ce type de méthode. Tout d'abord, il faut bien commencer par initialiser le processus par une première table de probabilités et cela ne peut être fait qu'en se basant sur les informations disponibles avant de commencer les recherches et donc nécessite un travail préliminaire. Ensuite, on doit estimer les efficacités de détection q_α , ce qui nécessite une certaine expertise. Enfin, un des problèmes majeurs dans ce

type de recherche est le coût en ressources qui est une fonction croissante de la durée des recherches. Comme la probabilité de succès d'une campagne de recherche est donnée par

$$p(D) = \sum_{\alpha} p(\alpha) q_{\alpha}, \quad (2.51)$$

on pourrait commencer par concentrer les efforts sur les cellules qui possèdent la probabilité de présence de l'épave la plus élevée. Sauf que si l'objet n'y est pas, on risque de rallonger les recherches ultérieures car il faudra ensuite investiguer plus de cellules. De plus, la recherche elle-même est contrainte de manière forte par exemple par la vitesse de déplacement du ou des véhicules automatiques portant les instruments de détection. La mise au point d'algorithmes d'optimisation du coût prenant en compte toutes les contraintes est un sujet de recherche à part entière [351] qui continue, encore maintenant, de faire l'objet de nouveaux développements [352].

2.4.3 Information mutuelle et inférence bayésienne

Information mutuelle de deux sources

On a déjà vu au cours de ce chapitre que si deux sources A et B sont corrélées entre elles, alors l'entropie de la source composée (A, B) vérifie les inégalités

$$\max(S[A], S[B]) \leq S[(A, B)] \leq S[A] + S[B]. \quad (2.52)$$

La borne inférieure est atteinte si les sources sont parfaitement corrélées. La borne supérieure lorsqu'elles sont parfaitement indépendantes. Une mesure du degré de corrélation entre ces deux sources est donnée par l'*information mutuelle* $I[A, B]$ définie par

$$I[A, B] = S[A] + S[B] - S[(A, B)]. \quad (2.53)$$

Les équations (2.25) et (2.31) entraînent que

$$0 \leq I[A, B] \leq \min(S[A], S[B]). \quad (2.54)$$

Pour des sources indépendantes, $I[A, B] = 0$, et pour des sources parfaitement corrélées, $I[A, B] = \min(S[A], S[B])$. Le schéma 2.8 permet de se faire une idée de ce que cette quantité représente. Si $S[A]$ représente l'information contenue dans la source A et $S[B]$ l'information contenue dans la source B , alors $I[A, B]$ représente l'information mutuelle contenue dans les deux sources A et B . À l'inverse, $S[(A, B)]$ représente la somme des informations exclusives à chacune des sources.

La corrélation entre des sources de messages, c'est-à-dire le fait que dans le cas général $p(i, j) \neq p(i)p(j)$ avec i un symbole de A et j un message de B , implique l'existence de cette information mutuelle. Cependant, on a déjà vu

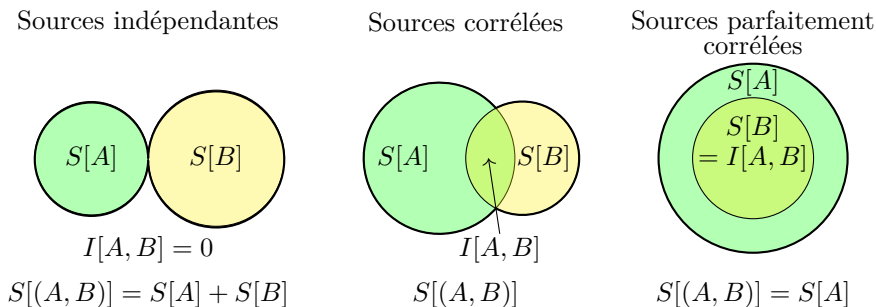


FIGURE 2.8 – Les différentes quantités informationnelles. $S[(A, B)]$ est égale à la somme des informations exclusives de chacune des sources, et correspond à la zone délimitée par le trait épais. Elle est égale à la somme des informations dans le cas de sources indépendantes, et au maximum des informations des deux sources pour des sources parfaitement corrélées (on a choisi ici que A était la source avec la plus grande information). $I[A, B]$ est l'information mutuelle contenue dans les deux sources, et correspond à la zone de recouvrement des informations de chaque source. On a $S[(A, B)] = S[A] + S[B] - I[A, B]$.

- *a contrario*, lorsque les sources A et B sont indépendantes, on a l'égalité $p(j|i) = p_B(j)$ pour tout message j de B et pour tout message i de A . Dans ce cas, l'entropie $S[B|i]$ est égale à $S[B]$ et représente le bruit au niveau de B .

L'entropie conditionnelle de B connaissant la source A est obtenue en pondérant les entropies conditionnées à un message émis par A sur la distribution de ceux-ci :

$$S[B|A] = - \sum_{i \in A} p(i) S[B|i]. \quad (2.56)$$

En utilisant la définition de la probabilité conditionnelle $p(j|i) = p(j, i)/p(i)$, on obtient alors :

$$S[B|A] = S[(A, B)] - S[A]. \quad (2.57)$$

L'entropie conditionnelle (ou l'équivoque [330] sur B connaissant la source A) mesure donc l'ajout d'information permis par B lorsque l'on connaît A . Si les deux sources sont indépendantes, alors $S[B|A] = S[B]$. Notons que les corrélations entre A et B ont pour effet de diminuer l'entropie conditionnelle $S[B|A]$ par rapport à l'entropie non conditionnée $S[B]$. On remarque alors que

$$S[A] - S[A|B] = S[A] + S[B] - S[(A, B)] = S[B] - S[B|A]. \quad (2.58)$$

On retrouve ici l'information mutuelle $I[A, B] = S[A] + S[B] - S[(A, B)]$ des deux sources de messages. On peut donc réinterpréter cette information mutuelle comme étant $S[A] - S[A|B]$ qui représente ce que l'on peut apprendre

de A en observant B . L'information mutuelle n'est rien d'autre que la divergence de Kullback-Leibler entre la distribution $p_{A,B}$ complète et le produit des marginales $p_A \times p_B$:

$$I[A, B] = \sum_{i,j} p(i, j) \log_2 \frac{p(i) p(j)}{p(i, j)} = D[p_{A,B} \| p_A \times p_B]. \quad (2.59)$$

Le gain d'information de l'inférence bayésienne

Maintenant que nous avons bien défini les notions d'information conditionnelle et d'information mutuelle, examinons le gain moyen d'information dans le processus d'inférence bayésienne. Pour cette étude, les sources que nous allons étudier sont donc une source de causes C et une source d'effets E . Bien entendu, ces sources sont corrélées, et les effets sont causalement postérieurs aux causes.

L'information initiale est celle associée à la distribution de probabilité des causes *a priori* : notons-la $S[C]$. L'information moyenne sur les causes après le processus d'inférence bayésienne est la moyenne sur les différents effets possibles de l'information associée à la probabilité *a posteriori* qui est conditionnée aux effets $p(\cdot|E)$. C'est donc l'entropie conditionnelle

$$S[C|E] = \sum_E p(E) S[p_f(\cdot|E)]. \quad (2.60)$$

Le gain d'information lié à l'inférence bayésienne est alors la différence entre l'information *a priori* et l'information *a posteriori*. Si les observations apportent quelque chose, on s'attend à ce que l'information après inférence bayésienne soit plus petite que l'information *a priori*. Le gain d'information tel que nous l'avons défini est donné par

$$\Delta I = S[C] - S[C|E] = I[E, C] \geq 0. \quad (2.61)$$

Il s'agit donc tout simplement de l'information mutuelle des deux sources. Cette expression peut se réécrire sous la forme

$$\Delta I = S[E] - S[E|C], \quad (2.62)$$

qui s'interprète en disant que le gain d'information n'est autre que l'information associée aux observations à laquelle est soustraite l'information $S[E|C]$ qui représente le bruit sur les observations (c'est-à-dire l'aléatoire des effets à cause parfaitement déterminée).

On a déjà vu que l'information mutuelle vérifiait les inégalités suivantes

$$0 \leq \Delta I \leq \min(S[C], S[E]). \quad (2.63)$$

En d'autres termes, le processus d'inférence bayésienne nous permet toujours de gagner de l'information. Cependant ce gain d'information peut être nul. C'est le cas si et seulement si l'information mutuelle est nulle ($I[C, E] = 0$), ce qui veut dire que les causes et les effets sont statistiquement indépendants. En clair quand on prend comme effets des choses qui n'ont rien à voir avec la cause, l'induction ne nous apprend rien.

Le cas opposé est celui où les effets sont totalement déterminés par les causes. Dans ce cas, $S[E|C] = 0$ et on a donc $\Delta I = S[E]$. On gagne alors toute l'information qui est contenue dans les observations : c'est le cas où le rendement du processus d'inférence bayésienne est le plus élevé. Notons que la borne (2.25) nous dit que dans ce cas $S[E] \leq S[C]$. Nous avons alors bien $I[E, C] = \min(S[C], S[E]) = S[E]$. Le cas où $S[C] = S[E]$ correspond au cas où les observations permettent de remonter de manière non ambiguë aux causes, chaque effet n'ayant pu être causé que par une seule cause.

Remarquons cependant que cette discussion porte sur le gain d'information moyen lors du processus d'inférence bayésienne. Cependant, lorsque l'on met en œuvre l'inférence bayésienne sur un cas concret, on part d'une distribution *a priori* des causes $p_i(C)$ pour arriver à une distribution *a posteriori* qui dépend des observations effectuées *via* la formule de Bayes. On obtient donc une distribution de probabilité *a posteriori* $p(C|E)$ où E est fixé. Le gain d'information associée à cette instance spécifique du processus d'inférence bayésienne est alors

$$\Delta I[E] = \sum_C p_f(C|E) \log_2 \left(\frac{p(C|E)}{p_i(C)} \right), \quad (2.64)$$

qui n'est autre que la divergence de Kullback-Leibler $D[p(\cdot|E)||p_i]$ entre la distribution de probabilité *a priori* et celle *a posteriori*. Moyenner sur un grand nombre d'instances redonne bien la quantité ΔI calculée précédemment. Mais nous voyons immédiatement la difficulté : si on considère une séquence d'observations successives qui aboutissent à la mise en évidence d'effets E_1, E_2 , etc., le gain d'information peut très bien diminuer en cours de route car nous n'avons aucun résultat qui nous dise que $D[p(\cdot|E_1, E_2)||p_i] \geq D[p(\cdot|E_1)||p_i]$. En particulier, le gain d'information au bout de la seconde étape n'est relié ni par une égalité, ni par une inégalité à la somme du gain d'information de la première étape et du gain d'information entre la première et la seconde $D[p(\cdot|E_1, E_2)||p(\cdot|E_1)]$. Néanmoins, lorsque l'on considère les gains d'information moyenne qui sont donnés par une information mutuelle, nous avons bien

$$I[C, \{E_1, E_2\}] = I[C, E_1] + I[C, E_2|E_1], \quad (2.65)$$

où la quantité $I[C, E_2|E_1]$ est définie comme la moyenne sur E_1 des informations mutuelles entre E_2 et C conditionnées à un événement dans E_1 (c'est donc une quantité positive). L'égalité (2.65) décrit le gain d'information moyen associé à l'accumulation d'éléments dans un processus d'inférence bayésienne.

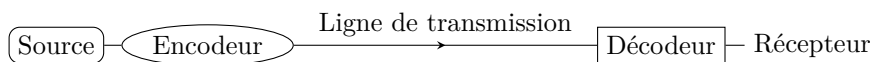


FIGURE 2.9 – Un canal de transmission classique permet de transmettre un signal encodé à partir d’une source jusqu’à un décodeur permettant de réceptionner le signal.

Si on se place du point de vue d’une chaîne particulière d’observations et d’actualisations de la probabilité des causes, un problème concret consiste à savoir quand s’arrêter. Pour cela, il faut un algorithme de décision qui peut être basé, par exemple, sur l’information qui reste à préciser.

2.5 Capacité de transmission d’un canal

Jusqu’à présent, nous avons analysé à l’aide de la théorie des probabilités deux éléments des systèmes de communication, les sources et les systèmes d’encodage. L’entropie de Shannon apparaît comme étant la bonne quantité informationnelle permettant de caractériser les ressources numériques nécessaires pour bien représenter une source.

Il nous reste encore cependant à analyser un élément majeur de tout système de communication, le canal de transmission de l’information, schématisé sur la figure 2.9. La question naturelle à se poser est de savoir s’il est possible de fidèlement transmettre à travers un canal potentiellement bruité l’information contenue dans une source et, si oui, existe-t-il une capacité limite d’un canal ? Les travaux de Shannon ont montré qu’il existait en effet une telle limite à la quantité d’information que peut véhiculer un canal de transmission sans connaître les détails de la construction de l’émetteur, du récepteur et même de la ligne. La clé pour pouvoir négliger les erreurs induites par le bruit causé par l’environnement réside dans la redondance de l’encodage de la source. C’est ce que nous allons étudier maintenant. Notons déjà que l’on commence à toucher ici à la notion de redondance de l’information qui va jouer un rôle très important dans l’émergence d’un monde classique à partir d’une théorie quantique en permettant d’éclairer la question de consensus entre différents observateurs.

2.5.1 Canal de transmission

Définition

Un canal de transmission se caractérise par trois objets. D’abord par deux ensembles de symboles, un en entrée Ω_e et un en sortie Ω_s que l’on supposera discret par commodité. Le troisième élément décrit les erreurs qui apparaissent lors de la transmission d’un symbole en entrée, qui donne donc

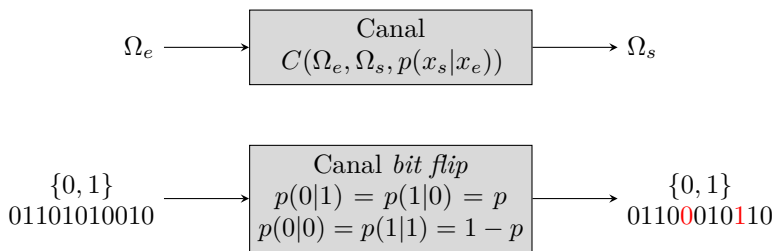


FIGURE 2.10 – Représentation d'un canal de transmission C , qui prend en entrée les symboles de l'ensemble Ω_e , et qui émet les symboles d'un alphabet Ω_s . La probabilité d'émettre un symbole donné x_s peut dépendre du symbole considéré en entrée x_e . L'origine de cette incertitude provient des erreurs engendrées par les bruits environnementaux. Dans le cas du canal numérique avec erreur dite de *bit flip*, les alphabets considérés sont $\{0, 1\}$ et le canal peut retourner l'état d'un bit avec probabilité p . Si on utilise plusieurs fois ce canal, il va retourner certains bits sans que l'on sache à l'avance lesquels, comme on peut le voir sur l'exemple, altérant ainsi l'information transmise.

lieu potentiellement à plusieurs symboles en sortie qui proviennent du bruit ajouté par le canal imparfait. Comme on ne tient pas compte de la construction physique de ce dernier, la transmission en présence d'erreurs est modélisée par des probabilités conditionnelles $p(x_s|x_e)$ de réception du symbole x_s sachant que le symbole x_e a été émis. Classiquement, un canal de transmission discret sans mémoire C est donc la donnée $C = (\Omega_e, \Omega_s, p(x_s|x_e))$.

Le canal bruité numérique élémentaire

Pour commencer, nous allons considérer l'exemple simple d'un canal de transmission de données binaires au sein duquel chaque bit est retourné (*bit flip*) avec une probabilité p . On dispose d'un message originel formé de N bits et on cherche à savoir combien de bits il faut rajouter à N pour être capable de détecter et de corriger les erreurs introduites par le canal de communication. Bien entendu, la question est posée statistiquement : on effectue un grand nombre de transmissions du symbole (message très long) et on cherche à rendre la probabilité d'erreur de décodage arbitrairement petite dans cette limite. La figure 2.10 donne un exemple particulier d'une telle erreur sur un message.

Notons N_C le nombre total de bits du message incluant les bits redondants permettant d'en effectuer le décodage. Du fait des erreurs introduites par la ligne, il faut mobiliser $N_C H(p)$ bits pour encoder l'information introduite par les erreurs dans la ligne de transmission (voir section 2.2.3). La différence $N_C - N_C H(p)$ représente alors le nombre de bits disponibles pour l'envoi du message. Si on veut pouvoir envoyer N bits de message, il faut donc envoyer N_C bits en tout avec la condition : $N_C - N_C H(p) \geq N$. Le quotient $R \simeq N/N_C$

représente le quotient de la taille du message au nombre de bits transmis en tout et mesure donc l'efficacité du dispositif de transmission⁷. Une condition nécessaire pour une bonne transmission à travers notre canal bruité s'écrit donc :

$$\frac{N}{N_C} \leq 1 - H(p) \equiv \mathcal{C}. \quad (2.66)$$

Un point important que nous ne prouverons pas ici est qu'il existe une méthode d'encodage et décodage qui permet d'atteindre ce taux de transmission [330].

Ce résultat donne donc la capacité $\mathcal{C}$ de transmission d'un canal bruité par des erreurs de type *bit flip* qui portent sur un bit. Le tableau suivant donne quelques valeurs de la capacité de transmission (bits utiles/bits transmis) en fonction du taux d'erreur p :

Taux d'erreur	N/N_C
10^{-3}	0,988 59
10^{-2}	0,919 21
10^{-1}	0,863 20
0,2	0,278 07
0,3	0,118 71

L'efficacité du canal R diminue au fur et à mesure que le taux d'erreur augmente entre $p = 0$ et $p \rightarrow 1/2$, ce qui veut dire que pour transmettre avec un taux d'erreur minimal, il faut augmenter la redondance. Lorsque $p = 1/2$, la capacité du canal devient nulle car cela revient à remplacer le message initial par une chaîne de bits totalement aléatoire. La symétrie de $1 - H(p)$ dans $p \leftrightarrow 1 - p$ s'explique en remarquant que si la ligne réalise un renversement du bit avec probabilité p , alors en appliquant un renversement systématique de tous les bits, on fabrique une ligne qui, de manière effective, retourne les bits avec probabilité $1 - p$.

Le raisonnement ci-dessus a été fait dans un cas très particulier (transmission de chiffres binaires avec erreurs sur un bit à la fois). Dans un canal de transmission réel, des erreurs à plusieurs bits sont possibles. De plus le canal peut être analogique et non purement numérique. Il faut donc étendre les raisonnements précédents pour obtenir un résultat universel exprimé uniquement en termes de concepts issus de la théorie de l'information.

2.5.2 Capacité de transmission d'un canal

Un bon moyen de tester un canal de communication consiste à écouter la sortie pour tous les émetteurs E possibles et imaginables en entrée. La capacité du canal dépend évidemment de la ligne mais aussi du récepteur R . Si on regarde uniquement l'information de R , on compte également l'information

7. On utilise parfois l'inverse $R = N_C/N$ qui représente le nombre de bits à transmettre par bit utile.

utilisée pour coder les fluctuations introduites par le bruit dans le canal de transmission. Cette information n'est autre que l'information conditionnée à l'émetteur E , à savoir $S[R|E]$. La quantité d'information réellement transmise par le canal de transmission est donc $S[R] - S[R|E] = I[R, E]$. La *capacité du canal de transmission* est donc obtenue en cherchant la valeur maximale que peut prendre l'information mutuelle pour tous les émetteurs E en entrée :

$$\mathcal{C}[C] = \max_E (I[E, R]) . \quad (2.67)$$

Examinons deux cas limites simples pour voir que cette notion correspond bien à l'idée intuitive qu'on se fait de la capacité d'un canal :

- *ligne coupée* : dans ce cas, E et R sont statistiquement indépendantes. Mais alors $I[R, E] = 0$ et donc la capacité du canal est nulle ;
- *ligne parfaite* : dans ce cas, chaque message reçu est complètement déterminé par chaque message en entrée et permet de l'identifier de manière unique. La probabilité conditionnelle $p(x_s|x_e)$ est de la forme $\delta_{x_s, f(x_e)}$ où f est injective. Ceci entraîne que $S[R|E] = 0$ et donc que $I[R, E] = S[R]$. De ce fait, la capacité du canal est donnée par l'information du récepteur placé en sortie.

En général, la capacité du canal est inférieure à la capacité d'un canal idéal dont le récepteur serait R

$$\mathcal{C}[C] = \mathcal{C}[C_{\text{parfait}}(R)] - \min_E (S[R|E]) \quad (2.68)$$

et donc, comme notre intuition le suggère, la capacité du canal est bornée ainsi :

$$0 \leq \mathcal{C}[C] \leq \mathcal{C}[C_{\text{parfait}}(R)] . \quad (2.69)$$

Il n'est pas surprenant de voir apparaître l'information mutuelle dans la capacité d'un canal et comme gain d'information dans un processus d'inférence bayésienne. Ce dernier peut en effet être vu comme un canal de transmission dans lequel l'émetteur dispose des causes et le récepteur délivre les effets. Ainsi, dans une enquête policière, un enquêteur cherche à optimiser sa capacité à remonter aux causes à partir des effets. Il cherchera donc à maximiser la capacité de transmission du canal qui amène des causes du crime aux observations qu'il a menées.

2.5.3 Second théorème de Shannon sur le codage

La capacité d'un canal de transmission trouve son application majeure dans le *second théorème de Shannon* [330] sur le code d'un canal de transmission bruité, dont nous avons vu une forme lors de notre analyse du canal bruité numérique. Nous ne le donnerons que sous une forme qualitative.

Pour un code binaire de longueur N et de cardinal $|\Omega|$, définissons son taux R par $R = \log_2 |\Omega|/N$. Ainsi si $|\Omega| = 2^N$ alors $R = 1$, ce qui veut dire que le code n'ajoute aucune redondance. Cela n'est pas le cas si $R < 1$.

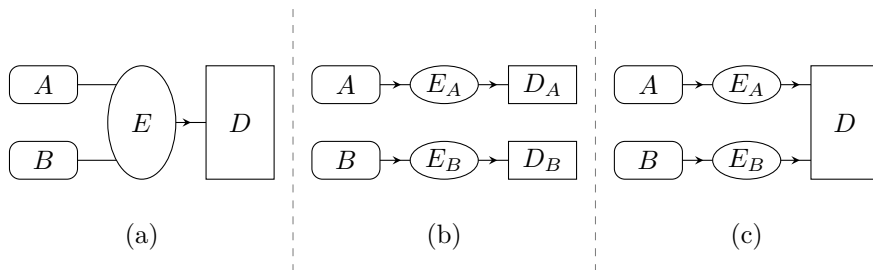


FIGURE 2.11 – Transmission de deux sources : (a) encodage et décodage joints (une seule ligne utilisée), (b) encodages et décodages indépendants, (c) encodages indépendants mais décodage joint exploitant les corrélations entre A et B .

Le second théorème de Shannon s'énonce alors de la façon suivante : pour un canal binaire de capacité C et pour tout $R \leq C$, il existe une suite de codes pour lesquels le taux s'approche arbitrairement de R et le taux d'erreur s'approche arbitrairement de zéro. Réciproquement, si $R > C$, le taux d'erreur de décodage est minoré par une constante non nulle dépendant de R et C . C'est ce théorème qui nous dit vraiment que la capacité C est le taux maximal que l'on peut espérer faire passer à travers un canal de transmission sans erreurs pour l'utilisateur. Le point absolument remarquable est que même en présence d'erreurs qui sont, par essence aléatoires, il existe une capacité qui peut être non nulle.

Sur l'exemple du canal numérique bruité, nous avons pu calculer la capacité exactement et nous avons montré ce théorème dans ce cas particulier. Les bons codes vérifient $R = N/N_C \leq C = 1 - H(p)$ où $H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$. En particulier, on constate que $C > 0$ sauf lorsque $p = 1/2$. Le cas $p = 1/2$ correspond à une aléatorisation complète des bits et il n'est donc pas surprenant que dans ce cas, on ne puisse rien récupérer. La vraie (bonne) surprise est qu'aussi proche de $1/2$ soit le taux d'erreur, il existe toujours un encodage qui permet d'avoir une capacité de transmission non nulle. Plus précisément, pour $p \simeq 1/2$, $C \simeq (1 - 2p)^2$.

2.5.4 La borne de Slepian-Wolf

Afin d'illustrer les concepts introduits dans ce chapitre, mentionnons un dernier résultat important. Nous avons vu que pour une source A de messages uniques, il est possible de la compresser en utilisant $S[A]$ bits par message. La question naturelle que l'on peut se poser est de savoir si on peut faire mieux avec une source distribuée, c'est-à-dire que l'on remplace A par (A, B) , deux sources de messages corrélées.

Le théorème de Shannon sur le codage nous dit que pour (A, B) , la compression maximale nécessite $S[A, B]$ bits par message complet. C'est ce qui se

passé lorsque l'on envoie les flux issus de A et de B dans un encodeur global pour récupérer ensuite le flux dans un décodeur global après passage dans un canal de transmission sans pertes (voir figure 2.11a). Si on opte pour une stratégie de parallélisation naïve où on encode et décode A indépendamment de B et donc deux canaux de transmission (voir figure 2.11b), on doit utiliser $S[A] + S[B]$ bits par message, ce qui est moins efficace que la stratégie précédente. Cependant, une telle stratégie peut être nécessaire si les canaux de transmission physiques dont on dispose possèdent une bande passante limitée et que $S[A, B]$ la sature alors que, isolément, $S[A]$ et $S[B]$ ne la saturent pas. La question posée ici est donc d'importance pratique.

Pour y répondre, on peut utiliser un circuit plus élaboré dans lequel deux lignes sont utilisées mais où on exploite les corrélations entre A et B . Pour cela, ils disposent chacun d'un encodeur mais les sorties des deux lignes de transmission sont envoyées dans un décodeur unique (figure 2.11(c)). Notons R_A et R_B le nombre de bits par symboles nécessaires en sortie des encodeurs de A et de B . L'encodeur de A transmet $S[A]$ bits par message et cela suffit à caractériser ses messages typiques. Imaginons que le décodeur sache exploiter les corrélations alors, quand le message typique y a été envoyé par A , l'encodeur de B peut se contenter d'utiliser $S[B|y]$ bits pour envoyer ses messages. En moyenne, la seconde ligne de transmission utilisera donc $S[B|A]$ bits pour transférer ses messages. Nous voyons donc que $R_A = S[A]$, $R_B = S[B|A]$ et donc $R_A + R_B = S[A] + S[B|A] = S[A, B]$ dans ce cas. On a récupéré l'optimum pour la source complète (A, B) .

Le *théorème de Slepian-Wolf* nous dit qu'asymptotiquement, il est toujours possible de trouver un encodage décrit par la figure 2.11(c) tel que

$$R_A \geq S[A|B], \quad (2.70a)$$

$$R_B \geq S[B|A], \quad (2.70b)$$

$$R_A + R_B \geq S[A, B]. \quad (2.70c)$$

Ces inégalités délimitent dans le plan (R_A, R_B) une zone représentée sur la figure 2.12 dont on voit qu'elle permet, du fait des corrélations entre A et B une performance supérieure au parallélisme naïf. Ce théorème donne ainsi une interprétation de l'entropie conditionnelle en termes de capacité de compression : c'est le taux de compression (sur A) que l'on peut atteindre lorsque le décodeur possède une information corrélée à la source (*via* B). L'information mutuelle $I[A; B]$ apparaît alors comme le gain entre le parallélisme naïf et ce que l'on peut espérer en exploitant au mieux les corrélations entre A et B .

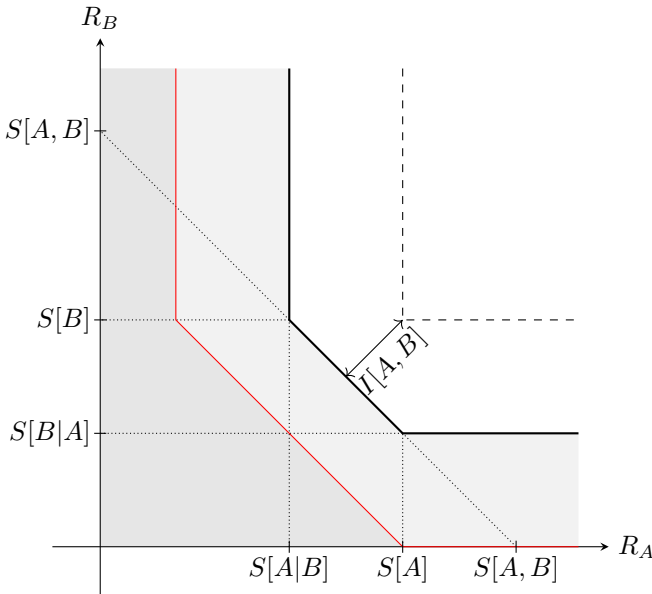


FIGURE 2.12 – Domaine accessible dans le plan (R_A, R_B) en bits par message pour la transmission à travers deux canaux de transmission parfaits. En tirets : cas d’encodages et décodages indépendants : Alice doit alors transmettre au moins $S[A]$ bits et Bob au moins $S[B]$. Traits pleins noirs : encodages exploitant la corrélation entre les sources (la zone interdite est en gris clair) pour un cas où $S[A] < S[A, B] < S[A] + S[B]$. Pour cet exemple, $S[B]/S[A] = 3/4$. La ligne rouge délimite l’extension maximale de la zone autorisée par l’inégalité (2.25). Elle correspond au cas où l’information mutuelle $I[A, B]$ qui mesure la corrélation entre A et B atteint la valeur maximale $\min(S[A], S[B])$. Le domaine en gris foncé reste donc inaccessible même en exploitant complètement les corrélations maximales possibles entre A et B . Comme ici, $S[B] < S[A]$, la corrélation maximale correspond au cas où la connaissance de A détermine parfaitement B et donc $S[A, B]$ serait égal à $S[A]$. La transmission de la source A suffirait et donc on pourrait avoir $R_A = S[A]$ et $R_B = 0$. En revanche, même en transmettant B intégralement, A devrait transmettre $S[A|B] > 0$.

Chapitre 3

Calculabilité et complexité

Il n'est pas facile de définir la notion de calcul alors que nous vivons dans une société où le calcul semble omniprésent, de la caisse du supermarché aux marchés financiers, à la prévision météo et la gestion du trafic aérien sans parler de tous les objets technologiques qui nous entourent. Cependant, malgré une grande diversité de forme, un calcul n'est finalement qu'un procédé qui permet d'évaluer la valeur d'une fonction définie dans le langage des mathématiques. Pour pouvoir l'étudier précisément, il est nécessaire de formaliser ce qu'est un calcul. Ce travail constitue le fondement de l'informatique.

Formaliser le calcul va permettre de définir ce qu'est une fonction calculable. Se posent alors deux questions fondamentales. Toutes les fonctions sont-elles calculables, ou bien existe-t-il des fonctions qui, par essence, ne le sont pas ? Est-ce que certaines fonctions sont plus difficiles à calculer que d'autres, et comment peut-on le quantifier ?

La première question définit le domaine de la calculabilité, qui s'attache à déterminer ce qui est calculable et ce qui ne l'est pas. La seconde est au cœur du domaine de la complexité algorithmique dont l'objectif est de quantifier le coût d'un calcul. La première difficulté que nous verrons apparaître sera le fait que pour formaliser le calcul, nous devons introduire un modèle de calcul. Il n'y a aucune raison *a priori* que les résultats de calculabilité et de complexité ne dépendent pas de ce modèle. Cependant, il est nécessaire de choisir un modèle pratique, qui soit physiquement réalisable. Il semble bien que tous les modèles de calcul raisonnables, à la fois assez puissants mais réalistes soient équivalents en termes de calculabilité. Cette hypothèse s'appelle la thèse de Church-Turing. En revanche, en termes de complexité, la réponse est plus subtile : nous verrons dans la seconde partie de ce livre qu'utiliser des ressources quantiques permet de calculer plus efficacement qu'avec des ressources classiques.

Pour présenter ces notions, nous avons encore une fois choisi une présentation pragmatique qui part de ce que des décennies d'ingénierie nous ont apportées, à savoir nos ordinateurs. Dans une première section, nous ferons émerger le modèle de calcul dû à Alan Turing de l'architecture de nos ordi-

nateurs¹. Dans la section suivante, nous montrerons qu’aussi puissant qu’il paraisse, ce modèle de calcul ne permet pas pour autant de tout calculer, loin de là ! Cela nous conduira à étudier des variantes équivalentes au modèle de Turing puis à examiner s’il n’existe pas d’autres modèles de calcul fondamentalement plus puissants. Nous verrons qu’il est toujours possible d’en imaginer mais que ceux-ci ne semblent pas réalisables dans un monde régi par la mécanique quantique. Nous arriverons alors à une discussion charnière de ce livre où seront présentées les thèses fondamentales qui motivent une approche purement informationnelle de l’Univers. Cela nous amènera enfin à discuter la question des ressources nécessaires pour effectuer un calcul à la section 3.4.

3.1 Avec quoi calcule-t-on ?

Nous avons annoncé vouloir formaliser la notion même de calcul dans ce chapitre. Cependant, avant d’arriver à cela, essayons de savoir ce que l’on souhaite calculer. Pour le physicien, il s’agira de prédire des résultats numériques à partir d’un modèle. La mécanique quantique telle que nous l’avons décrite au chapitre 1 est exactement cela, car elle permet de calculer une probabilité de transition en partant d’un modèle et d’une description des conditions initiales et finales. Pour le mathématicien, il s’agira de calculer le nombre de solutions d’une équation, d’obtenir une réponse binaire donnant la véracité d’un théorème ou bien de déterminer l’existence d’une solution à un problème mathématique.

En clair, quel que soit le domaine d’application envisagé, le calcul est alors un procédé qui, pour une entrée donnée, produit un résultat. Ce que l’on cherche à calculer n’est donc rien d’autre qu’une fonction. On peut néanmoins en dire plus et se rendre compte que de façon extrêmement générale, l’entrée comme le résultat sont donnés sous la forme d’une succession de symboles définis à l’avance. Par exemple, dans le cas d’une formule arithmétique, ces symboles ne sont rien d’autres que les chiffres, « + » et « × ». On appellera l’ensemble des symboles utilisés l’*alphabet*, noté Σ . L’alphabet est un ensemble fini. Une succession finie de symboles s’appelle un *mot*, un mot pouvant éventuellement être vide, c’est-à-dire ne contenir aucun symbole. On peut à l’instar de ce qui est fait dans un ordinateur actuel, se contenter d’un alphabet qui ne contient que deux symboles $\{0, 1\}$. Il suffit de coder de façon unique chaque élément de Σ comme un mot de taille n sur l’alphabet $\{0, 1\}$. Ce processus de codage fait que l’alphabet n’a pas d’importance sur les différents résultats que l’on peut obtenir en termes de calculabilité.

Si on a un alphabet Σ , on notera Σ^* l’ensemble des mots que l’on peut former à partir de Σ . Ce que l’on cherche à calculer, ce sont donc les fonctions de Σ^* dans Σ^* . On appellera une telle fonction un *problème*. Un sous-ensemble

1. Le modèle de nos ordinateurs est en fait plus proche du modèle dû à John von Neumann. Néanmoins, en s’intéressant aux caractéristiques très générales d’un ordinateur, il est possible de déterminer les prérequis nécessaires au modèle de Turing.

d'intérêt des problèmes sont les problèmes de décision, dont la réponse est binaire. Les problèmes de décision sont donc les fonctions de Σ^* dans $\{0, 1\}$. De manière équivalente, ceux-ci peuvent être décrits par le sous-ensemble des mots w de Σ^* tels que $f(w) = 1$. Un tel sous-ensemble est appelé un *langage*. Comme nous le verrons, les problèmes de décision sont particulièrement intéressants car malgré leur simplicité, ils capturent toutes les difficultés pour définir la calculabilité. Enfin, si on considère non pas un problème de façon générale, mais associé avec une entrée donnée, on parlera d'une *instance* du problème.

Notre objectif va être de dégager un modèle de calcul à la fois raisonnable et puissant. L'approche historique consiste à développer des modèles de calcul de plus en plus puissants [343, 382]. Cependant, nous connaissons tous une réalisation pratique d'un modèle de calcul : l'ordinateur. Nous allons donc essayer de construire un modèle à partir de cette réalisation, à travers une approche *bottom-up* qui ne sera rien d'autre qu'une démarche de rétro-ingénierie sur des objets que nous croisons quotidiennement. En partant des éléments de base d'un ordinateur, nous allons donc construire une façon de calculer des fonctions.

3.1.1 Du transistor à la fonction

Un ordinateur est avant tout constitué de transistors. Ces transistors servent à réaliser des portes logiques, qui elles-mêmes servent à calculer des fonctions. Nous allons donc décrire comment sont arrangés ces différents éléments, afin de donner une idée de l'architecture globale d'un ordinateur. Le lecteur souhaitant dépasser le cadre introductif que nous présentons ici pourra se référer à l'ouvrage de Tanenbaum [355], permettant d'avoir une vue plus globale de la construction d'un ordinateur.

Vers les portes logiques

Nous n'étudierons pas dans ce livre le fonctionnement physique des transistors, mais plutôt leur fonctionnement logique ainsi que la façon de les agencer dans un ordinateur. Ce paragraphe a un but plus culturel que pratique, la technologie à la base des ordinateurs pouvant être entièrement modifiée, comme elle l'a déjà été plusieurs fois. Cependant, un changement radical de technologie ne rend pas caducs les principes utilisés pour fabriquer un ordinateur.

Il peut sembler surprenant de constater qu'un objet aussi complexe qu'un ordinateur puisse être construit principalement autour de deux éléments. En effet, en ce début de XXI^e siècle, un ordinateur est majoritairement composé de deux types de transistors : les transistors NMOS et PMOS (voir figure 3.1). Ceux-ci ont des rôles complémentaires. Le transistor NMOS est passant uniquement lorsque l'on applique une tension de grille. Le transistor PMOS a un comportement exactement opposé : il est passant uniquement lorsque la tension de grille est nulle.

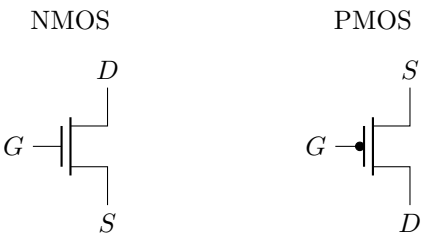


FIGURE 3.1 – À gauche : un transistor NMOS qui est passant lorsque l’on applique une tension sur la grille G . À droite : un transistor PMOS qui, à l’inverse, est passant lorsque la tension de grille est nulle. La source S et le drain D sont placés de façon à ce que le courant passe du bas vers le haut sur le schéma, leur inversion venant du fait que les porteurs de charge sont les électrons pour un transistor NMOS et les trous pour un transistor PMOS. Un ordinateur actuel est constitué essentiellement de ces deux transistors.

		x	y	$x \text{ nand } y$	x	y	$x \wedge y$	x	y	$x \vee y$
		0	0	1	0	0	0	0	0	0
0	1	0	1	1	0	1	0	0	1	1
1	0	1	0	1	1	0	0	1	0	1
		1	1	0	1	1	1	1	1	1

FIGURE 3.2 – Les tables de vérité des portes de base. De gauche à droite, on a les fonctions associées aux portes **not**, **nand**, **and** et **or**.

Pour des raisons de performances aussi bien énergétiques que fonctionnelles, on associera toujours des transistors NMOS et PMOS. C’est ce que l’on appelle la technologie CMOS, pour *Complementary Metal Oxyde Semiconductor*. En codant 0 par l’absence de tension et 1 par la présence d’une tension de l’ordre du volt, on peut ainsi créer la porte **not** et la porte **nand** (contraction de **not-and**, il s’agit d’une porte **and** suivie d’une négation) comme le montre la figure 3.3. On notera les opérations de ces portes respectivement $\neg x$ et $x \text{ nand } y$. Leurs tables de vérité sont données sur la figure 3.2.

Grâce aux portes **not** et **nand**, nous pouvons créer deux portes supplémentaires : la porte **and** ainsi que la porte **or**, dont les tables de vérité sont également données sur la figure 3.2. On notera respectivement $x \wedge y$ et $x \vee y$ les opérations effectuées par ces deux portes. On a alors $x \wedge y = \neg(x \text{ nand } y)$, et $x \vee y = \neg(\neg x \wedge \neg y)$.

Une question naturelle est alors de savoir si, à partir des portes simples que nous venons de présenter, nous pouvons réaliser toutes les opérations logiques.

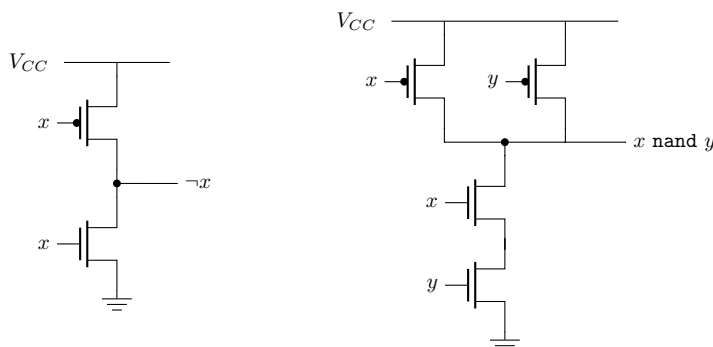


FIGURE 3.3 – Deux réalisations CMOS de portes logiques de base. À gauche : représentation d’une porte **not**. À droite : représentation d’une porte **nand**. Le 0 est codé par une tension nulle, et le 1 codé par la tension V_{CC} . Cette dernière est de l’ordre du volt. À titre d’exemple, regardons la réalisation de la porte **not**. Lorsque le signal d’entrée x a une tension nulle, le transistor du haut est passant et celui du bas bloquant. La tension de sortie est donc V_{CC} . Lorsque le signal d’entrée a une tension V_{CC} , le transistor du haut est bloquant et celui du bas passant, amenant la tension de sortie à 0. On a bien une inversion du signal.

En d’autres termes, peut-on générer toutes les portes logiques à n entrées et m sorties ? Comme nous allons le voir, c’est effectivement le cas.

Le théorème de Shannon

Une porte logique n’est rien d’autre qu’une représentation d’une fonction de $\{0, 1\}^n$ dans $\{0, 1\}^m$. Pour simplifier un peu le problème, on peut aussi voir cela comme m fonctions de $\{0, 1\}^n$ dans $\{0, 1\}$. Jusqu’à présent, nous avons considéré un ensemble réduit de portes, pour lesquelles $n \in \{1, 2\}$. Nous allons voir qu’il est possible de représenter n’importe quelle fonction de ce type à l’aide des portes logiques **or**, **and** et **not**, que nous avons construites précédemment à l’aide des transistors. Cela revient à « câbler » de nombreuses portes de bases ensemble pour en fabriquer de nouvelles.

Prenons par exemple une fonction $f : \{0, 1\}^3 \rightarrow \{0, 1\}$ définie par sa table de vérité. On peut alors faire correspondre de façon univoque sa table de vérité à son expression en termes d’opérations logiques. En partant de la table de vérité, on regarde pour quelles valeurs des arguments la fonction vaut 1. Pour chacune de ces valeurs, on construit une clause contenant chacune des variables (niée si la variable vaut 0) séparées par des **and**. En séparant chacune de ces clauses par un **or**, on obtient alors une expression pour f appelée forme normale disjonctive (c’est-à-dire qu’il s’agit d’une disjonction de conjonctions).

x	y	z	$f(x, y, z)$	
0	0	0	0	
0	0	1	0	
0	1	0	1	$f(x, y, z) = (\neg x \wedge y \wedge \neg z)$
0	1	1	0	$\vee (x \wedge \neg y \wedge z)$
1	0	0	0	$\vee (x \wedge y \wedge z)$
1	0	1	1	
1	1	0	0	
1	1	1	1	

La démonstration du cas général s'effectue simplement par récurrence. On suppose que l'on sait représenter les fonctions de $\{0, 1\}^n$ dans $\{0, 1\}$, et on considère $f : \{0, 1\}^{n+1} \rightarrow \{0, 1\}$. On a alors

$$f(x_1, \dots, x_{n+1}) = (\neg x_{n+1} \wedge f(x_1, \dots, x_n, 0)) \quad (3.2)$$

$$\vee (x_{n+1} \wedge f(x_1, \dots, x_n, 1)) \quad (3.3)$$

$(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n, 0)$ et $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n, 1)$ étant des fonctions de $\{0, 1\}^n$, donc calculables à l'aide de portes logiques, f est aussi calculable à l'aide de portes logiques. Nous venons donc de montrer le théorème suivant :

Théorème 3.1 (C.E. Shannon). *Il est possible de construire n'importe quelle fonction de $\{0, 1\}^n$ dans $\{0, 1\}^m$ à l'aide des trois portes **or**, **and** et **not**².*

Cela est un premier pas important dans la construction d'un ordinateur. Toutes les opérations de base d'un processeur, qu'il s'agisse d'opérations arithmétiques ou d'opérations logiques sont réalisées ainsi. Dans les ordinateurs actuels, n est le nombre de bits d'entrée et de sortie pour ces opérations, et il vaut en général 32 ou 64, bien que cela puisse varier selon l'architecture et les opérations demandées. Néanmoins, assembler des transistors de cette façon ne suffit pas pour faire un ordinateur. Un des problèmes est que l'on ne peut pas faire de l'arithmétique sur des nombres arbitrairement grands car cela demanderait un nombre de fils infinis. De manière générale, l'ensemble des fonctions de $\{0, 1\}^m \rightarrow \{0, 1\}^n$ est un sous-ensemble très restreint des fonctions que l'on peut penser comme calculables. Il va donc falloir affiner notre modèle de calcul. Un autre problème, que nous aborderons plus en détail dans le chapitre 8, est que l'existence d'un circuit n'est pas toujours suffisante. Il faut pouvoir construire le circuit en question. L'algorithme que nous avons vu permet en effet de construire une fonction à partir de sa table de vérité, ce qui suppose que l'on sait calculer la fonction en tout point.

2. De manière équivalente, il est possible de construire n'importe quelle fonction à partir de la seule porte **nand**, la porte **nand** permettant d'écrire les portes **or**, **and** et **not**. Le choix des portes de base est arbitraire. Les portes **or**, **and** et **not** sont bien adaptées pour écrire des formes normales disjonctives. La porte **nand** est plus adaptée pour une description en termes de transistors (en incluant le **not** si l'on cherche à en économiser le nombre).

3.1.2 Machine de Turing

Un ordinateur n'est pas seulement une unité de calcul constituée de transistors formant des portes logiques. Il comporte également des unités de mémoire qui permettent de stocker des informations. Ces informations peuvent être envoyées à l'unité de calcul pour traitement, et le résultat du calcul stocké à nouveau dans la mémoire. Cela nous donne un modèle dans lequel la limitation n'est plus basée uniquement sur le nombre de fils du processeur, mais où la taille de la mémoire doit également être prise en compte. Dans un ordinateur, celle-ci est bien sûr finie et chaque case mémoire est numérotée, de sorte que l'on puisse accéder rapidement à n'importe quel endroit de la mémoire³.

Le fait est qu'un ordinateur est trop complexe pour servir directement de modèle de calcul. Mais l'essence de l'architecture des ordinateurs permet d'imaginer un modèle simplifié plus adapté à l'étude théorique de ces questions. Ce dernier ne doit pas pour autant être simpliste, c'est-à-dire qu'il doit pouvoir calculer tout ce qu'un ordinateur peut calculer. C'est le modèle de la machine de Turing que nous allons maintenant présenter.

Remarque : Historiquement, l'ordinateur moderne est apparu bien après les travaux de Turing sur la calculabilité qui datent d'avant la seconde guerre mondiale. C'est le mathématicien John von Neumann qui proposa en 1945 l'architecture [285] qui est celle de nos ordinateurs. Sa principale originalité est l'accessibilité de la mémoire par adresse, ce qui permet de mettre en œuvre facilement le fait qu'un programme soit une donnée pour un autre programme. C'est une caractéristique essentielle que nous allons retrouver dans le modèle de calcul proposé par Turing.

Définition d'une machine de Turing

Si on simplifie à l'extrême l'architecture d'un ordinateur, on peut la décomposer en deux parties. D'une part, on a le processeur, qui est une unité de calcul associée à une très faible quantité de mémoire. Cette mémoire est conceptuellement importante : elle permet au processeur d'avoir un état interne. D'autre part, on a la mémoire, qu'il s'agisse de la RAM ou de la mémoire de masse (disque dur, SSD). Le processeur peut aller chercher des données en mémoire, effectuer des calculs, éventuellement modifier son état interne, et écrire le résultat en mémoire. Ce sont là les caractéristiques essentielles de la machine de Turing.

Celle-ci possède un état interne, donné parmi une liste finie. Elle a également accès à une mémoire, sous la forme d'un ruban infini constitué de cases, chaque case pouvant contenir un symbole parmi une liste finie. Celles-ci sont lues ou écrites par la tête de lecture/écriture de la machine. Le cycle habituel de fonctionnement est le suivant :

3. Cela est propre à l'architecture de von Neumann dont s'inspirent nos ordinateurs. La machine de Turing que nous allons présenter possède un modèle de mémoire linéaire, c'est-à-dire que pour passer de la case 3 à la case 425, il faut faire 422 pas de calcul.

1. La tête de lecture/écriture lit la case au-dessus de laquelle elle se trouve.
2. La tête de lecture/écriture écrit dans la case au-dessus de laquelle elle se trouve.
3. La tête de lecture/écriture se déplace d'une case, vers la gauche ou vers la droite.
4. La machine modifie éventuellement son état interne, selon ce qui a été lu et l'ancien état interne.

L'utilisation d'un ruban infini n'est pas problématique pour le réalisme du modèle car, en temps fini ou avec un nombre fini de pas de calcul, on n'a accès qu'à un nombre fini de cases du ruban.

Le modèle de la machine de Turing, malgré sa simplicité, permet de calculer tout ce qu'un ordinateur peut calculer. Avant de voir des exemples, formalisons cela plus précisément. Une machine peut lire et écrire dans ses cases des symboles d'un certain alphabet fini Σ , contenant un caractère spécial B (comme *blanc* qui signifie que la case est vierge ou a été effacée). À noter que l'on peut passer d'une machine qui travaille sur un alphabet quelconque fini Σ à une machine qui travaille sur l'alphabet $\{0, 1, B\}$ grâce au processus de codage que nous avons mentionné.

À l'état initial, le ruban n'est pas nécessairement vierge. Il contient les données initiales, comme l'argument de la fonction calculée par la machine de Turing, voire le programme à exécuter. Pour une machine qui effectue la somme de deux nombres par exemple, le ruban pourra contenir ces deux nombres séparés par un $\#$.

Formellement, on a donc la définition suivante :

Définition 3.2 (A. Turing). Une machine de Turing est un quadruplet (Q, q_0, Σ, δ) où :

- Q est un ensemble fini d'états ;
- $q_0 \in Q$ est l'état initial ;
- Σ est l'alphabet sur lequel la machine travaille. Il contient le caractère blanc B ;
- $\delta : Q \times \Sigma \rightarrow Q \times \Sigma \times \{<, >\}$ est la fonction de transition.

De plus, on suppose qu'à l'instant initial, le ruban ne contient que des symboles blancs à l'exception d'un mot fini constitué de symboles de $\Sigma \setminus \{B\}$, et que la tête de lecture-écriture est positionnée sur la case la plus à droite de ce mot. Nous considérerons que le ruban est infini à droite et à gauche.

L'élément central de la machine de Turing est la fonction de transition δ . Celle-ci se comprend de la manière suivante : $\delta(q_i, x_r) = (q_f, x_w, <)$ signifie « si la machine de Turing est dans l'état q_i et que sa tête de lecture-écriture lit la lettre x_r sur le ruban, alors elle écrit x_w à la place, se déplace d'une case vers la gauche, et l'état interne devient q_f ».

Dans le cas des problèmes de décisions, la machine de Turing doit donner une sortie binaire qui dit si le mot appartient au langage défini par le problème

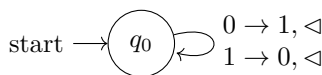
ou non. Il est toujours possible de faire des manipulations complexes sur le ruban pour écrire 0 ou 1 dessus. Cependant, il est beaucoup plus simple de rajouter un état acceptant q_a et un état rejetant q_r , qui permettent de dire si on a respectivement reconnu ou exclu le mot du langage que l'on nous a donné. Ces deux états sont les seuls états terminaux, c'est-à-dire que ce sont les seuls sur lesquels la machine peut s'arrêter. Il n'existe donc pas de transition depuis ceux-ci vers un autre état.

Si une machine de Turing s'arrête sur q_a pour tous les mots du langage L , et seulement pour ceux-ci, on dit que cette machine *reconnaît* L . Cependant, une telle machine ne suffit pas pour déterminer si un mot donné appartient à L : il peut exister des mots en dehors du langage pour lesquels la machine ne s'arrête pas. Tant que la machine n'est pas arrêtée, on ne peut pas savoir si le mot en entrée est dans le langage. Il faut donc de plus que la machine s'arrête sur q_r pour tous les mots n'appartenant pas à L . Dans ce cas, on dit que la machine *décide* L . Quand une machine décide un langage, on dit que ce langage est *décidable par machine de Turing*.

Quelques exemples de machines de Turing

You shall not pass Nous allons maintenant construire une machine de Turing capable de faire un **not** bit à bit, c'est-à-dire appliquer l'opération **not** sur chaque case du ruban. On va considérer pour cela la machine de Turing (Q, q_0, δ) , avec $Q = \{q_0\}$. La fonction de transition est alors définie de la manière suivante, donnant également une représentation graphique :

$$\begin{aligned} \delta(q_0, 0) &= (q_0, 1, \triangleleft) \\ \delta(q_0, 1) &= (q_0, 0, \triangleleft) \end{aligned}$$



On supposera que la machine commence sur la case non blanche la plus à droite. Alors, à chaque itération, on a une opération **not** et un décalage à gauche.

On a bien un **not** bit à bit. On peut noter que pour le couple (q_0, B) , la fonction δ n'est pas définie. Ce n'est pas une erreur ! En fait, si le cas rencontré n'est pas dans le domaine de définition de δ , la machine s'arrête.

Gimme one more Pour construire une machine qui ajoute un au nombre présent sur le ruban, nous utiliserons l'algorithme que nous utilisons habituellement lorsque nous posons une addition. On supposera de nouveau que l'on part à droite du nombre inscrit sur le ruban. L'idée est simple : tant que l'on tombe sur un 1, on écrit 0 et on passe à gauche. Si l'on tombe sur 0 ou B , on écrit 1 et on s'arrête.

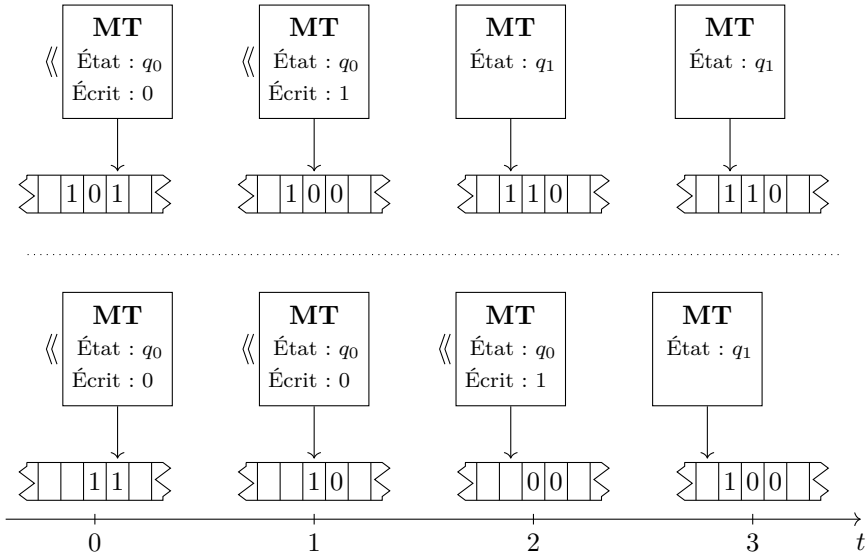
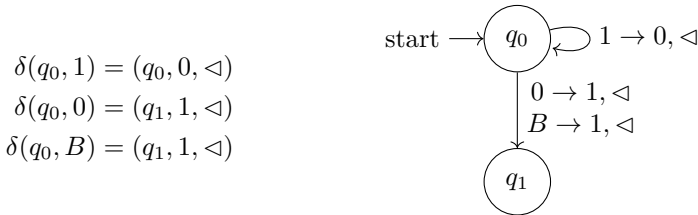


FIGURE 3.4 – Deux exemples d’évolution de la machine de Turing qui ajoute 1 au nombre inscrit sur son ruban. En haut, on inscrit 101 à l’instant initial. La machine de Turing écrit 0 jusqu’à tomber sur le premier 0, auquel cas elle écrit 1 et passe dans l’état q_1 , à la suite de quoi le calcul s’arrête. En bas, on inscrit 11 sur le ruban à l’instant initial. La machine écrit des 0 jusqu’à tomber sur la première case vide, à la suite de quoi elle écrit 1, passe dans q_1 et s’arrête.

Pour cela, nous allons avoir besoin de deux états q_0 et q_1 , q_0 étant l’état initial. La fonction de transition est alors :



Une autre manière de voir cette machine de Turing consiste à voir l’état q_0 comme « je dois reporter la retenue » et l’état q_1 comme « j’ai fini le calcul ». Cet exemple est très simple, mais grâce à cela, nous avons le début d’une machine de Turing permettant de calculer la somme de deux nombres.

Oh one On va illustrer ici un problème de décision simple : on veut construire une machine qui s’arrête sur q_a si l’entrée appartient au langage

$$L = \{0^n 1^n; n \in \mathbb{N}\} \subset \{0, 1\}^*, \quad (3.4)$$

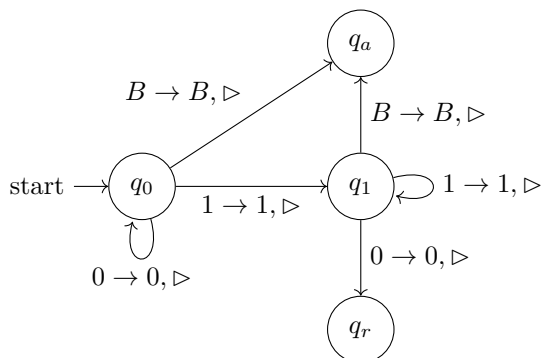


FIGURE 3.5 – Machine de Turing décidant qu’un mot est de la forme $0^m 1^n$ pour $(m, n) \in \mathbb{N}^2$. Les nœuds du graphe représentent les états internes de la machine. Chaque flèche est associée à une action décrite par le label de cette dernière : $x \rightarrow y$ signifie que si la tête lit x elle écrit y à la place et le symbole $\triangleleft$ qu’on déplace le ruban vers la droite.

qui contient l’ensemble des mots de $\{0, 1\}^*$ qui possèdent un certain nombre de zéros suivi du même nombre de uns. Par mesure de simplicité, on va d’abord vérifier que le mot est de la forme $\{0^m 1^n; m, n \in \mathbb{N}\}$. Cela est faisable par la machine de Turing dessinée sur la figure 3.5. Cette machine reste dans l’état q_0 tant qu’elle ne lit que des 0 et passe dans l’état q_1 à partir du moment où elle lit un 1. La machine passe dans l’état q_r si elle lit un 0 alors qu’elle est dans l’état q_1 , car dans ce cas, le mot contient au moins un 0 après un 1 et n’est donc pas dans L . Notons que cette machine est particulièrement simple, notamment parce qu’elle n’a jamais à revenir en arrière. Une fois que l’on a testé que le mot était de la forme $\{0^m 1^n; m, n \in \mathbb{N}\}$, il faut alors vérifier que l’on a autant de 0 que de 1. Pour cela, nous allons étendre notre alphabet à $\{0, 1, \emptyset, \bar{1}\}$. Comme la machine précédente nous amène à la fin du mot, nous allons commencer par la fin, en barrant alternativement un 1 puis un 0. C’est ce que fait la machine donnée sur la figure 3.6. On commence normalement par le 1 le plus à droite, que l’on barre. Ensuite, on va chercher le zéro non barré le plus à gauche (états q_1 et q_2). Une fois qu’on l’a trouvé (état q_3), on le barre, et on va chercher le 1 non barré le plus à droite (états q_4 et q_5). Puis on recommence, jusqu’à épuisement.

Machine de Turing universelle

Nous avons montré quelques exemples de machines de Turing spécialement conçues pour différents calculs. À chaque fois, on peut décrire la machine à travers sa fonction de transition, qui n’est autre que le programme de la

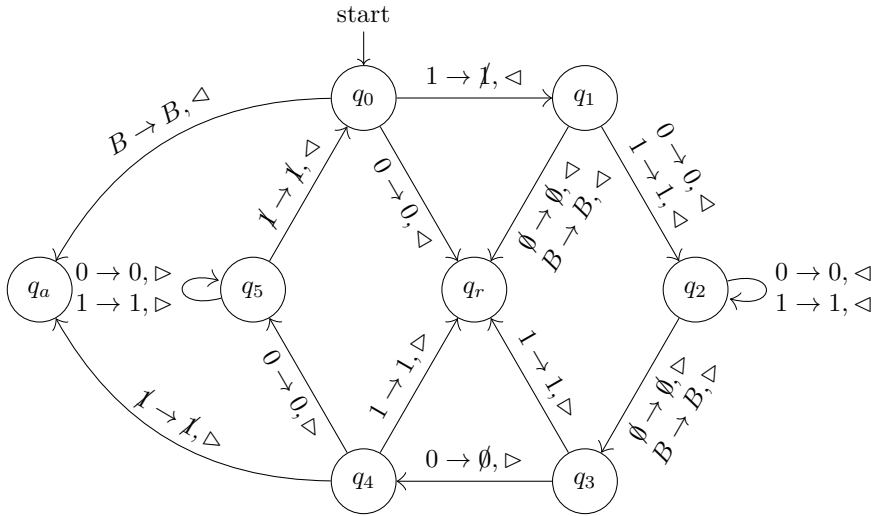


FIGURE 3.6 – Machine de Turing qui vérifie que pour un mot de la forme $0^m 1^n$ avec $(m, n) \in \mathbb{N}^2$, on a $m = n$. Les nœuds du graphe représentent les états internes de la machine. Chaque flèche est associée à une action décrite par le label de cette dernière : $x \rightarrow y$ signifie que si la tête lit x elle écrit y à la place et les symboles $\triangleleft$ et $\triangleright$ que l'on déplace le ruban respectivement vers la droite ou vers la gauche.

machine de Turing. Mais, comme à l'époque des machines Colossus⁴, on aimerait s'affranchir de la nécessité d'avoir à définir une machine différente pour chaque calcul.

L'idée est de considérer la fonction de transition, qui définit le programme de la machine de Turing, comme une donnée. Il faut donc spécifier la façon de décrire la fonction de transition⁵, ce qui nous donne une façon de coder les machines de Turing. On peut donc décrire n'importe quelle machine de Turing par un mot. Il est possible de voir ce mot comme un programme et le codage comme la syntaxe du langage de programmation. On peut alors imaginer une machine de Turing capable d'exécuter le programme (avec ses arguments) inscrit sur son ruban. Une telle machine est dite universelle, son universalité venant de sa capacité à exécuter n'importe quelle autre machine de Turing.

4. Colossus est une série de calculateurs utilisés pendant la seconde guerre mondiale pour le décryptage des codes secrets utilisés par le haut commandement allemand. Ces machines ne stockaient pas les programmes qu'elles exécutaient en mémoire mais dans leur câblage même.

5. N'importe quelle façon d'énumérer la fonction de transition convient. Pour l'argument (q_i, r) tel que $\delta(q_i, r) = (q_j, w, d)$, on pourra par exemple choisir le mot $i\#r\#j\#w\#d\$$, où l'on utilise $\#$ pour séparer chacun des éléments du quintuplé, et $\$$ pour marquer la fin du mot. Il suffit ensuite de concaténer ces mots pour l'ensemble des arguments possibles.

L'existence d'une machine de Turing universelle n'est pas quelque chose d'évident. C'est un des résultats majeurs obtenus par Turing [362]. La preuve consiste à définir une machine à plusieurs rubans (voir section 3.2.2 pour la définition) qui encode l'état interne de la machine de Turing que l'on veut simuler. C'est ainsi que, grâce à une machine avec un nombre d'états limité, on peut simuler le fonctionnement d'une machine ayant une complexité bien supérieure. Dans les années 1950, Shannon posa la question de savoir quelle était la plus petite machine de Turing universelle [329] et a montré qu'il fallait au moins deux états internes pour une machine à un ruban.

La machine de Turing universelle a joué un rôle crucial dans le développement de l'informatique en introduisant l'idée que le programme lui-même était une donnée. De nos jours, les microprocesseurs sont des machines universelles et, de fait, on retrouve les machines universelles partout.

3.2 Calculable et incalculable

Nous avons à présent tous les outils nécessaires pour définir et étudier la notion de calculabilité. Nous allons nous restreindre à l'étude des problèmes de décision.

Définition 3.3 (Langage décidable). Un langage L , ou problème de décision, est *calculable*, ou *décidable*, s'il est décidable par machine de Turing. Sinon, il est *indécidable*.

N'étudier que les problèmes de décision ne restreint pas la généralité de notre propos. En effet, à une fonction $x \mapsto f(x)$ qui se calcule par machine de Turing, nous pouvons associer simplement un problème de décision suivant :

Est-ce que pour x et y donnés, $y = f(x)$?

Si f est calculable par machine de Turing alors $f(x)$ est un mot de Σ^* . Pour x donné, on peut donc essayer chacun des y par ordre de taille⁶ car on est certain qu'il existe un nombre fini de y précédant $f(x)$ en utilisant cet ordre. Réciproquement, si f n'est pas calculable, il est évident que le problème de décision ne l'est pas non plus étant donné qu'il est nécessaire de calculer f .

Cette définition des fonctions calculables comme étant exactement celles calculables par machine de Turing est connue sous le nom de *thèse de Church-Turing*, que l'on énoncera et justifiera en seconde partie de ce chapitre. On va commencer par se demander si tous les problèmes de décision sont décidables par machine de Turing.

6. On peut ensuite classer les y de même taille par ordre alphabétique.

3.2.1 Problèmes indécidables

Problème de l'arrêt

Il existe un problème de décision pour lequel il est possible de montrer simplement l'indécidabilité : il s'agit du problème de l'arrêt. On peut formuler ce problème simplement de la façon suivante : est-ce qu'une machine de Turing donnée s'arrête sur un mot donné ? Nous allons voir que supposer ce problème décidable mène à une contradiction.

Plus précisément, le problème de l'arrêt est l'ensemble des mots (m, x) tels que le calcul de la machine M de programme ⁷ m sur l'entrée x s'arrête.

Théorème 3.4 (A. Turing). *Le problème de l'arrêt est indécidable.*

Pour le prouver, supposons le problème de l'arrêt décidable. Il existe donc une machine de Turing H de code h qui accepte une entrée (m, x) si et seulement si la machine de code m sur l'entrée x s'arrête. Nous allons maintenant créer à partir de H une machine de Turing paradoxale P grâce au procédé dit diagonal. Pour cela, considérons la machine P qui, sur une entrée e :

- simule H sur l'entrée (e, e) pour savoir si le calcul de la machine de code e sur l'entrée e s'arrête ;
- lorsque la réponse est oui, P se met à boucler indéfiniment et le calcul ne finit jamais ;
- si la réponse est non, alors le calcul de P s'arrête.

Nous avons donc défini une machine de Turing P dont la machine H est un sous-programme. Considérons maintenant p le code de P . Est-ce que le calcul de P sur l'entrée p s'arrête ? S'il s'arrête, c'est que par définition de P , le calcul de $H(p, p)$ donne 0. Par définition de H , le calcul de P sur p ne s'arrête donc pas. De la même manière, si nous supposons que le calcul de P sur p ne s'arrête pas, nous en déduisons que le calcul de P sur p s'arrête. La machine P est donc paradoxale, et cela provient du fait que la machine H ne peut exister. Le problème de l'arrêt est donc indécidable.

Cela ne veut pas dire qu'on ne peut jamais montrer qu'un programme s'arrête, et c'est heureux. Il est d'ailleurs relativement simple de montrer que les machines de Turing que l'on a construites à la section précédente s'arrêtent pour toutes les entrées possibles. Cependant, il n'existe pas de procédure qui puisse le déterminer pour n'importe quel programme.

Avant d'étudier plus en détail les problèmes indécidables, on peut se demander ce que l'on pourrait faire d'une machine capable de résoudre le problème de l'arrêt. En d'autres termes, est-ce que résoudre le problème de l'arrêt permettrait de résoudre tous les problèmes ? Nous pouvons commencer par montrer que cela nous donnerait accès à tous les problèmes qui sont réfutables en temps fini, par exemple la conjecture de Goldbach :

7. Un codage des machines de Turing ayant été choisi.

Tout entier pair supérieur ou égal à 4 peut s'écrire comme la somme de deux nombres premiers.

Dans le cas où la conjecture de Goldbach serait fausse, il suffirait de donner le premier entier qui ne la satisfait pas et de montrer qu'il est impossible de l'écrire comme une somme de deux nombres premiers. On peut écrire un programme qui teste tous les entiers les uns après les autres et qui s'arrête dès qu'il trouve un entier où la conjecture de Goldbach échoue. Si ce programme s'arrête, la conjecture de Goldbach est fausse, s'il ne s'arrête pas, elle est vraie. Ainsi, si nous disposions d'une machine capable de résoudre le problème de l'arrêt, la conjecture de Goldbach serait résolue par cette machine.

En revanche, l'ensemble des problèmes que l'on ne peut pas décider en temps fini resteraient inaccessibles même en ayant accès à une machine de Turing résolvant le problème de l'arrêt. En réalité, il y a toute une hiérarchie dans les problèmes indécidables [302].

Nombre de problèmes indécidables

Puisqu'il existe des problèmes indécidables par machine de Turing, on peut se demander quelle est la proportion de ces problèmes parmi l'ensemble des problèmes. Pour cela, on va « compter » le nombre de programmes et le nombre de fonctions de décision. Comme l'ensemble des programmes et l'ensemble des fonctions de décision sont infinis, il va falloir étendre la notion de cardinal pour des ensembles infinis. On dit que deux ensembles sont de même cardinal s'ils sont en bijection. On va commencer par s'intéresser à l'ensemble des programmes.

On peut considérer la machine de Turing universelle telle que l'ensemble des programmes est donné par l'ensemble des mots de $\{0, 1\}^*$. On peut donc écrire

$$P = \{\epsilon, 0, 1, 00, 01, 10, 11, \dots\}, \quad (3.5)$$

où P désigne l'ensemble des programmes et ϵ est le mot vide. Cette description, où l'on énumère les programmes ordonnés par taille puis par ordre alphabétique est complète. En d'autres termes, on peut associer un entier naturel à chaque programme et réciproquement, on a donc le même nombre de programmes que d'entiers naturels. On dit que l'ensemble des programmes est *dénombrable*. Il reste donc à comparer $\mathbb{N}$ et l'ensemble des fonctions de décisions.

Une fonction de décision est une fonction de $\{0, 1\}^* \rightarrow \{0, 1\}$. Comme on peut mettre en correspondance $\{0, 1\}^*$ avec $\mathbb{N}$, il s'agit en fait d'étudier l'ensemble des suites à valeurs dans $\{0, 1\}$. On va supposer que l'on peut compter cet ensemble, et nous verrons que cela nous mène à une contradiction. Cette supposition nous permet d'énumérer les suites : $(x_i^{(n)})_{i \in \mathbb{N}}$ désigne la n -ième suite de l'ensemble des suites. On peut alors imaginer que l'on a le tableau donné sur la figure 3.7. On va donc chercher à créer une suite qui ne

n	$x_0^{(n)}$	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$\dots$
0	0	0	0	1	$\dots$
1	0	1	0	0	$\dots$
2	1	1	0	1	$\dots$
3	0	0	1	0	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

FIGURE 3.7 – Si l'on suppose que l'ensemble des suites à valeurs dans $\{0, 1\}$ et $\mathbb{N}$ sont en bijection, on peut faire correspondre à chaque entier une suite de façon unique. On peut alors construire une suite qui n'est pas dans le tableau à partir de la « diagonale » en prenant la négation du n -ième bit de la suite n .

soit pas déjà dans le tableau à partir des suites du tableau. Pour cela, il suffit de considérer la suite donnée par :

$$u_n = 1 - x_n^{(n)} = \begin{cases} 0 & \text{si } x_n^{(n)} = 1, \\ 1 & \text{si } x_n^{(n)} = 0. \end{cases} \quad (3.6)$$

L'élément u_n est différent de $x_n^{(n)}$. La suite (u_n) est donc différente de toutes les autres suites que l'on a pu énumérer. On ne peut donc pas énumérer les suites. Cette démonstration, appelée *procédé d'extraction diagonal de Cantor*, est exactement celui qui permet de montrer qu'il y a bien plus de nombres réels que de nombres rationnels⁸.

Nous venons donc de montrer qu'il y a strictement plus de problèmes de décision que de programmes pour une machine de Turing universelle. Cela implique notamment que l'ensemble des problèmes de décision est indénombrable, et qu'il existe donc infiniment plus de problèmes de décision que de programmes. Si on tire une fonction « au hasard », celle-ci ne sera pas décidable. Il existe en fait un résultat très général, qui caractérise l'indécidabilité pour toute une classe de fonctions. Ce résultat est le théorème de Rice et peut s'énoncer de la façon suivante :

8. Pour cela, on fait correspondre à chaque élément de la suite binaire un nombre réel dans $[0, 1]$. On peut être tenté pour cela d'utiliser simplement la suite comme écriture binaire de ce nombre. Cependant, l'écriture dans une base donnée n'est pas unique (en binaire par exemple on a $0,111\dots = \sum_{n>0} 2^{-n} = 1$). Énumérer l'ensemble des suites va donc revenir à compter certains réels deux fois. On peut contourner la difficulté en utilisant la suite binaire comme l'écriture d'un nombre ternaire, qui assure de ne pas compter de réels deux fois, au détriment néanmoins de ne pas réaliser de surjection entre $[0, 1]$ et l'ensemble des suites binaires.

Théorème 3.5 (H.G. Rice). *Toute propriété sur les fonctions qui n'est pas soit vraie soit fausse pour toutes les fonctions est indécidable.*

Une propriété vraie (ou fausse) pour toutes les fonctions pouvant être qualifiée de triviale, cela revient à dire que seules les propriétés triviales sur les fonctions sont décidables.

Toutes les autres propriétés sont indécidables. Afin d'éclairer la discussion, insistons sur la différence entre une fonction, qui est un objet mathématique abstrait associant une valeur à une autre, et un programme qui est un mot destiné à être lu par une machine de Turing universelle et qui donne une procédure pour associer une valeur à une autre. Dans le cas d'une fonction calculable, il est possible de représenter la fonction par un programme. Notons que ce programme n'est pas unique : il existe de nombreuses façons de calculer la même fonction. Ce que le théorème de Rice nous apprend, c'est qu'étant donné un programme quelconque, on ne peut dire *en général* si la fonction qu'il calcule est constante, si elle est croissante, égale à 0... Notons qu'il s'agit uniquement du cas général. Ce théorème n'interdit pas de montrer des propriétés sur certaines fonctions. Simplement, il n'existe pas de façon automatique pour le faire qui marche pour n'importe quelle fonction.

Une précision importante est que la propriété en question doit être une propriété de fonction calculée, ce qui signifie qu'elle ne doit pas être spécifique au programme qui la calcule. Par exemple, il est tout à fait possible de demander si un programme donné s'arrête sur une entrée donnée en moins de dix pas de calcul. Si un tel programme existe, alors il en existe également un autre, calculant la même fonction, ne satisfaisant pas la propriété (il suffit de ne rien faire pendant les dix premiers pas de calcul puis d'appeler le premier programme).

Une conséquence simple de ce théorème est qu'il est notamment impossible de donner un algorithme général qui puisse déterminer si un programme donné est malveillant [160]. Quand bien même on arriverait à formaliser ce qu'est un programme ayant un comportement malveillant, il est impossible de déterminer de manière générale si un programme est malveillant ou pas. Cela n'empêche néanmoins pas d'avoir des programmes qui permettent de détecter certains de ces programmes malveillants, comme le font les anti-virus par exemple. Néanmoins, l'anti-virus parfait n'existe tout simplement pas.

Les nombres réels calculables et les autres Avant de quitter cette discussion de l'indécidabilité, revenons vers des problèmes de calcul plus familiers au lecteur physicien de cet ouvrage en abordant la question de la calculabilité des nombres réels. Cette discussion nous montrera sur cette question particulière la subtilité des questions portant sur la calculabilité.

Intuitivement, un nombre réel sera calculable si la suite de ses décimales dans une base donnée quelconque est calculable par une machine de Turing, c'est-à-dire par un de nos ordinateurs. Cela signifie que le réel r est calculable s'il existe une machine de Turing qui sur l'entrée n calcule le n -ième chiffre

de l'écriture de r dans la base considérée. Les mathématiques élémentaires fournissent un grand nombre de réels calculables : les nombres rationnels bien sûr mais aussi π et e que l'on peut exprimer comme des séries convergentes à coefficients rationnels ainsi que $\sqrt{2}$ qui possède un développement en fraction continue construit sur les entiers. Les nombres réels calculables ont des propriétés agréables : la somme et le produit de deux nombres calculables sont calculables. L'inverse d'un nombre calculable non nul est aussi calculable : les réels calculables forment donc un sous-corps du corps des réels.

On peut donc se réjouir du fait qu'il existe énormément de réels calculables. Cependant, comme ils peuvent être décrits par les programmes permettant leur calcul, ils sont également dénombrables. Les réels calculables sont donc beaucoup moins nombreux que les nombres réels. Cela signifie que si l'on prend un réel au hasard, celui-ci est incalculable. On aimerait trouver un exemple de nombre réel parfaitement défini mais qui ne soit pas calculable. Ce problème n'est pas évident car un tel nombre ne peut être défini par une « formule simple » comme un développement en série ou en fractions continues, ni par une équation ne faisant appel qu'à des fonctions calculables.

Il existe cependant un exemple très simple basé sur le problème de l'arrêt sur une machine de Turing universelle donnée. Les programmes tournant sur cette machine de Turing étant dénombrables, on peut les numéroté. On fabrique alors un nombre réel plus petit que 1 en écrivant ses décimales en base 2 : la n -ième décimale est 0 si le n -ième programme s'arrête et 1 s'il ne s'arrête pas. Ce nombre, appelé nombre de Turing, code donc la liste des programmes qui s'arrêtent. Il n'est donc pas calculable, car cela reviendrait à décider du problème de l'arrêt par une machine de Turing. On en connaît toutefois un nombre infini de décimales car on sait fabriquer un nombre infini de programmes qui s'arrêtent.

Nous verrons dans la prochaine section un autre nombre réel non calculable défini à partir du problème de l'arrêt. Mentionnons également que l'ensemble des réels calculables n'est pas fermé par borne supérieure : il existe des suites croissantes et bornées de réels calculables qui convergent vers un réel non calculable [346].

3.2.2 Variantes des machines de Turing

Notre définition des fonctions calculables, que l'on a identifiées comme celles pouvant être calculées par machine de Turing, peut sembler arbitraire dans le sens où celle-ci dépend *a priori* du modèle que l'on a choisi. Pour que cette définition ait un sens général, il faut donc s'assurer que le modèle de calcul sous-jacent importe peu. L'objet de cette section ainsi que de la section 3.2.3 va être d'explorer un ensemble de modèles de calcul « raisonnable », et de voir que ceux-ci sont équivalents, en termes de calculabilité, à la machine de Turing à un ruban. Dans cette section, nous allons étudier des variantes de machine de Turing.

Pour établir l'équivalence entre deux modèles de calcul, il faut pouvoir simuler le premier modèle à l'aide du second et réciproquement. Cela permet d'établir que n'importe quelle fonction calculable sur le premier modèle l'est aussi sur le second.

✱ Machines de Turing à rubans multiples

La variation la plus simple à laquelle on puisse penser est sans doute la machine de Turing à plusieurs rubans. La machine à un ruban étant un cas particulier de machines à plusieurs rubans, il suffit de montrer que l'on peut simuler une machine de Turing à plusieurs rubans à l'aide d'une machine de Turing à un ruban pour avoir équivalence des deux modèles de calcul. Construire une telle simulation de A à Z étant assez laborieux, on ne donnera que les idées de cette construction.

Pour simplifier, on va simuler la machine à deux rubans à l'aide de la machine à un ruban. Il suffira de réitérer ce processus pour généraliser à un nombre arbitraire de rubans. On va considérer que l'alphabet de la machine de Turing à deux rubans est $\{0, 1, B\}$. Pour la machine de Turing à un ruban, il va nous falloir un séparateur pour séparer les deux rubans, ainsi qu'une façon de repérer la tête de lecture de chaque ruban. On va donc choisir l'alphabet $\{0, 1, \dot{0}, \dot{1}, V, \dot{V}, \#, B\}$, où le point suscrit marque la position de la tête de lecture et le dièse fait la séparation entre les deux rubans. On rajoute également le caractère V qui désigne les cases qui ont été lues par la machine de Turing sans nécessairement être écrites. Il faut alors traduire la fonction de transition de la machine de Turing à deux rubans.

Pour cela, il faut d'abord aller lire le caractère sur le premier ruban, puis le caractère sur le second ruban et enfin effectuer la transition. Sur notre machine à un seul ruban, le premier ruban correspond à la zone située à gauche du caractère $\#$. On va donc scanner la zone à gauche du $\#$ pour chercher le caractère avec le point suscrit, qui marque la position de la tête de lecture. Une fois lu, on rentre alors dans un « sous-programme » spécifique à l'état de la machine de Turing et au caractère lu. Dans ce sous-programme, on va chercher le caractère possédant le point suscrit dans la zone à droite du $\#$. À partir de là, on peut effectuer la transition correspondant à chacun des rubans, en déplaçant également le point suscrit comme il convient. Si l'on rencontre un $\#$ en effectuant la transition côté ruban 1 ou côté ruban 2, il faut alors décaler la moitié correspondante du ruban. C'est là que le marqueur V est nécessaire, c'est lui qui donne la condition d'arrêt de cette recopie.

Cette méthode de simulation montre qu'utiliser des machines de Turing à rubans multiples ne change pas les propriétés de calculabilité.

Machine de Turing non déterministe

Une autre façon d'étendre les machines de Turing consiste à considérer une machine de Turing non déterministe : à chaque pas de calcul, la machine

« explore » plusieurs possibilités. Afin de préciser ce que l'on entend par « explorer », commençons par décrire le modèle. La première modification consiste à changer la fonction de transition

$$\delta : Q \times \Sigma \rightarrow \mathcal{P}(Q \times \Sigma \times \{<, >\}), \quad (3.7)$$

où $\mathcal{P}(E)$ désigne l'ensemble des parties de l'ensemble E . La fonction de transition renvoie donc non plus une unique transition, mais un ensemble de transitions possibles. Le langage accepté par la machine est alors défini de la manière suivante. La machine accepte un mot x s'il existe au moins un calcul de la machine qui l'accepte. Dans ce cas, s'il existe une telle possibilité, la machine retient la plus petite suite.

Une façon commode de représenter l'exécution d'une telle machine est sous la forme d'un arbre (voir figure 3.8). Partant de l'état initial (la racine de l'arbre), la machine effectue un ensemble de transitions : cela génère des nœuds fils, qui représentent les nouveaux états après chacune des transitions possibles. On recommence ensuite à partir de ces nœuds et cela soit jusqu'à tomber sur un état acceptant, soit jusqu'à ce que toutes les branches terminent sur un état rejetant. On peut dire qu'une telle machine explore simultanément tout l'arbre des calculs possibles et s'arrête en acceptant le mot en entrée si une des branches s'arrête sur un état acceptant. Cette définition impose notamment que la machine de Turing n'est pas probabiliste : les transitions ne sont pas choisies « au hasard ». Elles sont choisies de façon à trouver l'un des plus courts chemins acceptants, s'il existe.

Insistons sur le fait que la machine de Turing non déterministe est un modèle théorique. On peut imaginer deux manières de faire des transitions multiples. Par exemple, la machine peut être auto-répliquante et se répliquer à chaque fois qu'il y a une transition multiple. Si l'une des répliques atteint un état acceptant, alors elle envoie un signal pour arrêter l'ensemble du calcul. La machine auto-répliquante explore ainsi l'ensemble des possibles de manière exhaustive. Une autre idée, plus économique mais également plus mystique, consiste à dire qu'une machine de Turing déterministe dispose d'un outil pour faire la « bonne » transition à chaque fois. Cela revient à lui adjoindre un oracle, c'est-à-dire une boîte noire capable de déterminer à chaque fois la bonne transition. L'oracle contient donc toute la magie et le caractère non déterministe de la machine de Turing non déterministe. Cela illustre surtout qu'un modèle de calcul est avant tout un objet théorique qui n'a pas nécessairement d'interprétation physique évidente.

Ainsi, la machine non déterministe semble beaucoup plus puissante que sa contrepartie déterministe que nous avons utilisée jusqu'à présent. Il est néanmoins tout à fait possible d'utiliser une machine déterministe pour simuler une machine non déterministe.

L'idée est assez simple, bien que la mise en œuvre sous forme de machine de Turing soit assez compliquée : il « suffit » de simuler chaque branche du calcul l'une après l'autre. Comme il n'est pas garanti que toutes les branches

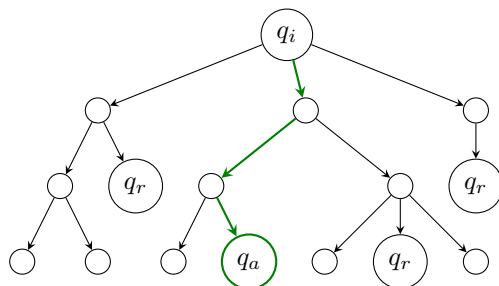


FIGURE 3.8 – Représentation de l'exécution d'une machine de Turing non déterministe. À chaque étape du calcul, la machine de Turing peut faire une transition vers différents états (nœuds de l'arbre). La machine s'arrête dès que l'une des branches rencontre un état acceptant, dans les cas où il existe une suite de transitions qui mène à un tel état. Si ce n'est pas le cas, la machine s'arrête et rejette le mot si et seulement si chacune des branches s'arrête (nécessairement sur l'état q_r).

s'arrêtent, on ne peut pas exécuter une branche jusqu'à la fin avant de passer à une autre branche. La bonne approche est de simuler toutes les branches jusqu'à un pas de calcul donné. Si aucune de ces branches ne termine sur un état acceptant, alors on recommence la simulation un pas de calcul plus loin. De cette façon, si la machine de Turing non déterministe accepte un mot, notre simulateur déterministe également. On peut procéder de même pour montrer que si toutes les branches de la machine de Turing non déterministe terminent (en d'autres termes, si cette machine décide un langage particulier), alors le simulateur déterministe s'arrête aussi.

On a donc montré que bien que la machine de Turing non déterministe semble incroyablement plus puissante, par sa capacité à « tester » un nombre de cas gigantesque de manière extrêmement rapide, que la machine de Turing déterministe ces deux machines sont équivalentes en termes de calculabilité.

Une machine non déterministe peut aussi être vue comme une machine probabiliste extrêmement chanceuse. À chaque pas de calcul, la machine tire au sort le résultat parmi les possibilités que lui donne sa fonction de transition. Elle est cependant chanceuse et la bonne transition est choisie à chaque fois. Nous reviendrons sur les machines probabilistes dans la section 3.4 consacrée à la complexité.

3.2.3 Autres modèles de calcul

Pour l'instant, nous avons étudié plusieurs modèles de machine de Turing qui s'avèrent pouvoir calculer exactement la même chose. Cela nous conforte dans une définition de la calculabilité basée sur cette machine de Turing déterministe à un ruban. Il existe néanmoins des modèles de calcul complètement différents des machines de Turing. Une discussion détaillée de ces modèles

dépasserait le cadre de cet ouvrage, nous allons juste donner une intuition de ce qu'ils recouvrent.

λ -calcul et fonctions récursives

λ -calcul Le λ -calcul a été inventé par A. Church [96]. Ce modèle de calcul est fondé sur des fonctions dont on aurait enlevé l'ensemble de définition, et que l'on appelle des λ -termes. Étant donné que l'ensemble de définition disparaît, on peut appliquer ces « fonctions » à des variables, comme on le ferait traditionnellement, mais également à n'importe quel autre λ -terme.

Dans le λ -calcul, tout est λ -termes, et le programme comme les données sont des λ -termes. Un calcul consiste à réduire le λ -terme qui contient à la fois le programme et les données jusqu'à une forme minimale, comme on le ferait en appliquant une fonction à son argument. Un pas de calcul consiste en une étape de réduction élémentaire, selon un ensemble de règles données.

Nous ne rentrerons pas dans la définition formelle de ce modèle, ni dans la démonstration de son équivalence en termes de calculabilité avec les machines de Turing [362]. Ce modèle est très proche de la programmation fonctionnelle, que l'on retrouve dans les langages comme Lisp, Caml ou Haskell, et dans laquelle l'accent est mis sur les fonctions. Tout comme dans le λ -calcul, en programmation fonctionnelle, les fonctions sont des objets de première classe, et peuvent être manipulées comme des variables habituelles.

Un autre intérêt du λ -calcul est que c'est l'approche privilégiée pour faire le lien entre un système logique formel et un programme. Ce lien, appelé isomorphisme de Curry-Howard [115, 116, 210], permet de mettre en correspondance une preuve dans un système logique donné et un programme dans un modèle de λ -calcul auquel on rajoute un système de typage correspondant au système logique. Cette correspondance permet donc de montrer les liens forts entre les mathématiques et l'informatique.

3.2.4 Hypercalcul

Les divers modèles que l'on a étudiés jusqu'à présent sont tous équivalents en termes de calculabilité. Cependant, au niveau conceptuel, rien n'empêche d'introduire des modèles plus puissants, capables de résoudre le problème de l'arrêt par exemple. Nous allons donc dans cette section étendre la machine de Turing dans ce but, ouvrant ainsi la possibilité de modèles de calcul fondamentalement plus puissants que le calcul par machine de Turing. Le point crucial est que la physique quantique semble interdire ces modèles (dits d'*hypercalcul*) d'exister dans notre univers, alors qu'il est parfaitement possible de réaliser une machine de Turing.

La discussion qui suit fera appel à des notions qui dépassent le contenu des chapitres précédents car nous allons mélanger des arguments venus de la physique quantique avec d'autres issus de la théorie de la relativité. C'est hélas le prix à payer pour étayer notre assertion selon laquelle l'hypercalcul

n'est pas réalisable dans l'Univers dans lequel nous vivons. Le lecteur qui ne souhaiterait pas s'approcher de ce gouffre pourra donc sauter cette section en première lecture et passer directement à la section 3.3, voire même ne jamais y revenir pour peu qu'il fasse sienne l'opinion de Martin Davis [120] selon laquelle l'hypercalcul n'est pas un sujet pertinent, opinion qui ne fut évidemment pas partagée par les promoteurs du sujet en question [348].

Machine de Turing munie de la probabilité d'arrêt Une première possibilité consiste à prendre une machine de Turing normale ayant accès à une donnée bien choisie. Ici, la donnée est la probabilité d'arrêt Ω , aussi appelée nombre de Chaitin. Celle-ci est définie comme la probabilité qu'un programme auto-délimité, c'est-à-dire contenant l'indication de sa propre fin⁹, tiré au hasard s'arrête sur notre machine de Turing universelle. Mathématiquement, ce nombre peut s'écrire comme une somme sur l'ensemble Stop_U des programmes auto-délimités pouvant être exécutés sur une machine de Turing universelle U et qui s'arrêtent

$$\Omega = \sum_{p \in \text{Stop}_U} 2^{-|p|}, \quad (3.8)$$

où $|p|$ désigne la longueur du programme en question encodé en binaire. Pour comprendre que cela définit bien un nombre plus petit que l'unité, organisons les programmes auto-délimités en un arbre binaire dont les branchements correspondent à chaque choix possible pour chaque bit à partir du début du programme. Un programme auto-délimité est alors une feuille de l'arbre (voir figure 3.9). Ce sont les points terminaux de l'arbre : le programme étant auto-délimité par une séquence particulière, on ne peut le rallonger en y rajoutant du code. Intuitivement, les feuilles situées à la profondeur N sont au maximum en nombre 2^N . On comprend donc qu'en affectant à un programme auto-délimité le poids $1/2^{|p|}$ où $|p|$ est sa longueur, on obtienne une mesure de sa probabilité dans un tirage aléatoire de tous les programmes auto-délimités.

Le tirage est réalisé en tirant au hasard le premier bit, puis le second, jusqu'à tomber sur une feuille de l'arbre des programmes. Le nombre Ω représente alors la fraction des feuilles correspondant aux programmes qui s'arrêtent sur la machine de Turing considérée.

Pour se convaincre que $\Omega < 1$, imaginons que nous considérons l'arbre jusqu'à la profondeur N . Dans un arbre binaire complet, en dessous d'une feuille située à la profondeur $l \leq N$, il y aurait 2^{N-l} sommets (en incluant la feuille en question). Si on somme maintenant sur les feuilles de l'arbre des programmes auto-délimités tronqué à la profondeur N , il lui manque $\sum_{p \in \text{Stop}_U} 2^{N-|p|}$ sommets par rapport à l'arbre binaire complet où la somme porte sur les p tels que $|p| < N$. Mais ce nombre est forcément inférieur à 2^N et donc, sur un arbre binaire de profondeur N , $\sum_f 2^{-|f|} \leq 1$. Cette inégalité simple sur les arbres binaires montre, en prenant la limite $N \rightarrow +\infty$, que le

9. Par exemple au moyen d'une séquence terminale particulière comme 00.

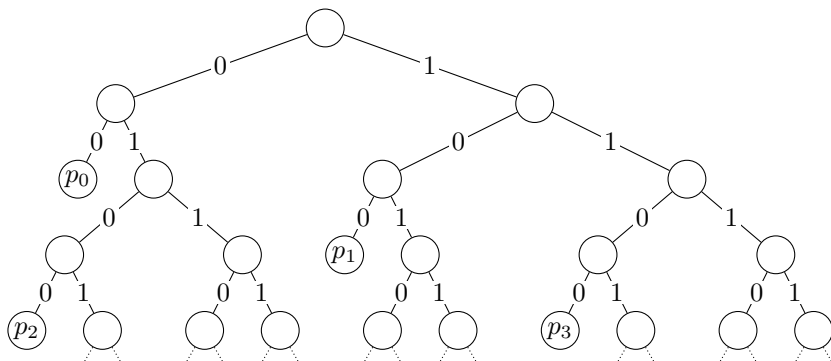


FIGURE 3.9 – Arbre binaire des programmes auto-délimités d'exécutant sur une machine de Turing universelle donnée. Seules certaines des feuilles sont des programmes qui s'arrêtent. Le nombre de Chaitin Ω est la fraction des feuilles correspondant aux programmes qui s'arrêtent.

membre de droite défini par (3.8) étant une somme sur une partie des feuilles de l'arbre des programmes auto-délimités, il est bien compris entre 0 et 1.

Cette discussion montre aussi que les N premiers bits de Ω donnent, en binaire, le nombre de programmes de N bits ou moins qui s'arrêtent. On peut alors imaginer l'expérience suivante : on exécute tous les programmes jusqu'à un certain pas de calcul, avant d'aller au pas de calcul suivant. Les programmes vont s'arrêter petit à petit jusqu'à ce qu'on tombe sur le nombre donné par la probabilité d'arrêt. À partir de là, on sait que tous les autres programmes ne s'arrêteront pas. Avec ce procédé, on a donc résolu le problème de l'arrêt en utilisant le nombre Ω ¹⁰.

Le nombre Ω n'est évidemment pas calculable sinon le programme qui le calcule nous permettrait de résoudre le problème de l'arrêt en utilisant une machine de Turing universelle¹¹. Cependant on peut très bien faire l'expérience de pensée dans laquelle ce nombre nous est transmis. Pour un mathématicien, c'est une supposition parfaitement légitime, mais là encore, on peut se demander si cette transmission est physiquement réalisable.

L'univers newtonien de la mécanique classique nous le permet : rien n'empêche de positionner sur un mètre étalon une gravure infiniment fine exactement à la position Ω . Y accéder revient donc à effectuer une mesure arbitrairement précise de la position, ce qui est en principe permis dans le cadre newtonien.

10. Le lecteur attentif aura remarqué que le nombre de Turing introduit précédemment fournit également la réponse au même problème mais il faut en connaître les 2^N premiers chiffres pour obtenir la liste des programmes de N bits ou moins qui s'arrêtent. Le nombre Ω est en fait une version compressée, au sens de la théorie de l'information, du nombre de Turing.

11. Tout comme le nombre de Turing. Cependant, contrairement à ce dernier, on peut en plus démontrer qu'il est impossible de connaître plus d'un nombre fini de décimales de Ω avec certitude. Pour en savoir plus, voir [124, 123].

Certes, il peut y avoir des difficultés techniques liées au bruit thermique sur les appareils de mesure, mais, sur le plan *théorique*, c'est parfaitement envisageable.

Cependant, cela n'est précisément pas possible dans l'univers quantique. Dans un monde quantique, on ne peut pas accéder à des échelles spatiales arbitrairement petites : pour résoudre des distances plus petites que l , il faut utiliser un rayonnement de longueur d'onde plus petite que l . Dans ce cas, on concentre l'énergie hc/l de ce rayonnement. À budget en énergie fini, il est donc fondamentalement impossible d'accéder à une mesure arbitrairement précise de position. La situation est en fait pire que cela : en utilisant un rayonnement de plus en plus énergétique, on concentre une masse $h/c\lambda$ dans une taille l . Lorsque l'on diminue l , il y a donc concentration d'une masse de plus en plus grande dans un rayon l . Or lorsque cette masse dépasse $M_S = lc^2/2G$, on crée un trou noir, une région de l'espace dont il est impossible (dans le cadre de la relativité générale) d'extraire de l'information. Il est donc impossible d'explorer des échelles spatiales en deçà de la distance l_P , appelée longueur de Planck et donnée par $h/c\lambda_P = l_P c^2/2G$, soit

$$l_P = \sqrt{\frac{2Gh}{c^3}}. \quad (3.9)$$

Pour la même raison, l'exploration d'échelles de temps en deçà de l'échelle de Planck $\tau_P = cl_P \simeq 10^{-43}$ s ne semble pas possible du fait d'effets liés à la coexistence de la gravité et de la mécanique quantique.

La machine de Zénon Les machines de Turing que nous avons définies jusqu'à présent ont un temps de calcul qui dépend linéairement du nombre de pas de calcul effectués. Inspirés par Zénon, nous pouvons imaginer une machine de Turing dont le temps de calcul serait divisé par deux à chaque pas. Le premier pas mettrait une seconde, le second une demi-seconde, le troisième un quart de seconde, etc. Quel que soit le calcul demandé, cette machine serait arrêtée au bout de deux secondes. À partir de là, il est très simple de résoudre le problème de l'arrêt : il suffit de lancer la machine et de regarder au bout de deux secondes si le nombre de pas de calcul réalisé est fini ou non. Si non, c'est que la machine ne s'arrêtait pas.

D'un point de vue conceptuel, la machine de Zénon est parfaitement valable. Cependant, la construire pose quelques problèmes comme nous allons le voir. Une littérature considérable existe sur ces questions et la rapide discussion ci-dessous ne prétend pas en faire une synthèse exhaustive. Néanmoins, le raisonnement que nous allons présenter suggère fortement qu'il serait impossible de construire une machine de Zénon dans un univers régi par les lois de la mécanique quantique.

Commençons déjà par raisonner dans un cadre classique. Une manière de construire une machine de Zénon consiste à prendre un calculateur habituel comme une machine de Turing ou un smartphone et à le faire fonctionner à

des vitesses de plus en plus rapides. Si on imagine une machine électrique, cela veut dire travailler avec des fréquences de plus en plus élevées. Comme précédemment, la mécanique quantique invalide cette possibilité, car utiliser des fréquences de plus en plus élevées revient à utiliser des photons d'énergie hf de plus en plus élevée et donc un budget énergétique divergent.

Dans un monde classique, cela n'est pas un problème car l'énergie dépend de l'amplitude de l'onde et peut être rendue arbitrairement faible. Cependant, la perte par rayonnement lié au mouvement accéléré des charges dégagerait une énergie de plus en plus grande¹², ce qui veut dire qu'une telle machine aurait un budget en énergie infini. Pour compenser cela, il faudrait donc diminuer l'amplitude des courants électriques et donc des champs électromagnétiques au sein de la machine de Zénon. On se heurterait alors au bruit thermique classique de Johnson-Nyquist. Le fonctionnement de cette machine nécessiterait de pouvoir la refroidir à une température aussi basse que l'on veut mais nous rencontrerions des problèmes avec les lois de la thermodynamique ! En effet, la troisième loi de la thermodynamique introduite par Nernst nous dit qu'il est impossible de réduire la température d'un système physique au zéro absolu en un nombre fini d'étapes. De plus, chaque étape de refroidissement prendrait de plus en plus de temps au fur et à mesure qu'on se rapprocherait du zéro absolu. Du coup, même dans un monde classique, s'affranchir du bruit thermique pour effectuer des mesures arbitrairement précises ou pour opérer la machine de Zénon de plus en plus vite avec des courants et des champs de plus en plus faibles nécessiterait probablement un temps infini.

3.3 Les thèses de Church, Turing, Landauer et Deutsch

Le contenu des sections précédentes nous a appris deux choses cruciales : en premier lieu, la seule manière de définir ce qui est calculable consiste à se donner un modèle de calcul. Ensuite, la discussion sur l'hypercalcul de la section précédente reflète notre incapacité à trouver un modèle qui soit à la fois plus puissant (au sens de la calculabilité) que la machine de Turing et physiquement réaliste.

Nous voyons donc que la question de la calculabilité, si elle peut être abordée uniquement du point de vue conceptuel de l'informatique théorique et des mathématiques, constitue un premier point de contact entre sciences physiques et sciences du calcul. Avant d'aller plus loin dans cet ouvrage, nous allons ici poser quelques points d'ancrage (les « thèses ») qui nous serviront dans toute la suite de cet ouvrage sur les relations qui lient physique, information et calcul.

12. Toute charge accélérée rayonne de manière proportionnelle à son accélération [156].

3.3.1 La thèse de Church-Turing

Le premier enseignement est une constatation propre à l'informatique théorique. Elle est formalisée par la thèse de Church-Turing qui fonde la théorie de la calculabilité que nous connaissons :

Les fonctions calculables sont les fonctions calculables par machine de Turing.

Il ne s'agit pas simplement d'une définition, puisqu'il s'agit de faire le lien entre la notion intuitive et vague de la calculabilité avec une définition formelle. La thèse de Church-Turing n'est pas un théorème non plus. C'est une thèse au sens étymologique du mot *thesis* (θέσις) qui en grec ancien signifie « poser », autrement dit une hypothèse philosophique. Ici la thèse de Church-Turing est une affirmation qui définit un domaine de l'informatique théorique.

Au niveau intuitif, elle est renforcée par plusieurs considérations. Tout d'abord, les mathématiciens ont fait plusieurs tentatives pour définir les fonctions calculables entre 1930 et 1936 : avec les machines d'Alan Turing [363] mais aussi avec le λ -calcul d'Alonzo Church [96] et les fonctions récursives de Jacques Herbrand et Kurt Gödel. Puis Stephen Cole Kleene a démontré que toutes ces définitions recouvrent la même notion mathématique [233]. D'autre part, comme nous l'avons vu plus haut, les modèles plus puissants imaginés jusqu'à présent sont physiquement irréalistes. Cela nous laisse penser que les machines de Turing nous donnent la bonne notion.

3.3.2 La thèse de Landauer

En 1961, Rolf Landauer, chercheur chez IBM, porté par des recherches sur les limites thermodynamiques mises en jeu dans un calcul, et plus généralement sur la performance des ordinateurs (voir chapitre 12, volume II), posa lui aussi une thèse qui porte cette fois sur l'univers physique. Cette thèse pose que toute information est portée par un support physique et donc que toute manipulation d'information est réalisée par un processus physique. Alors que nous avons présenté la notion d'information selon Shannon avec un point de vue purement mathématique en utilisant la théorie des probabilités, la thèse de Landauer va plus loin et énonce que, dans l'univers physique, l'information est nécessairement traitée par des systèmes physiques. La thèse de Landauer est développée dans sa publication [246] mais on retiendra la formulation choc :

Toute information est portée par des objets physiques.

Cette thèse ainsi que les observations que nous avons faites dans la section 3.2.4 conduisent naturellement à une question de physique à laquelle, *a priori*, l'informatique théorique ou les mathématiques n'ont pas vocation à répondre : quels sont les modèles de calcul physiquement réalisables dans notre univers ? Cette question revient à se demander quelles sont les limites physiques à la calculabilité. En ce sens, elle prolonge la démarche commencée avec Carnot [83]

et la thermodynamique qui fut la première théorie physique à poser des limites autres que des lois de conservations à des processus physiques.

3.3.3 La thèse de Church-Turing-Deutsch

On peut cependant renverser les rôles et adopter le point de vue d'un informaticien qui se demanderait alors ce que la puissance des différents modèles de calcul apporte en termes de compréhension de l'univers physique. La notion clé dans ce point de vue est celle de simulation discutée par David Deutsch [128]. Si on dispose d'un modèle de calcul, défini par une machine universelle M , on dira que cette machine simule un système physique S si et seulement s'il existe un programme p_S pour M qui rend la machine M physiquement équivalente à S . Concrètement, cela veut dire qu'un physicien ne pourrait pas distinguer les jeux de données produits par la machine M des résultats de mesures effectuées sur le système S , quelle que soit l'expérience effectuée sur S . Notons que dans ce point de vue d'informaticien, on ne se demande pas comment réaliser la machine M dans le monde physique.

Avec cette définition, la question « duale » de la précédente consiste à se demander quel modèle de calcul permet de simuler l'univers physique. Pour les physiciens, la question de l'existence ne se pose pas : le projet de la physique repose précisément sur la thèse qu'il existe un modèle de calcul permettant de décrire tous les phénomènes observables dans l'univers.

La relation entre les deux questions est naturelle si on confronte les points de vue qui leur ont donné naissance : si on admet la thèse de Landauer selon laquelle l'information et le traitement de l'information sont effectués dans l'univers physique (et pas en dehors par un être omniscient échappant aux lois de la physique), alors les machines M dont nous parlons sont également des systèmes physiques.

La réunion des deux questions est donc une question sur l'auto-cohérence du modèle de calcul servant à décrire l'univers physique : quel est le modèle de calcul physiquement réalisable qui permet de simuler l'ensemble des phénomènes physiques de l'univers dans lequel il est mis en œuvre ? La réponse à cette question est précisément la thèse avancée par David Deutsch en 1985 :

Tout sous-système physique fini de l'Univers peut être parfaitement simulé par un modèle de calcul basé sur la machine de Turing.

La thèse de Church-Turing-Deutsch est donc une thèse qui cette fois porte sur l'univers physique. Elle pose que l'ensemble des phénomènes naturels se produisant dans n'importe quel sous-système fini de l'Univers correspond à ce qui est effectivement calculable sur une machine de Turing. Elle affirme également que les modèles de calcul plus puissants que le modèle basé sur la machine de Turing ne peuvent pas être réalisés dans l'univers physique dans lequel nous vivons. *Exit* donc les modèles considérés dans la section 3.2.4. De plus, comme nous l'avons vu, l'univers classique de Newton fournit un modèle de calcul beaucoup plus puissant que celui de la machine de Turing

car il donne accès à tous les nombres réels, y compris ceux qui ne sont pas calculables par une machine de Turing. La machine de Turing n'est donc pas assez puissante pour simuler le monde classique sauf sur un tout petit ensemble de données initiales correspondant à des nombres réels calculables¹³. La thèse de Church-Turing-Deutsch réfutant toute description de l'Univers par un modèle de calcul qui soit plus puissant que le modèle de calcul de Turing, elle réfute donc la description de l'Univers classique. C'est finalement une thèse extrêmement forte qui fait bien évidemment l'objet de controverses.

3.3.4 La question des ressources

De manière plus concrète, la thèse de Church-Turing-Deutsch pousse à se poser la question des ressources nécessaires pour faire un calcul. En effet, nous avons vu qu'il existe plusieurs modèles de calcul équivalents à la machine de Turing déterministe à un ruban : on peut par exemple rajouter des rubans mais aussi considérer une variante non déterministe. Cette dernière variante correspond à la même notion de calculabilité mais on comprend intuitivement qu'elle est considérablement plus efficace que la machine habituelle. Cependant, l'une et l'autre de ces machines ont pour caractéristique d'être réalisables dans un univers newtonien discret (même si c'est une abstraction, c'est par une telle image classique qu'on les définit).

Les thèses de Landauer puis Church-Turing-Deutsch nous disent que le domaine de calculabilité de la machine de Turing est identique à celui de toute machine réalisée dans un univers physique, que nous savons par l'expérience ne pas être classique mais quantique. La question naturelle qui se pose maintenant est donc celle des ressources nécessaires pour faire un calcul. Dans le cas présent, est-ce que l'univers physique, régi par les étranges lois quantiques, ne recèle pas quelque part la possibilité de machines qui permettraient de calculer la même chose qu'une machine de Turing mais plus efficacement que n'importe laquelle d'entre elles ?

Pour aborder cette question, il nous faut discuter de la question des ressources qui sont mobilisées pour faire un calcul. Cette question absolument cruciale définit ce qu'on appelle la théorie de la complexité en informatique, que nous allons aborder dans la prochaine section du présent chapitre.

Une autre question naturelle consiste à se demander s'il existe une machine physique réalisée dans notre univers et utilisant de manière essentielle les principes de la mécanique quantique qui, en principe, serait apte à simuler n'importe quel phénomène s'y déroulant. Ainsi que David Deutsch l'a montré [128], une telle machine existe : c'est l'ordinateur quantique dont nous parlerons en détail dans la seconde partie de cet ouvrage.

13. En pratique, lorsque des physiciens font des simulations numériques dans un cadre classique, ils calculent des approximations discrètes du comportement d'un système classique jusqu'à un niveau de précision inférieur aux incertitudes expérimentales. Le point que nous soulevons est plus fondamental et n'invalide absolument pas le travail qu'ils effectuent.

3.4 Complexité

Dans la section précédente nous nous sommes intéressés à la théorie de la calculabilité. Savoir si une fonction est calculable ou non est de toute première importance en pratique, puisqu'il s'agit de savoir si un problème est soluble ou non. Dans cette section, en nous intéressant à la complexité, nous allons en quelque sorte être encore plus pragmatiques. Il ne s'agit plus de savoir si on peut résoudre le problème, mais si on peut le résoudre *efficacement*. Il est en général assez peu utile de savoir qu'un problème a une solution que l'on peut calculer, s'il faut plusieurs milliards d'années pour effectuer ce calcul.

Dans tous les exemples que nous avons vus, les ressources qui sont utilisées pour le calcul sont de deux types. Il y a d'abord le temps d'exécution, qui peut simplement s'exprimer en secondes dans le cas d'une architecture réelle ou bien en nombre d'opérations élémentaires pour les modèles de calcul. Le deuxième type de ressources important est la quantité de mémoire utilisée par l'algorithme qui, dans le cas d'une machine de Turing, est le nombre de cases différentes lues sur le ruban une fois la machine arrêtée à la fin du calcul.

Ces deux ressources sont donc le temps (d'exécution) et l'espace (mémoire) dont l'algorithme va avoir besoin pour mener à bien sa tâche. Ce sont les deux ressources les plus étudiées de la théorie de la *complexité algorithmique*. Il existe cependant d'autres sortes de ressources auxquelles nous pouvons nous intéresser, comme la complexité en requêtes, que nous verrons au chapitre 8, ou la complexité de communication. Le but de la théorie de la complexité est de quantifier les ressources nécessaires pour résoudre un problème donné. La première difficulté qui saute aux yeux est qu'il faut définir des quantités indépendantes des détails de l'architecture sous-jacente. À l'aide de ces quantités, nous pourrions organiser les différents problèmes dans différentes classes qui vont nous dire si ces problèmes sont réalisables en pratique. Tout comme dans la section précédente, nous nous intéresserons avant tout aux problèmes de décision. La raison est que d'ordinaire, si un problème de décision est « facile », il en va de même pour les problèmes dont il dérive.

3.4.1 Complexité algorithmique

Avant d'aller plus loin, nous allons mettre en place le décor qui va nous permettre de poser un certain nombre de problèmes. On va supposer que l'on a un certain nombre d'objets qui possèdent chacun un poids et une valeur (voir figure 3.10).

Afin de pouvoir classer les algorithmes, il est nécessaire de définir une notion quantitative à la fois suffisamment précise et suffisamment générique pour que cette définition soit indépendante des détails de l'architecture sous-jacente. Cette notion s'appelle la complexité d'un algorithme, et elle peut concerner le temps, l'espace ou une autre ressource utilisée par l'algorithme. Afin de simplifier la discussion ici, nous ne parlerons que de complexité en temps, mais toute la discussion est parfaitement transposable en termes d'espace.

					
Valeur v	4	3	13	22	30
Poids p	2	5	17	25	30

FIGURE 3.10 – Une liste d’objets possédant chacun un poids et une valeur, qui va nous servir de base pour présenter divers problèmes algorithmiques.

Afin de définir la notion de complexité, nous allons regarder le temps de calcul de quelques opérations simples sur l’exemple que nous avons défini plus haut. Pour cela, on va introduire un modèle de calcul simpliste, qui a pour opérations élémentaires les quatre opérations arithmétiques de base, l’addition, la soustraction, la multiplication et la division¹⁴. Chacune de ces opérations a un temps associé τ_{add} , τ_{sub} , τ_{mult} et τ_{div} . De plus, on suppose que ce sont les seules opérations qui prennent du temps¹⁵. Nous allons reprendre notre exemple donné sur la figure 3.10, et on va s’intéresser au temps nécessaire pour calculer différentes quantités sur notre modèle jouet, en fonction du nombre n d’objets considérés. Par exemple, si on souhaite calculer la moyenne des valeurs, il est nécessaire d’effectuer $n - 1$ additions et une division, le temps nécessaire est donc

$$t_{\text{moy}}(n) = (n - 1)\tau_{\text{add}} + \tau_{\text{div}} = n\tau_{\text{add}} + \tau_{\text{div}} - \tau_{\text{add}}. \quad (3.10)$$

Prenons un autre exemple. Si l’on souhaite calculer la variance de l’ensemble des valeurs v , on peut se servir du fait que $\text{var}(v) = \langle v^2 \rangle - \langle v \rangle^2$. Le premier terme nécessite n multiplications, $n - 1$ additions et une division, et le second n’est rien d’autre que la moyenne multipliée par elle-même. Enfin, il faut soustraire le second terme au premier. On a donc

$$\begin{aligned} t_{\text{var}}(n) &= n\tau_{\text{mult}} + (n - 1)\tau_{\text{add}} + \tau_{\text{div}} + t_{\text{moy}}(n) + \tau_{\text{mult}} + \tau_{\text{sub}} \\ &= n(\tau_{\text{mult}} + 2\tau_{\text{add}}) + 2\tau_{\text{div}} - 2\tau_{\text{add}} + \tau_{\text{mult}} + \tau_{\text{sub}}. \end{aligned} \quad (3.11)$$

Une quantité indépendante des détails de l’architecture n’a pas de dépendance en τ_{add} , τ_{sub} , τ_{mult} et τ_{div} . Ici, on a deux termes, un d’ordre un et l’autre d’ordre

14. Étant donné que l’on travaille avec des entiers, il s’agira de la division euclidienne.

15. En fait, ce modèle n’est pas si éloigné de ce que l’on va trouver dans une architecture moderne, où les opérations arithmétiques de base sont câblées et nécessitent chacune un certain nombre de cycles. Il faut néanmoins faire attention à trois choses. Tout d’abord, pour que le temps de chacune de ses opérations soit constant, il est impératif de travailler avec des entiers de taille fixe, en général la taille des registres du processeur. D’autre part, une architecture moderne est plus complexe. Notamment, on ne peut négliger le coût des accès à la mémoire et au disque, qui sont dépendants d’un certain nombre de facteurs. Enfin, pour des raisons de simplicité nous allons copieusement négliger les surcoûts liés aux détails des mini-algorithmes que nous allons écrire.

zéro en n . Pour des opérations plus compliquées, il y a souvent des termes d'ordres supérieurs, chacun avec des facteurs de proportionnalité différents, qui dépendent de temps dictés par l'architecture. La solution est de regarder ce qui se passe de manière asymptotique quand n est grand. Dans notre cas, on a une complexité linéaire en temps, de manière asymptotique, pour le calcul de la valeur moyenne et de la variance. En utilisant la notation $\mathcal{O}$ (dite grand-O) des mathématiques, on pourra écrire ¹⁶

$$t_{\text{moy}}(n) = \mathcal{O}(n), \quad (3.12)$$

$$t_{\text{var}}(n) = \mathcal{O}(n). \quad (3.13)$$

C'est exactement la notion de complexité à laquelle nous voulions aboutir. On dira que le calcul de la moyenne et de la variance ont une complexité en temps qui est linéaire, si on regarde les opérations arithmétiques de base. On appelle cette complexité la *complexité arithmétique*, car notre ressource est le nombre d'opérations arithmétiques de base.

La première des choses que l'on peut objecter à cette définition est que la majorité des algorithmes n'ont pas une dépendance simple du nombre d'opérations élémentaires en la taille de l'entrée. Dans la figure 3.10, si on demande s'il y a un objet de poids 2, il est nécessaire de n'effectuer qu'une seule comparaison puisque cet objet est le premier de la liste. En revanche, si on demande s'il existe un objet de poids 30, il est nécessaire d'effectuer cinq comparaisons, alors que la taille de l'instance du problème est la même. Pour régler ce problème, il y a deux approches possibles. La première est de regarder la complexité « en moyenne » pour toutes les entrées d'une taille n donnée. Cette approche a l'avantage d'être pragmatique dans le sens où, en général, les instances des problèmes de la vie courante (ici, le nombre d'objets, leur poids et leur valeur) ont de bonnes chances d'être des instances « moyennes ». Ici, nous allons adopter une approche pessimiste, où on va regarder quel est le temps le plus long que l'on peut avoir pour une taille d'entrée donnée. Nous allons donc nous intéresser à la complexité dans le pire des cas. Il faut noter que cette approche pessimiste est la même que celle que nous avons adoptée dans la section précédente, lorsqu'il était question de calculabilité. C'est une approche courante en informatique.

Une seconde difficulté vient de la notion d'opération élémentaire que nous avons utilisée. En effet, celle-ci dépend encore de l'architecture que l'on considère. Dans un ordinateur, les opérations élémentaires sont celles qui sont câblées dans le processeur. L'opération élémentaire sur une machine de Turing est un cycle de lecture-écriture-déplacement. Le gros point faible de notre approche, où nous avons considéré les opérations arithmétiques comme étant

16. On a aussi $t_n = \mathcal{O}(n^p)$ pour tout $p \geq 1$. La notation $\mathcal{O}$ permet donc de trouver une « majoration » asymptotique. Dans un certain nombre de cas, il est possible d'être plus précis, à condition bien sûr de montrer que la complexité obtenue par l'algorithme est optimale. Dans la littérature, cela est formalisé par la notation Θ , où l'on a $f(n) = \Theta(g(n))$ si et seulement si $f(n) = \mathcal{O}(g(n))$ et $g(n) = \mathcal{O}(f(n))$.

des opérations de base, est que cela n'est vrai que dans la limite où les nombres manipulés restent petits. Si la somme des valeurs se met à excéder la taille des registres du processeur, par exemple, il devient faux de dire que le temps nécessaire pour une addition est constant. Étant donné que l'on souhaite regarder les régimes asymptotiques, où le nombre d'objets est grand, il va de soi que faire de telles approximations n'est pas très rigoureux. Pour éviter de tels écueils, nous allons donc nous baser sur l'architecture de la machine de Turing pour associer une complexité à un algorithme. Il faut noter néanmoins que pouvoir quantifier la complexité en termes d'une opération en particulier (une opération arithmétique, l'échange de deux éléments dans la mémoire) est très utile quand on doit comparer des algorithmes entre eux.

Enfin, il est important d'insister sur le fait que la complexité d'un algorithme est une notion asymptotique. Ce n'est pas parce qu'un algorithme a une meilleure complexité qu'un autre qu'il sera meilleur dans tous les cas. Notamment, pour les petites tailles d'entrées, il est fréquent de préférer des algorithmes simples, qui ne sont pas les meilleurs asymptotiquement, mais qui seront beaucoup plus rapides car avec des constantes multiplicatives plus faibles.

3.4.2 Classes de complexité

La notion de complexité que nous avons vue permet de s'abstraire du matériel et déjà de classer les algorithmes entre eux. Toutefois, cela n'est pas suffisant pour classer les problèmes. En fait, il y a certains algorithmes que l'on ne peut pas mettre en œuvre sans avoir le matériel adéquat. Si la calculabilité ne dépend pas du modèle choisi, en revanche la complexité en dépend. Nous allons l'illustrer ici sur les machines de Turing.

Deux versions différentes de la machine de Turing calculant la même fonction peuvent avoir des complexités différentes, selon le nombre de rubans. Pour cela, on va reprendre l'algorithme qui permet d'identifier les mots de la forme $0^a 1^a$ sachant qu'un mot est de la forme $0^a 1^b$. La machine de Turing 3.6 fait cela avec un seul ruban. Étant donné que pour une entrée valide de taille n (le pire des cas ici) on doit balayer $n/2$ fois le ruban de longueur n , la complexité est quadratique. Pour un problème de ce type, on ne pourra pas faire mieux avec un seul ruban. En revanche, si l'on a accès à deux rubans, il suffit de se servir du second ruban comme d'une pile. Au début, on écrit $\#$ sur le second ruban, comme marqueur de début. Puis on se déplace à droite à chaque fois que l'on rencontre un 0 et à gauche à chaque fois que l'on rencontre un 1. Un tel algorithme a une complexité linéaire en la taille de son entrée. Étant donné cette dépendance, on va introduire la notion de classe de complexité. Il va s'agir de regrouper certaines complexités de façon à pouvoir classer non plus les algorithmes mais les problèmes en fonction de la difficulté que l'on a à les résoudre. Cette vision plus grossière devrait donc permettre de s'abstraire des menus détails du modèle de calcul tels que le nombre de rubans.

Nous allons néanmoins introduire certaines distinctions dans les différentes variantes de machine de Turing que nous avons vues, notamment entre les modèles déterministes et les modèles non déterministes. La machine de Turing non déterministe semble radicalement différente des autres. Si cela n'a aucune influence sur ce que l'on est capable de calculer, on a l'intuition qu'il y a un certain nombre de problèmes qui seront plus faciles à calculer sur une telle machine. Afin de vérifier si cette intuition est vraie, nous allons séparer les machines déterministes et les machines non déterministes quand nous définirons les différentes classes de complexité.

Commençons donc par la première classe que nous pouvons définir, celle des problèmes pour lesquels il existe un algorithme déterministe dont le temps d'exécution est polynomial en la taille de l'entrée. Pour ces problèmes on peut toujours trouver une machine de Turing déterministe dont la complexité est en $\mathcal{O}(n^p)$ où p est un entier. Cette classe s'appelle P .

Définition 3.6 (P). Un problème appartient à la classe P (pour polynomial) s'il existe une machine de Turing déterministe qui résout ce problème et dont le nombre de pas d'exécution est polynomial en la taille de l'entrée.

On admettra que cet ensemble est véritablement une classe, c'est-à-dire que si un problème est dans P pour une certaine variante de machine de Turing déterministe, il l'est aussi pour toutes les autres variantes déterministes.

On considère en général que P est la classe des problèmes qui peuvent être résolus en pratique, sur des ordinateurs actuels, à condition de ne pas avoir recours à une source d'aléatoire¹⁷. Étant donné que la définition de la complexité est asymptotique, cela s'entend pour des tailles d'entrées importantes. La plupart des calculs que l'on fait tous les jours sont dans cette classe : tri ou recherche dans une liste, trouver le plus court chemin entre deux points, calcul d'intégrale, etc. Cela ne signifie pas pour autant que résoudre ces problèmes soit facile dans l'absolu. Tout d'abord, un problème dont la complexité posséderait un exposant et une constante élevés ne pourrait pas être résolu en pratique, même pour de petites instances. Cependant, dans la plupart des cas, on a des complexités qui sont linéaires, quadratiques ou cubiques, avec des constantes raisonnables. Enfin, même dans ces derniers cas, il est parfois nécessaire de regarder des tailles d'entrées extrêmement importantes, ce qui rend le calcul non trivial.

Afin de comparer les machines de Turing déterministes et leurs analogues non déterministes, nous pouvons dès à présent construire l'analogue de la classe P pour des machines de Turing non déterministes.

Définition 3.7 (NP). Un problème appartient à la classe NP (pour polynomial non déterministe) s'il existe une machine de Turing non déterministe qui résout ce problème et dont le nombre de pas d'exécution est polynomial en la taille de l'entrée.

17. Nous reviendrons sur les apports de l'aléatoire au calcul à la section 3.4.4.

Un exemple célèbre de problème dans NP est le problème du sac-à-dos. Dans ce problème, nous avons une liste d'objets avec des poids et des valeurs, similaire à celle de la figure 3.10. De plus, on a un sac-à-dos qui peut porter au plus un certain poids. La question posée est alors « peut-on trouver une combinaison d'objets ayant au moins la valeur v qui rentre dans le sac-à-dos ? ». En partant de là, il n'est pas évident de voir que ce problème est dans NP. Cependant, ce problème a une structure caractéristique : si on se donne la liste des objets que l'on choisit de mettre dans le sac-à-dos (on parlera d'une configuration), l'algorithme qui permet de vérifier si ces objets ont un poids inférieur à celle que le sac peut supporter et une valeur supérieure à v est dans P. C'est cette structure qui va nous permettre de trouver un algorithme polynomial sur une machine de Turing non déterministe. Il suffit pour cela de générer l'ensemble des configurations (dont le nombre est 2^n pour n objets) et de les tester. Générer l'ensemble des configurations est en réalité très facile sur une machine non déterministe. En un pas de temps, la machine de Turing non déterministe peut avoir deux transitions possibles : il est possible d'écrire 0 puis de se décaler vers la droite, ou bien d'écrire 1 et de se décaler vers la droite. En effectuant cette transition n fois, on explore, au sens de la machine de Turing non déterministe, l'ensemble des 2^n mots de taille n .

À partir d'un mot de taille n , on peut vérifier que la configuration correspondante est une solution du problème en temps polynomial. La machine de Turing non déterministe peut effectuer cette vérification pour chacun des mots possibles « en parallèle ». S'il existe une solution valide, alors par définition la machine accepte et, dans le cas contraire, la machine refuse. Tout s'est donc déroulé en temps polynomial, en utilisant le déploiement des 2^n branches de calcul lors de la génération de l'ensemble des configurations possibles du sac-à-dos.

En fait, la structure que nous avons exploitée est celle de tous les problèmes NP. Une autre façon de définir les problèmes NP est de dire qu'il s'agit des problèmes dont on peut vérifier en temps polynomial si une solution potentielle est valide ou non. Cela signifie qu'un problème est dans NP s'il existe une machine (que l'on appelle un *vérifieur*) qui, étant donné une instance et une de ses potentielles solutions (de longueur polynomiale en la taille de l'instance), est capable de vérifier en temps polynomial si la solution est valide ou non. Une solution valide est appelée un *certificat* et est analogue à la « configuration » du sac-à-dos. Le certificat n'existe que si la réponse au problème est « oui » pour l'instance considérée.

Cette équivalence est illustrée sur la figure 3.11. Si l'on considère un problème dont la réponse est donnée en temps polynomial par une machine de Turing non déterministe, il est possible d'étiqueter le chemin d'exécution de la machine non déterministe par un mot de taille polynomiale. Ce chemin est notre certificat, que l'on peut vérifier de manière déterministe en temps polynomial. Inversement, si on a un vérifieur polynomial, on peut construire une machine de Turing non déterministe qui s'exécute en temps polynomial. Pour

sommet une seule fois, les arêtes du graphe portant la distance entre deux sommets. La version décisionnelle, qui revient à se demander s'il existe un chemin plus court qu'une certaine distance D , est un problème NP : étant donné un trajet, il est facile (complexité linéaire) de vérifier que ce trajet passe par l'ensemble des sommets et parcourt une distance inférieure à D .

Satisfiabilité d'un circuit Dans la section 3.1.1, nous avons vu qu'il était possible de décrire n'importe quelle fonction de $\{0, 1\}^n \rightarrow \{0, 1\}$ comme un circuit composé de portes logiques **or**, **and** et **not**. On peut se demander si, pour un tel circuit, il existe une entrée telle que la sortie est 1. Vérifier qu'une entrée donne 1 est facile, car il suffit de simuler le calcul du circuit sur l'entrée par une machine de Turing déterministe, ce que l'on peut faire en temps polynomial en la taille du circuit. Un problème proche, nommé SAT (pour *satisfiability*), qui consiste à vérifier la satisfiabilité d'une formule booléenne est également NP. Il en va de même pour le problème 3SAT, dans lequel la formule booléenne est une conjonction (ensemble de clauses séparées par des **and**) de clauses, chacune de ces clauses étant une disjonction de trois variables, éventuellement niées (séparées donc par des **or**). Par exemple, déterminer s'il existe $(X_1, \dots, X_5)$ tel que

$$(X_1 \vee X_2 \vee \neg X_3) \wedge (\neg X_1 \vee X_4 \vee \neg X_5) \wedge (\neg X_2 \vee X_3 \vee X_5) \quad (3.14)$$

est une instance du problème 3SAT.

Factorisation Le théorème fondamental de l'arithmétique nous assure que tout nombre entier n a une décomposition en produit de facteurs premiers. Il est alors naturel de se demander quelle est la liste des facteurs premiers, avec leur multiplicité, de n'importe quel nombre entier. La version décisionnelle de ce problème revient à se demander s'il existe un facteur premier entre 0 et $m < n$. Si on se donne la liste des facteurs, il suffit de les multiplier pour retrouver le nombre, ce qui est polynomial en la taille du nombre en binaire (qui est $\log_2 n$).

☞ NP et au-delà

La classe NP peut être définie comme la classe des problèmes pour lesquels on peut fournir une preuve d'appartenance vérifiable en temps polynomial. Un problème X est dans NP si et seulement s'il existe un problème Y dans P et un polynôme p tels que pour tout mot d'entrée $x \in \{0, 1\}^*$.

$$x \in X \Leftrightarrow (\exists y \in \{0, 1\}^{p(|x|)}, (x, y) \in Y) . \quad (3.15)$$

Ici, un y satisfaisant $(x, y) \in Y$ est une preuve d'appartenance de x à X . Cette preuve n'est pas trop longue (sa longueur est bornée par un polynôme

p en la longueur de x) et elle peut être vérifiée en temps polynomial par un algorithme résolvant Y .

Dans le cas de la primalité par exemple, x est le nombre que nous cherchons à décomposer en facteurs premiers et y l'un de ses facteurs. Le problème Y est l'ensemble des couples (x, y) tels que x soit divisible par y pour un y non trivial.

On peut alors voir une machine de Turing non déterministe polynomiale comme une machine qui détermine en temps polynomial s'il existe un certificat pour une certaine propriété d'un mot ou d'un entier. En revanche, peut-on, avec une machine de Turing non déterministe seule, répondre aux questions de la forme suivante ?

$$x \in X \Leftrightarrow \left(\forall y \in \{0, 1\}^{p(|x|)}, (x, y) \in Y \right). \quad (3.16)$$

Décider cette propriété revient à décider sa négation, c'est-à-dire à trouver un y tel que $(x, y) \notin Y$: c'est un certificat de réfutation de $x \in X$. On dit que ce problème X appartient à la classe **coNP** : c'est la classe des problèmes dont le complémentaire est **NP**.

L'intersection de **NP** et **coNP** est non vide, et contient **P**. On peut s'intéresser par exemple au test de primalité. Celui-ci est dans **coNP** : si un nombre n'est pas premier, un certificat de réfutation est par exemple un de ses facteurs non triviaux. En 1975, Pratt a montré que ce problème était également dans **NP** [303], en démontrant qu'il existe un certificat de primalité de taille polynomiale. Enfin, Agrawal, Kayal et Saxena ont montré en 2002 que la primalité est dans **P** [8].

En revanche on ne sait pas si **NP** et **coNP** sont égales ou différentes. L'hypothèse la plus probable à l'heure actuelle est que **NP** et **coNP** sont différentes. Dans ce cas, soulignons l'étrangeté de ces classes, notamment par rapport à la classe **P**. Une propriété importante de la classe **P** est que celle-ci est stable par sous-programme. Si une machine calculant en temps polynomial appelle un nombre polynomial de fois un programme qui s'exécute en temps polynomial, le programme obtenu reste dans la classe **P**. D'un point de vue programmation, cela à dire que si une fonction fait un nombre polynomial d'appels à une autre fonction qui elle même s'exécute en temps polynomial, alors le programme résultant s'exécutera aussi en temps polynomial. En théorie de la complexité, la machine appelée est dénommée *oracle* : il s'agit d'une boîte noire ayant certaines propriétés (ici, de résoudre les problèmes d'une classe donnée).

Si une machine de Turing non déterministe peut faire appel à une autre machine de Turing non déterministe, alors la machine appelante peut résoudre les problèmes **NP** en temps polynomial, mais également les problèmes **coNP**. Pour cela, il lui suffit d'inverser la réponse donnée par l'oracle. Cela implique que si **NP** $\neq$ **coNP**, alors la classe **NP** n'est pas stable par sous-programme.

Nous n'avons pour l'instant pas utilisé le non-déterminisme de la machine appelante. En l'utilisant, la machine appelante peut alors générer un ensemble

de configurations en utilisant le non-déterminisme. Elle peut passer chacune de ces configurations à l'oracle, de la même façon que ce que nous avons obtenu précédemment. Cet emboîtement de machines non déterministes permet alors de résoudre l'ensemble des problèmes où l'on imbrique un quantificateur « $\exists$ » et un quantificateur « $\forall$ » :

$$x \in X \Leftrightarrow \left(\exists y_2 \in \{0, 1\}^{p_2(|x|)}, \forall y_1 \in \{0, 1\}^{p_1(|x|)}, (x, y_1, y_2) \in Y \right). \quad (3.17)$$

On peut alors étendre cette construction, dans le cas où on autorise chaque oracle à appeler un autre oracle. Si on a n appels emboîtés, on a alors une alternance de n quantificateurs, commençant par le quantificateur « $\exists$ ». On peut donc construire une hiérarchie de classes de complexité. L'union de l'ensemble de ces classes, avec l'ensemble des possibilités pour l'alternance des quantificateurs, s'appelle la hiérarchie polynomiale, notée PH [269].

Réduction et complétude

Avant d'examiner plus avant la relation entre la classe P et la classe NP, nous allons essayer de voir si on ne peut pas introduire des distinctions dans la classe NP elle-même. Grâce à la notion de réduction, nous allons voir qu'il est possible de définir un sous-ensemble de NP qui est, en un certain sens, l'ensemble des problèmes NP les plus difficiles à résoudre.

Une action relativement courante, lorsque l'on a un problème, est de le transformer en un autre problème que l'on connaît mieux. On appelle l'action de transformer un problème A en un problème B une réduction et on dit que A se réduit à B . Plus précisément, c'est une fonction f qui à un mot x associe $f(x)$ tel que x appartient à A si et seulement si $f(x)$ appartient à B .

Pour peu que cette transformation soit efficace, si A se réduit à B alors B est au moins aussi dur que A . L'idée est que si l'on sait résoudre efficacement le problème B et que l'on peut transformer A en B de manière tout aussi efficace, alors on sait résoudre efficacement A . Réciproquement, ne pas savoir résoudre efficacement A implique de ne pas savoir résoudre efficacement B si on sait réduire efficacement A à B .

Toute la subtilité se cache dans ce que l'on entend par « efficacement ». Nous avons vu qu'en général ce qui est calculable en pratique ce sont les problèmes P. Nous allons donc nous intéresser aux cas où la réduction s'effectue en temps polynomial, sur une machine de Turing déterministe. Une première chose que l'on peut remarquer, c'est que si A se réduit polynomialement à B et que B est polynomial, alors A est également polynomial. De cette façon, nous allons pouvoir classer les problèmes dans NP. Pour cela, nous allons introduire les problèmes NP-durs. Il s'agit de problèmes qui sont liés par réduction polynomiale à tous les problèmes de la classe NP. En d'autres termes, si on sait résoudre efficacement un problème NP-dur, alors on sait résoudre efficacement n'importe quel problème de la classe NP. Pour cela, il suffit de prendre notre problème NP, de le traduire en temps polynomial en notre problème NP-dur,

et de résoudre le problème NP-dur. Les problèmes NP-durs sont donc au moins aussi durs que le problème NP le plus difficile.

Définition 3.8 (NP-dur). Un problème appartient à la classe NP-dur si n'importe quel problème NP peut être réduit de façon polynomiale à celui-ci.

Nous ne démontrerons pas ici l'existence d'un problème NP-dur. La difficulté de la démonstration vient du fait qu'il est nécessaire de réduire n'importe quel problème NP au problème que l'on a trouvé et que l'on pense NP-dur. Néanmoins, une fois que l'on a trouvé un problème NP-dur, il suffit de montrer que celui-ci se réduit à un autre problème pour montrer que cet autre problème est lui-même NP-dur. L'astuce consiste à encoder les machines de Turing dans un problème en particulier. En l'occurrence, le théorème de Cook-Levin [108] utilise le problème de la satisfiabilité des circuits pour cela, et ainsi montre que ce problème est NP-dur.

Étant donné leur définition, les problèmes NP-durs sont très nombreux : la seule contrainte est qu'ils ne peuvent être plus « faciles » que les problèmes NP. Ce qui rend ces problèmes aussi intéressants est le fait qu'il existe des problèmes à la fois NP et NP-durs. De tels problèmes sont dits NP-complets.

Définition 3.9 (NP-complet). Un problème appartient à la classe NP-complet s'il est à la fois dans NP et dans NP-dur.

De fait, les problèmes NP-complets sont les problèmes NP les plus durs. Si on sait résoudre un problème NP-complet de manière efficace (soit parce qu'on a une architecture adaptée à ce problème, soit parce qu'on a trouvé un algorithme efficace sur une architecture conventionnelle), alors on sait résoudre efficacement n'importe quel problème NP. Le problème du sac-à-dos que nous avons abordé est également un problème NP-complet. C'est également le cas de la satisfiabilité des circuits, et du voyageur de commerce [170].

3.4.3 P *versus* NP

Nous allons maintenant nous intéresser à ce qui est sans doute la plus grande question de la théorie de la complexité, la relation entre P et NP. Nous avons vu dans la section 3.2.2 qu'une machine déterministe n'était qu'un cas particulier de machine non déterministe. En d'autres termes, tout problème dans P est également dans NP. Cela signifie donc que

$$P \subseteq NP. \quad (3.18)$$

Se pose alors la question réciproque de savoir si NP est inclus dans P. Cela revient à se demander si, en pratique, les problèmes NP sont aussi faciles à résoudre que les problèmes P. Une autre façon de poser cette question est celle de savoir si les machines de Turing non déterministes sont plus puissantes que les machines déterministes. Pour illustrer la difficulté de résoudre un problème NP, essayons de le faire « naïvement » en partant des deux définitions

que nous avons pour un problème NP. La première définition utilise la notion de machine de Turing non déterministe. Nous pouvons déterminer cette machine à l'aide de la procédure que nous avons vue dans la section 3.2.2 : la simulation de toutes les branches du calcul par une machine déterministe. Comme le nombre de ces branches est *a priori* une fonction exponentielle du temps de calcul de la machine non déterministe, cela nous donne un algorithme déterministe mais en temps exponentiel pour résoudre notre problème. Voyons si la deuxième définition que nous avons donnée de NP nous permet d'obtenir un algorithme déterministe plus efficace. Dans ce cas, il existe une machine de Turing déterministe, le vérifieur, qui nous permet de décider si un mot est une solution de notre instance, et donc si cette instance appartient au problème. Pour résoudre le problème, il suffit donc de tester tous les mots pour voir s'ils sont une solution. Mais le nombre de ces mots est exponentiel, et nous obtenons encore une fois un algorithme en temps exponentiel pour résoudre notre problème. En utilisant ces approches naïves, il est extrêmement difficile de résoudre, en pratique, n'importe quel problème NP, même sur des tailles d'entrées raisonnables. La véritable puissance du non-déterminisme est de rendre facile la recherche d'une solution d'un problème, si sa vérification est également facile.

Ainsi, si l'on part de l'exemple de la figure 3.10, on peut essayer de construire naïvement des solutions. On va se demander s'il existe une configuration du sac-à-dos ayant une valeur d'au moins 38. La réponse est oui, car si on met le second, le troisième et le dernier objet, on a une valeur de 39 pour un poids inférieur à 45. Peut-on arriver facilement à ce résultat ? En fait, un exemple d'algorithme naïf consisterait à trier la liste selon un certain critère et à essayer de faire rentrer des objets dans le sac jusqu'à ce qu'il soit rempli. Dans notre exemple, on peut essayer de trier les objets par poids, par valeur ou bien par densité de valeur, et on se rend compte que l'on n'arrive jamais à remplir correctement le sac sans repartir en arrière. On est obligé d'enlever des objets et d'essayer d'autres combinaisons. Cela revient à épuiser les combinaisons possibles jusqu'à ce que l'on en trouve une qui corresponde. Cet argument cependant ne montre pas que le problème est intrinsèquement difficile, juste qu'un algorithme polynomial n'est pas aussi trivial.

À l'heure actuelle, on ne sait toujours pas si $P = NP$. C'est d'ailleurs un des sept problèmes de mathématiques (et le seul en informatique !) qui peut rapporter un million de dollars à la personne qui le résoudrait¹⁸. Le lecteur intéressé par les différentes voies explorées pour résoudre ce problème, les progrès effectués, les espoirs obtenus, souvent suivis de déceptions, pourra se référer à [2].

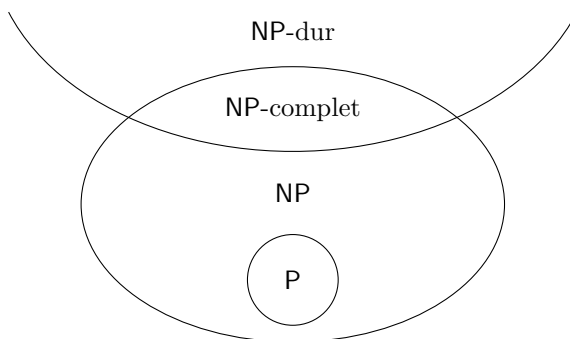
Autrement dit, on ne sait toujours pas si cette puissance est intrinsèque au non-déterminisme ou bien si nous n'avons pas été assez malins pour trouver un algorithme polynomial pour chacun des problèmes NP. Trouver un algorithme pour chacun des problèmes d'une classe semble être une tâche absolument

18. <http://www.claymath.org/millennium-problems/p-vs-np-problem>

titanesque, mais en fait il suffit de trouver un algorithme polynomial pour n'importe lequel des problèmes NP-complets, cela entraînant la chute de tous les problèmes NP. En revanche, un résultat important, le théorème de Ladner, montre que si $P \neq NP$, alors il existe des problèmes intermédiaires, c'est-à-dire qui ne sont ni P ni NP-complets [243]. Le problème de la factorisation est sans doute un bon candidat : actuellement, on ne connaît aucun algorithme polynomial, et on n'a toujours pas réussi à montrer qu'il était NP-complet. La relation entre les différentes classes que nous avons vues est résumée sur la figure 3.12.

L'intuition que les machines de Turing non déterministes sont plus puissantes que leurs variantes déterministes reste donc au stade de l'intuition. Cependant, il existe un faisceau d'indices qui pousse les gens à penser que $P \neq NP$, et que le non-déterminisme apporte réellement quelque chose. Le premier indice est que, malgré des années de recherches, personne n'a encore trouvé d'algorithme polynomial pour les problèmes NP-complet. Cependant, ce n'est qu'un indice assez mince, étant donné que la théorie de la complexité est relativement jeune. Comme trouver une preuve directe se montre extrêmement difficile, les gens se sont donc penchés sur les conséquences de l'hypothèse $P = NP$. En fait, les quelques classes de complexité que nous avons vues (et les quelques autres que nous allons voir par la suite) ne sont qu'une toute petite partie de toutes les classes que l'on peut définir. L'égalité $P = NP$ impliquerait qu'un certain nombre de classes de complexité, que l'on pense distinctes, sont en fait égales. C'est notamment le cas de l'ensemble de la hiérarchie polynomiale, qui s'effondrerait sur la classe P.

Enfin, on peut se pencher sur les conséquences philosophiques d'une telle égalité. Si on associe P aux problèmes faciles, les problèmes NP sont ceux pour lesquels il est facile de vérifier si une solution donnée est juste ou fausse. Dans le cas où $P = NP$, il devient facile de trouver une solution pour peu que sa vérification soit simple. En dehors du fait qu'une bonne partie du monde s'effondrerait (les protocoles de chiffrement et de paiement sont basés sur l'hypothèse que $P \neq NP$), faire un sudoku deviendrait aussi facile que de vérifier si sa solution est correcte. Mais surtout, comme le fit remarquer Gödel à von Neumann [179], trouver une preuve de mathématiques deviendrait aussi facile que de la vérifier. De même, identifier des modèles à partir de grands jeux de données expérimentales serait aussi facile que de vérifier que les prédictions d'un modèle sont conformes aux données. Tout problème d'optimisation se ramenant à un problème NP-complet pourrait être facilement résolu avec les conséquences que l'on imagine dans le domaine de l'ingénierie ou de l'économie. En clair, l'égalité $P = NP$ implique que quasiment tout ce qui est considéré comme un processus créatif, c'est-à-dire un saut important dans la connaissance, pourrait être réalisé facilement par un ordinateur [378]. C'est peut-être la conviction la plus forte, bien que la moins mathématique, que l'on puisse avoir sur le fait que $P \neq NP$.

FIGURE 3.12 – Les relations entre les classes de complexité si $P \neq NP$.

Une dernière possibilité est que la question $P = NP$ soit indécidable, comme l'est par exemple l'hypothèse du continu. Il existerait alors un monde mathématique dans lequel l'égalité est vraie et un monde mathématique dans lequel elle est fausse...

3.4.4 L'aléatoire ou le rôle du hasard en complexité

Jusqu'à présent, nous nous sommes contentés d'algorithmes n'ayant pas recours à l'aléatoire. Pourtant pour décrire un système physique, l'aléatoire est omniprésent. On le retrouve dans le bruit des appareils de mesure, dans la description du mouvement brownien, ainsi qu'en physique statistique. On peut alors utiliser cet aléatoire comme ressource pour le calcul. C'est ce que nous allons faire ici, en introduisant le modèle de la machine de Turing probabiliste. Tout comme précédemment, nous allons voir quelles sont les classes raisonnables pour une telle machine.

Machines de Turing probabilistes

Pour introduire le modèle de la machine de Turing probabiliste, revenons à la machine de Turing non déterministe. Nous l'avons introduite comme un modèle mathématique. Ce dernier est un modèle formel, qui ne répond pas à la question suivante : comment interpréter le calcul d'une machine de Turing non déterministe ? On peut naturellement la voir comme une machine qui sait choisir le bon calcul parmi tous les calculs possibles. Mais on peut aussi considérer que la machine fait des choix aléatoires avec une probabilité $1/2$ ¹⁹.

19. De façon plus générique, on peut supposer que l'on associe à chaque transition une probabilité, sous réserve qu'elle soit calculable de manière efficace. Chaque probabilité de transition est un réel dont on peut calculer une approximation à 2^{-n} près en temps polynomial en n . Sans cette précaution, on pourrait cacher de la puissance de calcul dans les probabilités de transition.

Dans cette interprétation, la machine de Turing non déterministe est chanceuse à chaque transition.

Un modèle probabiliste réaliste n'aura pas une telle chance. Cela implique notamment qu'une machine probabiliste se trompera parfois. Il est donc nécessaire de fixer un seuil à la probabilité de se tromper, et ce de façon à ce que nous soyons raisonnablement sûrs de la réponse en répétant le calcul.

Une première idée consiste à dire qu'on peut résoudre efficacement un problème avec de l'aléatoire si on peut trouver une machine de Turing probabiliste qui se trompe avec une probabilité strictement inférieure à $1/2$. Plus précisément, si la réponse au problème est oui, la machine accepte avec une probabilité supérieure à $1/2$ et si la réponse au problème est non, elle rejette avec une probabilité supérieure à $1/2$.

Définition 3.10 (PP). Un problème appartient à la classe PP (pour *probabilistic polynomial*) s'il existe une machine de Turing probabiliste ayant un temps d'exécution polynomial qui accepte avec une probabilité supérieure à $1/2$ si la réponse est « oui » et avec une probabilité strictement inférieure à $1/2$ si la réponse est « non ».

Le problème canonique de la classe PP est de savoir si un circuit accepte au moins la moitié de ses entrées. De plus, ce problème est complet pour la classe PP.

La classe PP est en fait extrêmement large. Elle contient notamment l'ensemble des problèmes NP. En effet, si un problème est dans NP, il suffit d'essayer un certificat potentiel, pris au hasard. Si c'est vraiment un certificat, la machine accepte, sinon elle répond « non » avec une probabilité $1/2 + 2^{-2p(n)}$, où $p(n)$ est la taille d'un certificat potentiel. Étant donné qu'il y a au plus $2^{p(n)}$ certificats potentiels, la probabilité de répondre correctement est strictement supérieure à $1/2$.

Pour cette raison, la classe PP est trop large par rapport à notre intuition de ce qu'est un problème calculable en temps polynomial à l'aide de l'aléatoire. Si la probabilité d'accepter un mot est très proche de $1/2$, il y a de grands risques que quelques simulations de plus nous fassent basculer de l'autre côté de ce seuil. Dans ces conditions, comment savoir quand nous devons accepter un tel mot en pratique ?

Plus formellement, si la probabilité de donner la bonne réponse est en $1/2 + \varepsilon$, où ε est un nombre positif assez petit, pour obtenir une probabilité donnée (par exemple $2/3$) d'obtenir la bonne réponse, il va falloir lancer la machine un nombre de fois proportionnel à $1/\varepsilon$. Une machine dont la probabilité de réussite dépendrait de la taille n de l'entrée en $1/2 + 2^{-n}$ nécessiterait donc de répéter un nombre de fois exponentiel le calcul.

Ce problème peut être aisément résolu, en imposant que la machine de Turing probabiliste réponde juste avec une probabilité fixée (ne dépendant donc pas de la taille de l'entrée) strictement supérieure à $1/2$. Cela nous donne la classe BPP.

Définition 3.11 (BPP). Un problème appartient à la classe BPP (pour *bounded-error probabilistic polynomial*) s'il existe une machine de Turing probabiliste dont le temps d'exécution est polynomial en la taille de l'entrée et qui accepte avec une probabilité supérieure à $2/3$ si la réponse est « oui » et avec une probabilité inférieure à $1/3$ si la réponse est « non ».

À partir de cette définition, il est possible de trouver un algorithme polynomial qui donne la bonne réponse avec une probabilité p aussi proche de 1 que l'on veut. Il suffit de lancer un certain nombre de fois la machine de Turing et de répondre comme la majorité d'entre elles (afin d'éviter les problèmes en cas d'égalité, on lancera la machine un nombre impair de fois). Ainsi, avec 9 lancers, $p > 0,85$, avec 49 lancers $p > 0,99$ et avec 99 lancers, $p > 0,9995$. On peut montrer que p s'approche de 1 de façon exponentielle, donc il est relativement facile de gagner en précision. Par ailleurs, notons que la fraction $2/3$ dans cette définition est arbitraire. Tant que la probabilité de trouver la bonne réponse est strictement supérieure à $1/2$ et que celle-ci ne dépend pas de la taille de l'instance considérée, alors on peut se ramener à la définition que nous avons donnée en relançant la machine un nombre fixe de fois.

Le test de Miller-Rabin

Il n'est pas évident de voir quels problèmes peuvent bénéficier de l'utilisation de l'aléatoire. Afin d'illustrer cela, nous allons revenir sur le problème de la primalité. Nous allons donner les ingrédients derrière le test de Miller-Rabin, qui est un test de primalité en temps polynomial utilisant l'aléatoire, et qui permet donc de classer le problème de la primalité dans BPP.

La façon naïve, déterministe, de tester la primalité d'un nombre k consiste à effectuer la division euclidienne de ce nombre par l'ensemble des nombres compris entre 2 et $\sqrt{k}$. Il faut donc faire de l'ordre de $\sqrt{k}$ divisions euclidiennes, ce qui implique que la complexité est polynomiale en $\sqrt{k}$. Pour autant, cet algorithme n'est pas polynomial, car la taille n du nombre k est $n = \log_2 k$.

Pour faire mieux, on va se baser sur le petit théorème de Fermat. Celui-ci dit que pour un nombre premier p et pour tout entier a tel que $1 < a < p$, on a

$$a^{p-1} \equiv 1 \pmod{p}. \quad (3.19)$$

Une première idée est alors de choisir un nombre a au hasard, de tester s'il est premier avec k et de regarder si $a^{k-1} \equiv 1 \pmod{k}$. Ce test est appelé test de primalité de Fermat. Si on a $a^{k-1} \not\equiv 1 \pmod{p}$, alors on peut conclure que k est composé : a est appelé un témoin de Fermat du nombre k . Afin d'avoir un test concluant, au niveau probabiliste, il faut que pour tout nombre composé, il existe suffisamment de témoins de Fermat. Si c'est le cas, en « interrogeant » suffisamment de témoins, on peut alors être raisonnablement sûr que le nombre k est premier ou non. Cependant, le test de Fermat ne permet pas cela : certains nombres composés, appelés nombres de Carmichael,

vérifient également le théorème de Fermat. Il est donc nécessaire de raffiner le test.

Pour cela on utilise le fait que si p est premier, alors l'équation $x^2 \equiv 1 \pmod{p}$ a deux solutions possibles :

$$\begin{cases} x \equiv 1 \pmod{p}, \\ x \equiv -1 \pmod{p}. \end{cases} \quad (3.20)$$

Étant donné que l'on souhaite vérifier si k est premier, on peut ne considérer que les cas où k est impair. Dans ce cas on peut écrire $k-1 = 2^r d$. En prenant successivement la racine de l'équation $a^{k-1} \equiv 1 \pmod{k}$, on montre que l'on a l'une des deux possibilités suivantes, si k est premier :

$$\begin{cases} a^d \equiv 1 \pmod{k} & \text{ou,} \\ a^{2^s d} \equiv -1 \pmod{k} & \text{avec } 0 \leq s < r. \end{cases} \quad (3.21)$$

Ce test de primalité est appelé test de Miller-Rabin. Si la propriété ci-dessus n'est pas valide pour l'entier k , alors on dit que a est un témoin de Miller [17, 271]. Contrairement au test de Fermat, si k est composé, alors non seulement il existe un témoin de Miller mais le théorème de Rabin [306] assure que si k est composé, la proportion de témoins menteurs est inférieure à $1/4$.

On peut donc construire un algorithme qui tire le nombre a au hasard et répond que k est composé si a est un témoin de Miller, et qu'il est premier sinon. Cet algorithme permet de montrer que le problème de la primalité est dans BPP. D'une part, les opérations nécessaires (exponentiation modulaire et division euclidienne) sont polynomiales en la taille de k . De plus, l'algorithme ne renvoie jamais de résultat faux si k est premier, et la probabilité que le résultat soit faux s'il est composé est inférieure à $1/4$.

Néanmoins, si on effectue le test, on peut se demander quelle est la probabilité que k soit premier sachant que l'on a effectué l tests inconclusifs. Il s'agit là d'une probabilité de connaissance et on utilise pour cela la loi de Bayes. On obtient alors que la probabilité d'avoir un nombre composé après l tests inconclusifs est majorée par

$$\left(\frac{1}{4}\right)^l \times (\ln k - 1), \quad (3.22)$$

où le facteur logarithmique provient du fait que la probabilité que k soit premier vaut $1/\ln k$. Il faut noter que cette probabilité augmente avec la taille du nombre. Néanmoins, cela a peu d'impact car il s'agit d'un facteur linéaire en la taille alors que la probabilité évolue exponentiellement avec l .

Notons que l'on peut non seulement vérifier la primalité en temps polynomial de façon probabiliste, mais la densité des nombres premiers est suffisamment importante pour que l'on puisse également engendrer des nombres premiers en temps polynomial et ce de manière probabiliste. La probabilité

de tirer au hasard un nombre premier dont la taille est d'ordre n est $1/n$. Engendrer des nombres premiers peut donc se faire en temps polynomial : il faut tirer de l'ordre de n nombres au hasard avant de les passer au test de Miller-Rabin pour trouver un nombre aléatoire.

Bayes et BPP

La définition de BPP que nous avons vue énonce que la probabilité conditionnelle que la machine de Turing réponde « vrai » sachant que la réponse au problème est « vrai » doit être au moins $2/3$. De même, la probabilité de répondre « faux » est d'au moins $2/3$ lorsque la réponse au problème est « faux ». On formalise cela en écrivant

$$p(M_V|P_V) \geq 2/3, \quad p(M_F|P_F) \geq 2/3, \quad (3.23)$$

où $M_{V,F}$ désigne les évènements où la machine répond respectivement vrai ou faux, et $P_{V,F}$ désigne les évènements où la réponse au problème est respectivement vraie ou fausse.

On a également vu qu'en répétant l'exécution de la machine, on pouvait faire tendre ces probabilités vers 1 de façon exponentielle. Si on a une machine M' qui utilise l fois la machine M et affiche la réponse majoritaire, on a

$$p(M'_V|P_V) \geq 1 - \epsilon_l, \quad p(M'_F|P_F) \geq 1 - \epsilon_l. \quad (3.24)$$

En revanche, tout comme dans le test de Miller-Rabin, on est en général plutôt intéressé par la probabilité que la réponse soit vraie en fonction du résultat donné par la machine M' . En utilisant le théorème de Bayes, on peut majorer la probabilité que la machine se trompe. Dans le cas où la machine répond « vrai », on obtient

$$p(P_F|M'_V) \leq \epsilon_l \frac{p(P_F)}{p(P_V)}. \quad (3.25)$$

On voit ainsi que la probabilité d'erreur peut dépendre de la taille de l'instance par le ratio $p(P_F)/p(P_V)$. Bien évidemment, lorsque dans une instance de taille n il n'y a aucun élément pour lequel la réponse est « vrai », alors si la machine M' répond « vrai », celle-ci se trompe à coup sûr. Néanmoins, la probabilité que la machine se trompe est extrêmement faible et peut être rendue arbitrairement faible en augmentant l . Par ailleurs, s'il n'y a qu'un seul élément pour lequel la réponse est vraie, on a alors $p(P_F)/p(P_V) \approx 2^n$. Étant donné que ϵ_l a un comportement exponentiel en l , on peut réduire le taux d'erreur en augmentant l de façon linéaire en n . Ainsi, le temps total d'exécution dans le pire des cas reste polynomial.

Relations entre PP, BPP et les autres classes

Une question légitime est alors de connaître la relation entre BPP et les différentes classes de complexité que nous avons définies. D'une part, une

machine de Turing déterministe n'est qu'un cas particulier de machine de Turing probabiliste, qui retourne la bonne réponse avec une probabilité 1. Ainsi, un problème dans P est aussi dans BPP . De plus, d'après les définitions que l'on a données, il est trivial de voir que BPP est inclus dans PP . Pour résumer :

$$P \subseteq BPP \subseteq PP. \quad (3.26)$$

En fait, ce sont là les seules relations démontrées entre BPP et les différentes classes que nous avons vues. À première vue, on pourrait penser que BPP est strictement plus grand que P , et c'était effectivement l'hypothèse privilégiée jusqu'au début des années 2000.

Actuellement l'hypothèse privilégiée est plutôt $P = BPP$. Cela vient du fait que certains problèmes que l'on pense difficiles ne peuvent l'être que si ces deux classes sont égales. L'égalité peut sembler étrange de prime abord : cela revient à rendre déterministe un calcul probabiliste. De façon équivalente, cela revient à générer de façon déterministe des nombres qui semblent aléatoires. En physique classique, cela est permis par les processus chaotiques. Ceux-ci permettent, à partir d'un processus déterministe, de générer un phénomène qui semble aléatoire. Toute la question est alors de savoir s'il est facile ou non de distinguer ce qui semble aléatoire de ce qui ne l'est pas.

En informatique, le processus chaotique est un générateur de nombre pseudo-aléatoire. Ces derniers sont des processus entièrement déterministes, qui prennent un mot en entrée et génèrent un mot plus long en sortie. Le mot en entrée est appelé la *graine* (en anglais *seed*). Celle-ci est le petit bout de véritable aléa dont on a besoin pour initier le processus. Étant donné que le processus est déterministe, si on utilise deux fois le générateur avec la même graine, celui-ci produira la même sortie. De plus, on souhaite que, ne sachant pas quelle est la graine, il soit difficile de distinguer le mot produit par le générateur de nombre aléatoire d'un mot vraiment aléatoire. Pour cela, il faut donner certaines limites aux moyens de calcul donnés pour discerner le véritable aléatoire du faux. Ici, nous allons par exemple supposer que l'on ne peut discerner la distribution de probabilité des mots issus du générateur de la distribution uniforme, en temps polynomial en la taille de ces mots, sur une machine de Turing déterministe.

En fait, si l'on a accès à un générateur de nombres pseudo-aléatoires qui permet d'étendre une graine de façon exponentielle, on peut alors montrer que $P = BPP$. Pour une entrée de taille n , la machine de Turing non déterministe a besoin de tirer au hasard un nombre polynomial de transitions. Au lieu de tirer ces transitions à partir du vrai hasard, on peut les tirer à l'aide du générateur de nombre pseudo-aléatoire. Celui-ci a donc besoin d'une graine dont la taille est de l'ordre de $\log n$, afin de couvrir l'ensemble des transitions possibles. Comment choisir correctement la graine ? Pour cela, on peut simplement itérer sur l'ensemble des graines possibles, ce qui est possible en temps polynomial. Si le problème que l'on a est dans BPP , la seule possibilité est alors que la machine de Turing réponde correctement. Sans cela, on pourrait l'utiliser

comme moyen pour distinguer entre la distribution fournie par notre générateur et la distribution uniforme, ce qui contredirait notre hypothèse.

Toute la question est alors de savoir s'il existe un tel générateur de nombres pseudo-aléatoires. Pour l'instant cette question est ouverte. Néanmoins, certains problèmes conjecturés difficiles impliquent l'existence d'un tel générateur. Cela penche donc pour l'égalité $P = BPP$.

D'un point de vue physique, et plus philosophique, l'existence de générateur de nombres pseudo-aléatoires que nous avons supposée indique qu'il est parfois impossible de faire la différence entre chaos classique et aléatoire, tout du moins avec des moyens de calcul raisonnables. Certaines tâches ne peuvent pas être effectuées sans aléatoire du tout : si l'on souhaite générer une suite qui soit aléatoire (au moins aux yeux d'un autre observateur), il faut *a minima* initialiser le générateur avec une graine aléatoire. Cependant, ce besoin est *a priori* beaucoup plus faible que ce que l'on pourrait penser. Ainsi, pour le calcul des problèmes de décision, il semble que le chaos suffise, et que l'on n'ait pas besoin d'appeler le véritable aléatoire à la rescousse.

3.4.5 Thèse de Church-Turing étendue

Maintenant que nous avons classé les problèmes en différentes catégories, nous pouvons nous poser la question de savoir quels sont les problèmes que l'on peut résoudre efficacement en pratique. C'est une question analogue à celle que l'on s'est posée dans la section 3.3, mais pour la théorie de la complexité. La *thèse de Church-Turing étendue* consiste à poser une des classes de complexité comme la classe des problèmes calculables en pratique.

Paradoxalement, si la thèse de Church-Turing peut s'interpréter de multiples manières, il est clair que la thèse de Church-Turing étendue est en rapport avec le monde physique. En se demandant quels problèmes on peut résoudre en temps polynomial sur une machine physique, cela revient à se demander quelle variante de la machine de Turing on peut réaliser dans le monde physique. Pour l'instant, nous en avons vu trois : la machine de Turing déterministe, la machine de Turing probabiliste et la machine de Turing non déterministe. Pour chacune de ces machines, il y a la classe que l'on peut calculer efficacement, à savoir P , BPP et NP respectivement. La question est donc de savoir si le monde est intrinsèquement déterministe, probabiliste ou non déterministe²⁰.

Afin de donner une première version de la thèse de Church-Turing étendue, nous allons regarder quels ingrédients peut utiliser un ordinateur actuel. Ce dernier est principalement régi par la physique classique (dans les limites que nous avons établies dans la section précédente). Il est donc extrêmement proche d'une machine de Turing déterministe. Dans une certaine idéalisation, un ordinateur est donc parfaitement déterministe. Toutefois, celui-ci est connecté au monde physique, à l'aide de différents instruments : horloge, interface réseau,

20. Dans le cas où certaines de ces classes sont égales, cela supprime les différences, en tout cas d'un point de vue complexité, entre ces différentes visions du monde.

etc. Ceux-ci, ainsi que des périphériques spécifiques, peuvent être utilisés pour servir de source d'aléa, soit pour donner une graine à un générateur de nombre pseudo-aléatoire, soit pour générer de véritables nombres aléatoires. Mais il existe également des moyens d'engendrer de vraies suites aléatoires en utilisant les principes de la physique quantique. Depuis quelques années, on trouve dans le commerce des périphériques de la taille d'une clé USB produisant un flux d'aléa en utilisant des photons envoyés à travers une lame semi-réfléchissante puis détectés.

En d'autres termes, nous faisons l'hypothèse ici d'un monde probabiliste. Dans un tel monde, la thèse de Church-Turing étendue prend la forme suivante :

Une machine de Turing probabiliste peut simuler efficacement
n'importe quel modèle de calcul réaliste.

Le statut de cette thèse est beaucoup plus fragile que celui de la thèse de Church-Turing. Cette dernière est une thèse fondatrice de l'informatique théorique notamment justifiée par l'équivalence d'une multitude de modèles, alors que la thèse de Church-Turing étendue est plutôt une hypothèse sur les capacités de mise en œuvre en pratique dans le monde physique. Une machine capable de résoudre rapidement et avec un nombre limité de ressources des problèmes en dehors de BPP remettrait en cause celle-ci.

Jusqu'à présent, nous avons uniquement utilisé des ressources classiques pour le calcul. Cependant, le monde quantique n'est pas seulement probabiliste : la mécanique quantique autorise les superpositions d'états et met donc en œuvre une sorte très particulière de parallélisme. Un ordinateur quantique pourrait donc *a priori* résoudre efficacement plus de problèmes qu'un ordinateur classique. Sous réserve de la possibilité pratique de construire un ordinateur quantique, et sous certaines hypothèses raisonnables de théorie de la complexité, la thèse de Church-Turing étendue serait donc invalidée.

Nous aborderons ces questions au cours du chapitre 8. Nous y verrons une nouvelle classe de complexité, analogue à BPP mais pour une machine de Turing exploitant les lois de la théorie quantique : la classe BQP.

Chapitre 4

Systèmes quantiques simples

Disposant du cadre conceptuel de la théorie quantique, de la théorie de l'information et des notions de calculabilité et de complexité, il nous faut maintenant pouvoir modéliser les systèmes physiques qui pourront être utilisés pour communiquer et calculer avec des ressources quantiques.

De façon assez étonnante, il est possible de décrire la physique contenue dans la plupart des modèles en utilisant deux systèmes quantiques simples : le système à deux niveaux, encore appelé qubit (*quantum bit*) et l'oscillateur harmonique.

Le qubit est l'objet fondamental de l'information et du calcul quantique. Les exemples de systèmes à deux niveaux (exacts ou effectifs) manipulables sont nombreux : la molécule d'ammoniac [156, Tome mécanique quantique], certains couples de niveaux d'atomes de Rydberg [194] ou encore la polarisation lumineuse. Leur simplicité et leur importance pour l'information quantique en font un parfait exemple pour illustrer les concepts de base de la physique quantique introduits au chapitre 1.

L'oscillateur harmonique revêt aussi une importance fondamentale. Il décrit non seulement le comportement en première approximation de tout système physique autour d'une position d'équilibre stable mais permet aussi de décrire la dynamique du champ électromagnétique quantique. On dispose donc de tout un ensemble de systèmes physiques manipulables se comportant comme des oscillateurs harmoniques.

Bien sûr, il ne faut pas imaginer qu'utiliser ou assembler des composants simples nous cantonne nécessairement à une physique simple. Au contraire, la physique du XX^e siècle a bien compris qu'avec la complexité grandissante des systèmes émergeait une physique nouvelle.

Enfin, nous commencerons à aborder des questions conceptuelles importantes qui seront développées plus précisément par la suite : comment retrouver un comportement classique à partir d'un modèle quantique grâce, par exemple, aux états cohérents ? Comment appréhender les différences essentielles entre information classique et quantique grâce, par exemple, au protocole de tomographie ?

4.1 Systèmes à deux niveaux

4.1.1 De l'état au vecteur de Bloch

Par définition, l'espace des états $\mathcal{H}$ d'un système à deux niveaux ou qubit est de dimension deux sur $\mathbb{C}$. N'importe quel état $|\psi\rangle$ se développe sous la forme d'une combinaison linéaire des éléments d'une base $(|0\rangle, |1\rangle)$ orthonormée de $\mathcal{H}$ comme

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle \quad (4.1)$$

avec $|\alpha|^2 + |\beta|^2 = 1$. D'autres bases peuvent être considérées suivant la situation physique que l'on souhaite examiner. C'est particulièrement vrai dans le cas de la polarisation de la lumière où il existe des polarisations linéaires selon n'importe quelle direction orthogonale à la direction de propagation ou des polarisations circulaires.

Nous allons construire une représentation géométrique simple et très utile des états d'un système à deux niveaux. Pour cela, il est en premier lieu nécessaire de donner la forme générique des observables agissant sur un qubit.

Pour un système à deux niveaux, tout opérateur hermitien peut être écrit comme combinaison linéaire de quatre opérateurs particuliers. Sous forme matricielle, ces opérateurs sont les *matrices de Pauli* accompagnées de l'identité

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (4.2)$$

Elles vérifient toutes l'égalité

$$\sigma_i^2 = \mathbb{1}. \quad (4.3)$$

Chacune des matrices de Pauli peut être diagonalisée et l'ensemble des vecteurs propres forme à chaque fois une base orthonormée de l'espace de Hilbert. Les bases associées à ces trois opérateurs sont ¹

$$|X_{\pm}\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} \quad |Y_{\pm}\rangle = \frac{|0\rangle \pm i|1\rangle}{\sqrt{2}} \quad |Z_{+}\rangle = |0\rangle, \quad |Z_{-}\rangle = |1\rangle. \quad (4.4)$$

Un état quelconque peut être écrit indifféremment dans n'importe laquelle de ces bases. Les probabilités d'observer $+1$ ou -1 pour ces différentes observables sont alors obtenues par la règle de Born,

$$\begin{aligned} p(z = 1|\psi) &= |\alpha|^2 & p(z = -1|\psi) &= |\beta|^2 \\ p(x = 1|\psi) &= \frac{|\alpha + \beta|^2}{2} & p(x = -1|\psi) &= \frac{|\alpha - \beta|^2}{2} \\ p(y = 1|\psi) &= \frac{|\alpha - i\beta|^2}{2} & p(y = -1|\psi) &= \frac{|\alpha + i\beta|^2}{2}. \end{aligned} \quad (4.5)$$

1. On utilisera également la notation $|\pm\rangle \equiv |X_{\pm}\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$.

Sachant cela, les valeurs moyennes des observables σ_a pour $a = x, y, z$ dans l'état $|\psi\rangle$ sont données par

$$\langle \sigma_z \rangle = |\alpha|^2 - |\beta|^2, \quad (4.6a)$$

$$\langle \sigma_x \rangle = \frac{|\alpha + \beta|^2 - |\alpha - \beta|^2}{2} = 2 \Re(\alpha\beta^*), \quad (4.6b)$$

$$\langle \sigma_y \rangle = \frac{|\alpha - i\beta|^2 - |\alpha + i\beta|^2}{2} = 2 \Im(\alpha^*\beta). \quad (4.6c)$$

Puisque $|\alpha|^2 + |\beta|^2 = 1$, le paramétrage suivant ($0 \leq p \leq 1$)

$$\alpha = \sqrt{p} e^{i\theta_+} \quad \beta = \sqrt{1-p} e^{i\theta_-} \quad (4.7)$$

nous permet d'écrire

$$\langle \sigma_z \rangle = 2p - 1, \quad (4.8a)$$

$$\langle \sigma_x \rangle = 2\sqrt{p(1-p)} \cos(\theta_+ - \theta_-), \quad (4.8b)$$

$$\langle \sigma_y \rangle = 2\sqrt{p(1-p)} \sin(\theta_- - \theta_+). \quad (4.8c)$$

Le vecteur $\mathbf{n}[|\psi\rangle]$ associé à l'état $|\psi\rangle$ est défini par

$$\mathbf{n}[|\psi\rangle] = \langle \sigma_x \rangle \mathbf{e}_x + \langle \sigma_y \rangle \mathbf{e}_y + \langle \sigma_z \rangle \mathbf{e}_z \quad (4.9)$$

où $(\mathbf{e}_a)_{a=x,y,z}$ forme une base orthonormée de $\mathbb{R}^3$. L'équation (4.8) montre que $\mathbf{n}[|\psi\rangle]$ est de norme 1 et qu'il ne dépend que de la différence des phases $\theta_- - \theta_+$ entre les deux coefficients α et β mais pas de la phase moyenne $(\theta_+ + \theta_-)/2$. Il correspond donc véritablement à l'état physique associé à l'état $|\psi\rangle$. Le paramétrage $p = \cos^2 \theta/2$ avec $\theta \in [0, \pi]$ montre que le vecteur $\mathbf{n}[|\psi\rangle]$ a pour coordonnées sphériques (θ, φ) avec $\varphi = \theta_- - \theta_+$ (voir figure 4.1). Il est appelé *vecteur de Bloch* associé à l'état $|\psi\rangle$ tandis que la sphère unité dans $\mathbb{R}^3$ est la *sphère de Bloch*.

La figure 4.1 montre géométriquement les vecteurs de Bloch associés aux différents vecteurs propres des matrices de Pauli. Remarquons ainsi que $\mathbf{n}[|X_+\rangle] = \mathbf{n}\left[\frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle)\right]$ n'est pas combinaison linéaire de $\mathbf{n}[|0\rangle]$ et de $\mathbf{n}[|1\rangle]$. Cette remarque est vraie de manière générale : le vecteur de Bloch associé à une superposition d'états n'est pas une combinaison linéaire simple des vecteurs de Bloch de chaque membre de la superposition. Une base orthonormée de l'espace des états d'un qubit correspond dans la représentation de Bloch à un couple de points diamétralement opposés de la sphère. Ceci découle des équations (4.6). En effet, si $\mathbf{n}$ désigne le vecteur de Bloch associé à un état $|\psi\rangle$ normé, alors

$$|\langle \psi_1 | \psi_2 \rangle|^2 = \frac{1 + \mathbf{n}_1 \cdot \mathbf{n}_2}{2}. \quad (4.10)$$

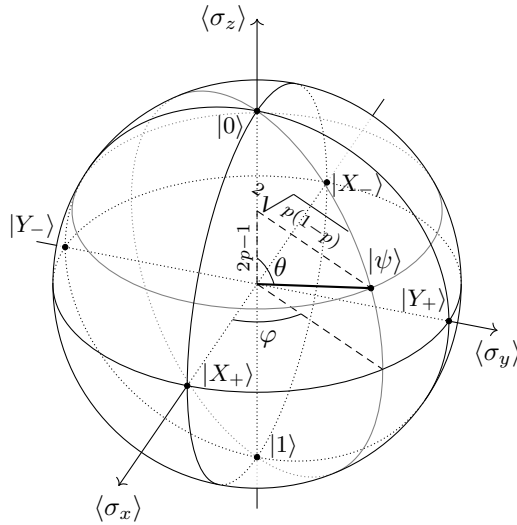


FIGURE 4.1 – Détermination du vecteur de Bloch associé à l'état $\sqrt{p}e^{i\theta_+}|0\rangle + \sqrt{1-p}e^{i\theta_-}|1\rangle$. Les coordonnées sphériques de ce vecteur sont (θ, φ) telles que $p = \cos^2(\theta/2)$ et $\varphi = \theta_- - \theta_+$. Les vecteurs de Bloch correspondant aux vecteurs propres des trois matrices de Pauli σ_x , σ_y et σ_z sont également indiqués sur la sphère.

4.1.2 Du vecteur de Bloch à l'état

Qu'en est-il de la réciproque, c'est-à-dire retrouver un état à partir d'un vecteur $\mathbf{n}$ de la sphère de Bloch ? Les formules (4.7) et (4.8) fournissent la réponse. En effet, elles montrent que $|\alpha|$ et $|\beta|$ sont parfaitement déterminés car, d'une part, $p = (n^z + 1)/2$ et d'autre part, la différence des phases $\theta_- - \theta_+$ correspondant à la longitude du vecteur de Bloch est totalement fixée. La demi-somme $(\theta_+ + \theta_-)/2$ ne l'est pas mais une phase globale ne change pas l'état physique. On peut donc, pour chaque vecteur $\mathbf{n}$ de la sphère de Bloch, trouver un état physique tel que $\langle \boldsymbol{\sigma} \rangle = \mathbf{n}$ où $\boldsymbol{\sigma}$ désigne le vecteur formé des trois matrices de Pauli.

Si à chaque vecteur $\mathbf{n}$ de la sphère de Bloch, on associe un unique état physique, il y a en revanche une ambiguïté de phase dans la reconstruction d'un vecteur normé dans l'espace de Hilbert. On doit donc faire un choix. Une possibilité naturelle consiste à choisir $\theta_+ + \theta_- = 0$. Dans ce cas, si nous introduisons les coordonnées sphériques (θ, φ) du vecteur $\mathbf{n}$

$$n_x = \cos \varphi \sin \theta, \quad n_y = \sin \varphi \sin \theta, \quad n_z = \cos \theta, \quad (4.11)$$

où $\theta \in [0, \pi]$ et $\varphi \in [0, 2\pi]$, nous posons :

$$|\mathbf{n}(\theta, \varphi)\rangle = e^{-i\varphi/2} \cos \theta/2 |0\rangle + e^{i\varphi/2} \sin \theta/2 |1\rangle. \quad (4.12)$$

Cette application associe effectivement au vecteur $\mathbf{n}$ un état $|\mathbf{n}\rangle$ tel que $\langle \sigma \rangle_{|\mathbf{n}\rangle} = \mathbf{n}$. Remarquons que $(\theta, \varphi) \mapsto |\mathbf{n}(\theta, \varphi)\rangle$ n'est pas continue puisque $|\mathbf{n}(\theta, \varphi + 2\pi)\rangle = -|\mathbf{n}(\theta, \varphi)\rangle$.

4.1.3 Les transformations du qubit

Ayant approfondi la structure de l'espace des états d'un qubit, il reste à comprendre quelles sont les transformations unitaires que l'on peut lui appliquer. En l'absence de mesure, ces transformations forment un groupe $U(2)$ appelé *groupe unitaire* sur $\mathbb{C}^2$. Son sous-groupe $SU(2)$ des matrices à déterminant unité permet de le décomposer en $U(2) = U(1) \times SU(2)$. La partie $U(1)$ agit comme une phase globale et ne change pas les états physiques. La forme générale d'un élément de $SU(2)$ est

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \text{ avec } |\alpha|^2 + |\beta|^2 = 1. \quad (4.13)$$

Nous pouvons paramétrer ces deux nombres complexes au moyen de trois angles $(\theta, \varphi_+, \varphi_-)$ avec $\alpha = e^{i\varphi_+} \cos \theta/2$ et $\beta = e^{i\varphi_-} \sin \theta/2$.

Calculons maintenant l'action correspondante sur le vecteur de Bloch $\mathbf{n}(|\psi\rangle)$. Pour cela, il nous suffit de calculer les opérateurs $U^\dagger \sigma_a U$ pour $a = x, y, z$ et de les exprimer comme combinaisons linéaires des $(\sigma_a)_{a=x,y,z}$:

$$U^\dagger \sigma_a U = \sum_b R_{ab}(U) \sigma_b. \quad (4.14)$$

Considérons le vecteur de Bloch $\mathbf{n}(|\psi\rangle) = \sum_a n_a \mathbf{e}_a$ avec $n^a = \langle \sigma_a \rangle_{|\psi\rangle}$. L'équation (4.14) montre que l'action de U sur le vecteur de Bloch correspond à l'action de $R(U)$ en tant que matrice sur ses composantes :

$$\mathbf{n}[U|\psi\rangle] = R(U) \mathbf{n}[|\psi\rangle]. \quad (4.15)$$

La matrice $R(U)$ caractérise donc bien l'action de $SU(2)$ sur les états physiques. Nous prouverons dans la suite que $R(U)$ est une matrice de rotation. Cet ensemble des rotations d'un vecteur dans l'espace forme lui aussi un groupe $SO(3)$ appelé groupe des rotations dans $\mathbb{R}^3$. Une propriété importante que possède l'application R qui envoie $U \in SU(2)$ sur $R(U) \in SO(3)$ est celle de morphisme de groupe. En effet,

$$(U_1 U_2)^\dagger \sigma_a (U_1 U_2) = \sum_{b,c} R_{ab}(U_1) R_{b,c}(U_2) \sigma_c = \sum_c (R(U_1) R(U_2))_{ac} \sigma_c. \quad (4.16)$$

Pour obtenir l'expression de $R(U)$, nous pouvons utiliser une décomposition de la matrice U comme le produit de trois éléments de $SU(2)$

$$U = \exp\left(i \frac{\varphi_+ + \varphi_-}{2} \sigma_z\right) \exp\left(i \frac{\theta}{2} \sigma_y\right) \exp\left(i \frac{\varphi_+ - \varphi_-}{2} \sigma_z\right). \quad (4.17)$$

Pour un élément de $SU(2)$ de la forme $e^{i\theta\boldsymbol{\sigma}\cdot\mathbf{n}/2}$, il est facile de démontrer, en développant l'exponentielle, que

$$U(\mathbf{n}, \theta) \equiv e^{i\theta\boldsymbol{\sigma}\cdot\mathbf{n}/2} = \cos \theta/2 \mathbb{1}_2 + i \sin \theta/2 \mathbf{n} \cdot \boldsymbol{\sigma}. \quad (4.18)$$

De plus, pour les éléments de $SU(2)$ de la forme $\exp(i\varphi\sigma_a/2)$ avec $a = x, y, z$, on démontre que la matrice $R(U)$ n'est autre que la rotation d'angle φ autour de l'axe défini par le vecteur unitaire $\mathbf{e}_a$ que nous notons $R_a(\varphi)$. On a donc en général $R(U(\mathbf{n}, \theta)) = R_{\mathbf{n}}(\theta)$, ce qui se traduit en disant que l'on a une rotation d'angle θ autour de l'axe $\mathbf{n}$. Ceci permet donc de conclure, par la propriété de morphisme de groupe, que

$$R(U) = R_z(\varphi_+ + \varphi_-) R_y(\theta) R_z(\varphi_+ - \varphi_-). \quad (4.19)$$

Outre le fait que cela confirme que $R(U) \in SO(3)$, cette formule permet de trouver facilement le noyau du morphisme de $SU(2)$ dans $SO(3)$: il s'agit des $U(\theta, \varphi_+, \varphi_-)$ qui donnent une rotation identité. Cela impose d'une part que $\theta = 0$ pour que $\mathbf{e}_z$ soit laissé fixe et d'autre part que $\varphi_+ + \varphi_-$ et $\varphi_+ - \varphi_-$ soient nuls modulo 2π . Nous obtenons donc que $\varphi_{\pm} = (n \pm m)\pi$ où $(n, m) \in \mathbb{Z}^2$. Si n et m sont tous deux de même parité, alors les deux angles $\varphi_{\pm}$ sont tous les deux nuls et $U = \mathbb{1}$. Par contre si n et m sont de parités différentes, nous voyons que cela donne $U = -\mathbb{1}$. Le noyau contient deux éléments $\pm\mathbb{1}$. Pour n'importe quelle rotation $R \in SO(3)$, il existe bien un élément $U \in SU(2)$ tel que $R(U) = R$. L'application R de $SU(2)$ dans $SO(3)$ est donc un morphisme surjectif dont le noyau est formé des deux matrices $\pm\mathbb{1}$. Le groupe $SU(2)$ est en fait difféomorphe à la sphère S_3 sur laquelle il agit naturellement par multiplication à gauche : cette action correspond à celle de U sur un état normé dans $\mathbb{C}^2$. Ce que nous venons de faire revient donc à descendre cette action de $SU(2)$ sur S_3 au niveau de la sphère de Bloch (*grosso modo* S_2), ce que résume la figure 4.2.

4.1.4 La polarisation du photon

L'importance du qubit provient du fait que non seulement il existe dans la nature des systèmes ayant deux degrés de liberté comme les photons (pour leur polarisation) ou les particules de spin $1/2$, mais aussi que de nombreuses situations physiques peuvent être décrites de façon effective par un système à deux niveaux (c'est-à-dire en négligeant ou en tenant compte de façon approximative de l'ensemble des autres niveaux). Cette section est dédiée à la présentation de quelques réalisations standard d'un qubit servant actuellement d'ingrédient de base à l'étude expérimentale de l'information quantique.

De la polarisation lumineuse à la sphère de Bloch

L'analyse des équations de Maxwell décrivant la lumière montre que le champ électromagnétique véhicule deux degrés de liberté caractérisant la

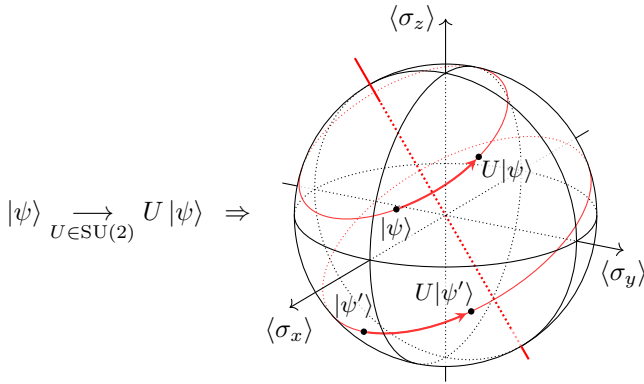


FIGURE 4.2 – À gauche : action d'un élément U de $SU(2)$ sur un état normé de $\mathbb{C}^2$. À droite : l'action correspondante sur le vecteur de Bloch se fait *via* une rotation de $SO(3)$ agissant sur la sphère de Bloch plongée dans l'espace $\mathbb{R}^3$.

polarisation de la lumière. Au niveau quantique, la lumière est décrite par une excitation élémentaire, le photon, lui-même possédant deux degrés de liberté. Les détails de la théorie quantique de la lumière ne seront pas abordés ici mais nous utiliserons tout de même la terminologie de la théorie quantique pour parler du champ électromagnétique.

On considère un photon se propageant dans la direction (Oz) . Sa polarisation est alors dans le plan équatorial (Oxy) , ce qui correspond bien à un système à deux degrés de liberté. Deux polarisations rectilignes naturelles existent (données par la direction des composantes du champ électrique) et sont définies par

$$|\uparrow\rangle : \begin{cases} \xi_x(t) = 0 \\ \xi_y(t) = \xi_0 \cos \omega t \end{cases} \quad |\leftrightarrow\rangle : \begin{cases} \xi_x(t) = \xi_0 \sin \omega t \\ \xi_y(t) = 0. \end{cases} \quad (4.20)$$

La première est appelée polarisation verticale (notée $|\uparrow\rangle$) et la seconde polarisation horizontale (notée $|\leftrightarrow\rangle$). Dans le cas général, une polarisation rectiligne $|p_\theta\rangle$ vibrant dans une direction définie par un angle θ par rapport à l'horizontale, comme on le voit sur la figure 4.3, s'écrit comme

$$\begin{cases} \xi_x(t) = \xi_0 \cos \theta \cos \omega t \\ \xi_y(t) = \xi_0 \sin \theta \cos \omega t. \end{cases} \quad (4.21)$$

Mais à quels états quantiques sont-elles associées ? Pour y répondre, procédons heuristiquement en donnant une « recette » qui nous permette de mettre en correspondance des amplitudes du champ électrique classique avec des états quantiques. Dans le cas de la polarisation linéaire, l'amplitude décrite par

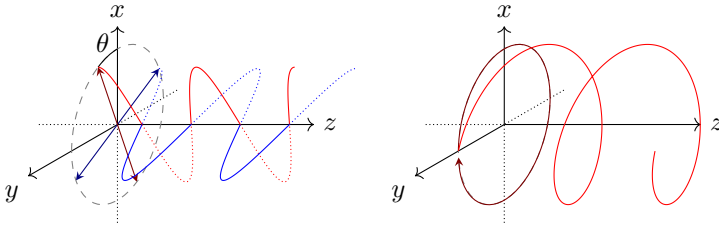


FIGURE 4.3 – Une représentation de différents états possibles de polarisation pour un photon se propageant selon la direction (Oz) . À gauche : les polarisations linéaires selon θ et selon $\theta + \pi/2$. À droite : la polarisation circulaire droite.

l'équation (4.21) est obtenue en sommant les amplitudes des polarisations horizontales et verticales avec des pondérations réelles $\cos \theta$ et $\sin \theta$. Cela nous amène donc à écrire

$$|p_\theta\rangle = \cos \theta |\leftrightarrow\rangle + \sin \theta |\updownarrow\rangle. \quad (4.22)$$

Notons que dans ce cas, la probabilité qu'un photon polarisé linéairement selon θ soit vu comme horizontal (resp. vertical) est alors donnée par $\cos^2 \theta$ (resp. $\sin^2 \theta$), ce qui est cohérent avec les lois de Malus décrivant la manière dont un faisceau de lumière classique avec cette polarisation linéaire est filtré par un polariseur horizontal (resp. vertical).

Mais il existe d'autres polarisations, comme les polarisations circulaires. Toujours dans le plan équatorial, on définit la polarisation circulaire droite $|D\rangle$, représentée sur la figure 4.3, ainsi que la polarisation circulaire gauche $|G\rangle$ comme suit

$$|D\rangle : \begin{cases} \xi_x(t) = \xi_0 \cos \omega t \\ \xi_y(t) = \xi_0 \sin \omega t \end{cases} \quad |G\rangle : \begin{cases} \xi_x(t) = \xi_0 \cos \omega t \\ \xi_y(t) = -\xi_0 \sin \omega t. \end{cases} \quad (4.23)$$

Remarquons maintenant que la polarisation circulaire droite peut être vue comme la superposition d'une onde polarisée horizontalement et d'une onde polarisée verticalement mais avec un déphasage de $\pi/2$ qui transforme le cosinus en sinus. Ainsi, au niveau de l'état quantique, notre règle empirique suggère donc que

$$\begin{cases} |D\rangle = \frac{|\leftrightarrow\rangle + i|\updownarrow\rangle}{\sqrt{2}} \\ |G\rangle = \frac{|\leftrightarrow\rangle - i|\updownarrow\rangle}{\sqrt{2}}, \end{cases} \quad (4.24)$$

où le déphasage est introduit sous la forme du facteur $\pm i$. De manière plus générale, cherchons à quoi correspond une phase arbitraire $(|\leftrightarrow\rangle + e^{i\phi}|\updownarrow\rangle)/\sqrt{2}$ dans la superposition de deux états de polarisation horizontale et verticale. Pour cela, appliquons notre principe empirique pour trouver l'amplitude classique en fonction des coefficients de la superposition quantique. On superpose donc

les champs associés avec les ondes polarisées horizontalement et verticalement mais en les déphasant par ϕ , ce qui donne

$$\begin{cases} \xi_x(t) = \xi_0 \cos \omega t \\ \xi_y(t) = \xi_0 \cos(\omega t - \phi) = \xi_0 \cos \phi \cos \omega t + \xi_0 \sin \phi \sin \omega t. \end{cases} \quad (4.25)$$

Pour reconnaître à quelle polarisation classique nous avons affaire, considérons des axes tournés de $\pi/4$. Les composantes du champ électrique le long de ces deux axes $\xi_{\pm}(t) = \frac{1}{\sqrt{2}} (\xi_x(t) \pm \xi_y(t))$ sont alors données par

$$\xi_{\pm}(t) = \pm \xi_0 (1 \pm \cos \phi) \cos \omega t \pm \xi_0 \sin \phi \sin \omega t, \quad (4.26)$$

que nous pouvons réécrire sous la forme plus parlante

$$\begin{cases} \xi_+(t) = \sqrt{2} \xi_0 \cos \frac{\phi}{2} \cos \left(\omega t - \frac{\phi}{2} \right) \\ \xi_-(t) = \sqrt{2} \xi_0 \sin \frac{\phi}{2} \sin \left(\omega t - \frac{\phi}{2} \right). \end{cases} \quad (4.27)$$

Nous voyons donc apparaître des dépendances temporelles en quadrature comme pour une polarisation circulaire. Cependant, les amplitudes des deux quadratures ne sont pas identiques : le cercle est aplati et, de manière générique, on décrit donc une polarisation elliptique dont les axes principaux sont les axes inclinés de $\pm\pi/4$ par rapport à l'axe horizontal. Le cas particulier de la polarisation rectiligne s'obtient pour $\phi \equiv 0 \pmod{\pi}$ et le cas circulaire est obtenu pour $\phi \equiv \pi/2 \pmod{\pi}$.

À ce stade, nous n'avons pas justifié la règle empirique qui a permis de construire la correspondance entre les polarisations des ondes électromagnétiques classiques et les points de la sphère de Bloch, c'est-à-dire les états d'un système à deux niveaux quantiques. Cependant, cette discussion tout à fait heuristique nous montre comment interpréter en termes de polarisation lumineuse tous les points de la sphère de Bloch, ce qui est résumé sur la figure 4.4.

En effet, les polarisations linéaires horizontale et verticale correspondent respectivement aux vecteurs $\pm \mathbf{e}_z$. Les polarisations circulaires droite et gauche correspondent respectivement aux vecteurs $\pm \mathbf{e}_y$. La polarisation linéaire p_θ correspond au vecteur $\cos(2\theta) \mathbf{e}_z + \sin(2\theta) \mathbf{e}_x$: la latitude le long du méridien reliant les pôles $\pm \mathbf{e}_z$ et passant par $\mathbf{e}_x$ n'est autre que le double de l'angle donnant la direction de polarisation, ce qui est logique car tourner un polariseur linéaire de 180 degrés ne change pas son action. Deux directions de polarisation linéaires orthogonales différant de $\pi/2$ correspondent donc à des vecteurs opposés sur la sphère de Bloch.

Enfin, l'état $\frac{1}{\sqrt{2}} (|\leftrightarrow\rangle + e^{i\phi} |\uparrow\rangle)$ correspond à un vecteur de Bloch situé dans le plan équatorial et dont la longitude, comptée à partir de $\mathbf{e}_x$ n'est autre que l'angle ϕ . Nous venons d'identifier chacun de ces points avec une polarisation

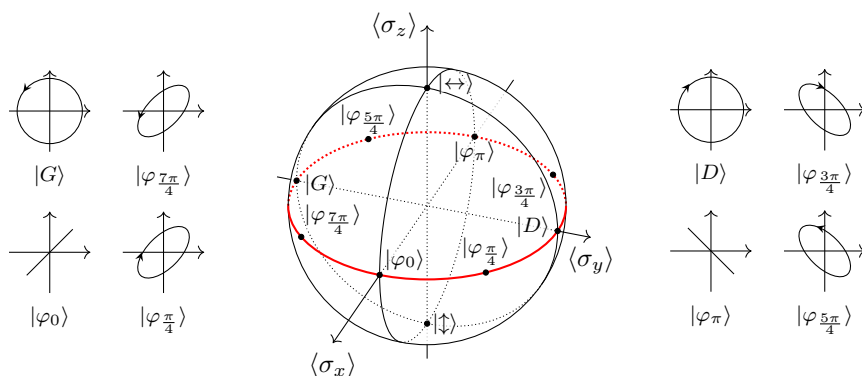


FIGURE 4.4 – Les polarisations sur l'équateur de la sphère de Bloch correspondent aux polarisations elliptiques dont les axes principaux sont à angle $\pm\pi/4$ par rapport à l'axe horizontal.

elliptique (ou exceptionnellement circulaire ou linéaire). La figure 4.4 représente toutes ces différentes polarisations sur l'équateur de la sphère de Bloch.

D'une manière générale, les polarisations correspondant à n'importe quel cercle passant par les deux polarisations circulaires D et G s'interprètent de la même manière : ce sont toutes les polarisations elliptiques dont les axes principaux correspondent aux deux points d'intersection de ce cercle avec le méridien des polarisations linéaires. Nous laissons au lecteur le soin de se convaincre que la règle empirique que nous avons formulée permet exactement d'aboutir à cette description pour tous ces cercles.

Comment manipuler la polarisation d'un photon ?

Encoder de l'information quantique dans l'état du qubit que forme la polarisation d'un photon ne peut être utile que si l'on sait manipuler son état. Comment donc réaliser une transformation unitaire arbitraire sur l'état de polarisation d'un photon ?

Avant de parler du problème général, on peut déjà voir comment réaliser certaines transformations unitaires simples à partir de composants optiques ordinaires. Les matériaux biréfringents sont des matériaux anisotropes qui ont la propriété de propager la lumière de manière différente selon sa polarisation. Ils peuvent donc être utilisés pour réaliser des composants dans lesquels deux polarisations linéaires perpendiculaires de la lumière voient un indice de réfraction différent et accumulent donc un retard de phase sur une certaine distance de propagation. Dans la base correspondant à ces deux polarisations rectilignes, l'action d'une telle lame s'apparente donc à une matrice unitaire

de la forme

$$e^{i\theta\sigma_z/2} = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}, \quad (4.28)$$

pour une longueur d'onde donnée. Les lames demi-ondes sont celles pour lesquelles le déphasage atteint π tandis que pour une lame quart d'onde, il atteint $\pi/2$. La lame quart d'onde permet donc de transformer une polarisation linéaire en une polarisation circulaire (et réciproquement). De même une lame demi-onde fait tourner les polarisations de $\pm\pi/4$ et échange les polarisations circulaires droites et gauches.

Les matériaux biréfringents servent également à fabriquer un prisme de Wollaston, qui est un composant séparant un faisceau lumineux en deux faisceaux de polarisations orthogonales, ce qui permet de filtrer spatialement une polarisation particulière. Le prisme de Wollaston n'est qu'un cas particulier de ce qu'on appelle les séparateurs de polarisation.

En combinant ces éléments, il est possible de réaliser n'importe quelle transformation unitaire sur l'état de polarisation d'un photon. De même, l'utilisation d'une lame quart d'onde et d'un polariseur permet de déterminer l'état de polarisation d'un faisceau lumineux classique monochromatique.

✱ Les degrés de liberté du photon

L'identification des degrés de liberté de polarisation du photon avec un système à deux niveaux trouve son origine dans le fait que, pour le champ électromagnétique, les seuls degrés de liberté dynamiques sont les degrés de liberté transverses. Ce résultat provient d'une analyse classique de la théorie de l'électromagnétisme, le champ étant décrit par le quadrivecteur potentiel. Un dénombrement naïf conclurait à quatre degrés de liberté en chaque point de l'espace (ou de l'espace réciproque), soit huit coordonnées et moments associés en formalisme hamiltonien. Cependant, l'invariance de jauge de la théorie vient restreindre ce nombre. Celle-ci stipule que l'on peut changer le potentiel vecteur et scalaire de manière locale sans que cela altère la physique.

Classiquement, les champs électrique et magnétique nous fournissent $3 + 3 = 6$ quantités dynamiques physiques en chaque point. Mais on sait que l'équation de Maxwell $\text{div } \mathbf{B} = 0$ ainsi que la loi de Gauss $\text{div } \mathbf{E} = -\rho/\varepsilon_0$ réduisent ce nombre à $2 + 2 = 4$ quantités physiques indépendantes en chaque point. Or, il s'avère qu'une analyse hamiltonienne montre que le potentiel électrique n'a pas de moment conjugué (c'est un multiplicateur de Lagrange associé à la contrainte de Gauss). Ne restent que trois degrés de liberté. C'est ici que l'invariance de jauge intervient dont l'ambiguïté inhérente doit être fixée en imposant une condition de jauge comme celle de Coulomb $\text{div } \mathbf{A} = 0$.

Il reste au final deux degrés de liberté dynamiques par point de l'espace direct (ou réciproque), ce qui correspond aux deux polarisations physiques du photon.

Ayant ainsi extrait les degrés de liberté physiques du champ électromagnétique, il est possible d'appliquer les méthodes de quantification canonique

pour obtenir l'électrodynamique quantique (en jauge de Coulomb) dont les vérifications expérimentales sont à ce jour les plus précises en physique. Le lecteur intéressé par la quantification canonique de l'électrodynamique pourra consulter [105, Chapitre 2] pour une approche pas à pas du problème. Le lecteur intéressé par un lien plus explicite avec la théorie générale des systèmes contraints consultera avec profit [216, Chapitre 9] mais aussi l'incontournable petit livre de Dirac [138]. En pratique, la méthode de quantification utilisée par les physiciens des hautes énergies vise à conserver l'invariance relativiste de manière explicite et ne passent pas par la quantification canonique. On se base sur une approche d'intégrale de chemins complétée par une astuce due à Faddeev et Popov [301] qui permet d'éviter un sur-comptage lié à l'invariance de jauge. Un exposé clair de cette méthode dans le contexte des théories de jauge non abéliennes est présenté dans [216, Chapitre 12] et, de manière plus générale dans [389].

4.1.5 Le spin 1/2

Du spin de l'électron à la sphère de Bloch

L'autre exemple de système décrit par un qubit élémentaire est ce qu'on appelle un spin 1/2. C'est ce qui décrit les degrés de liberté internes d'une particule comme un électron dans une fonction d'onde orbitale donnée.

Les matrices de Pauli sont alors directement reliées au moment angulaire intrinsèque de la particule $S_a = \hbar \sigma_a / 2$ pour $a = x, y, z$. Dans ce cas, le vecteur de Bloch associé à un état $|\psi\rangle$ est directement relié à la valeur moyenne du spin

$$\langle \mathbf{S} \rangle = \frac{\hbar}{2} \mathbf{n} [|\psi\rangle]. \quad (4.29)$$

Cet exemple a un intérêt direct dans le cas de qubits construits sur un moment angulaire $S = 1/2$ comme les moments magnétiques nucléaires du proton ou du noyau ^{13}C , ce qui a été utilisé pour la mise en œuvre d'algorithmes quantiques par résonance magnétique nucléaire.

Voir un qubit comme un spin 1/2 est intéressant car cela permet de discuter les transformations unitaires que l'on peut faire agir sur celui-ci en termes géométriques simples, à savoir des rotations dans l'espace tridimensionnel ordinaire, ce qui est développé dans la section 4.1.3. Comme pour le cas du photon, il est nécessaire de savoir manipuler le spin d'une particule pour pouvoir espérer l'utiliser comme une ressource informationnelle.

Les oscillations de Rabi

Les oscillations de Rabi sont particulièrement pertinentes pour l'information quantique : elles sont utilisées pour piloter nombre de systèmes à deux niveaux utilisés comme qubits. Pour employer un langage qui est celui de la résonance magnétique nucléaire, elles se produisent lorsqu'un spin 1/2 placé dans un

champ magnétique $\mathbf{B}$ selon z est irradié par un champ transverse oscillant à sa fréquence de résonance correspondant à l'énergie $\hbar\omega_0 = \mu\hbar|\mathbf{B}|$ où μ désigne le rapport gyromagnétique de la particule considérée (électron ou noyau). Dans son cours de physique [156], Feynman discute ce phénomène en prenant une molécule d'ammoniac soumise à un champ électrique oscillant. On pourrait également considérer les deux sous-niveaux 2p et 1s de l'atome d'hydrogène soumis à une radiation électromagnétique classique (raie Lyman, transition dipolaire électrique) ou bien les sous-niveaux hyperfins $F = 1$ et $F = 0$ du niveau 1s du même atome (raie à 21 cm, couplage dipolaire magnétique). Le contexte est différent mais c'est le même problème. La dynamique de notre système est décrite par un hamiltonien qui est de la forme

$$H(t) = \frac{\hbar\omega_0}{2} \sigma_z + \hbar\Omega \cos \omega t \sigma_x, \quad (4.30)$$

où Ω est directement proportionnel à l'amplitude de l'onde qui irradie l'objet considéré et ω est sa pulsation. La première étape consiste à supprimer le caractère oscillant de la perturbation extérieure en se plaçant dans ce qu'on appelle le référentiel tournant. On considérera donc un état $|\psi'(t)\rangle = e^{i\omega t \sigma_z/2} |\psi(t)\rangle$ qui obéit alors à l'équation de Schrödinger dépendant du temps

$$i\hbar \frac{d|\psi'(t)\rangle}{dt} = H'(t) |\psi'(t)\rangle \quad (4.31)$$

avec le hamiltonien effectif

$$H'(t) = \frac{\hbar(\omega_0 - \omega)}{2} \sigma_z + \hbar\Omega \cos(\omega t) (\cos \omega t \sigma_x - \sin \omega t \sigma_y) |\psi'(t)\rangle. \quad (4.32)$$

L'approximation clé, dite de l'onde tournante, consiste à ne retenir que les termes séculaires, c'est-à-dire les termes dépendant du temps qui ne se moyennent pas à zéro sur une période $2\pi/\omega$. On est alors ramené à un problème avec un hamiltonien effectif indépendant du temps et donné par

$$H_{\text{Rabi}} = \frac{\hbar}{2} ((\omega_0 - \omega) \sigma_z + \Omega \sigma_x). \quad (4.33)$$

Lorsque l'on est à résonance ($\omega = \omega_0$), on retrouve d'ailleurs le hamiltonien qui décrit le système considéré à la section 1.3.4 du chapitre 1. Ici nous avons une situation un peu plus générale avec un hamiltonien statique qui dans le langage de la résonance magnétique correspond à un champ magnétique effectif arbitraire B_e dans le plan (xz) . Ses deux états propres correspondent aux deux vecteurs de Bloch opposés $\mathbf{n}_{\pm}$, qui sont colinéaires à ce champ,

$$\mathbf{n}_+ = -\mathbf{n}_- = \cos \theta_R \mathbf{e}_z + \sin \theta_R \mathbf{e}_x, \quad (4.34a)$$

$$\cos \theta_R = \frac{\omega_0 - \omega}{\sqrt{\Omega^2 + (\omega_0 - \omega)^2}}, \quad (4.34b)$$

$$\sin \theta_R = \frac{\Omega}{\sqrt{\Omega^2 + (\omega_0 - \omega)^2}}, \quad (4.34c)$$

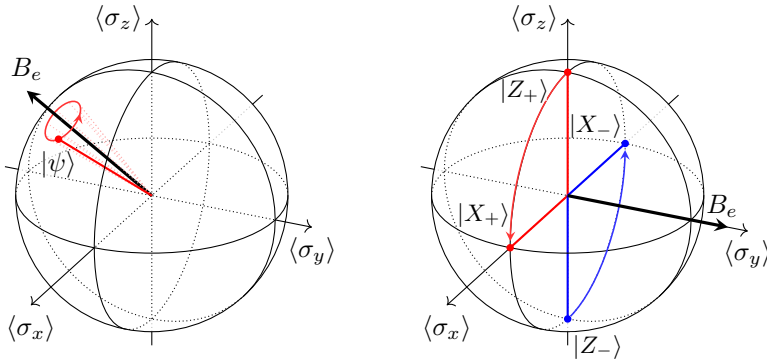


FIGURE 4.5 – Évolution d'un système à deux niveaux sous l'action du hamiltonien H_{Rabi} . L'état quantique suit un cercle sur la sphère autour d'une direction donnée par le hamiltonien, à la manière d'une précession du type précession de Larmor liée à un champ magnétique effectif B_e . Une impulsion $\pi/2$ permet de transformer les états $|Z_{\pm}\rangle$ en $|X_{\pm}\rangle$.

et les énergies correspondantes sont alors $E_{\pm} = \pm \hbar \sqrt{\Omega^2 + (\omega_0 - \omega)^2}/2$. Un calcul direct montre que dans le référentiel tournant, le vecteur de Bloch $\mathbf{n}'(t)$ correspondant à l'état $|\psi'(t)\rangle$ va décrire un cône autour de l'axe du champ effectif B_e à la vitesse angulaire $\Omega_R = \sqrt{\Omega^2 + (\omega_0 - \omega)^2}$, appelée pulsation de Rabi. Ce mouvement de précession, représenté sur la figure 4.5, est tout à fait analogue à la précession de Larmor pour des moments magnétiques classiques et décrit complètement l'évolution de l'état du système quantique irradié.

L'interféromètre de Ramsey

En utilisant une onde classique résonante pendant la bonne durée, on peut donc réaliser très facilement une rotation du vecteur de Bloch autour de l'axe (Oy) avec un angle $\pi/2$. En résonance magnétique, une telle opération est appelée une impulsion $\pi/2$. Elle peut être effectuée en un temps très court pour peu que l'on prenne une onde suffisamment intense : le temps est inversement proportionnel à la pulsation Ω_R qui, à résonance, est directement proportionnelle à l'amplitude de l'onde classique. On peut alors laisser évoluer le spin durant un temps T et ensuite effectuer la rotation inverse et nous pouvons chercher quel est l'état ainsi obtenu. Pour simplifier, nous négligerons les durées des impulsions devant T . La première impulsion $\pi/2$ est décrite par la matrice unitaire

$$U_{\pi/2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \quad (4.35)$$

Durant la phase d'évolution entre les deux impulsions $\pi/2$, l'évolution est décrite par la matrice diagonale $\exp(i\omega_0 T \sigma_z/2)$. Le processus complet est donc

décrit par

$$U_{\text{tot}}(T) = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot e^{i\omega_0 T \sigma_z/2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (4.36)$$

$$= \begin{pmatrix} \cos \omega_0 T/2 & i \sin \omega_0 T/2 \\ i \sin \omega_0 T/2 & \cos \omega_0 T/2 \end{pmatrix}. \quad (4.37)$$

Si nous préparons l'état $|Z_+\rangle$, nous voyons se produire des oscillations dans la probabilité de sortir dans le même état : $P(+, T|+, 0) = \cos^2 \omega_0 T/2$. Ces oscillations permettent d'utiliser ce dispositif comme un interféromètre. En effet, la première impulsion $\pi/2$ envoie le spin initialement dans l'état $|Z_+\rangle$ sur la superposition $(|Z_+\rangle + |Z_-\rangle)/\sqrt{2}$. Ensuite chaque composante de la superposition accumule une phase différente, à savoir $e^{\pm i\pi\omega_0 T/2}$. Enfin, les deux chemins sont recombinaés par la seconde impulsion $\pi/2$. Cette méthode est la base de l'*interférométrie de Ramsey*.

Analogie avec l'optique

Un tel interféromètre existe en optique : c'est l'*interféromètre de Mach-Zehnder* représenté sur la figure 4.6. Sur la plage de longueurs d'onde où on les utilise, chacune des deux lames semi-réfléchissantes $\mathcal{L}_{\text{in}}$ et $\mathcal{L}_{\text{out}}$ est décrite par une matrice de diffusion 2×2 qui contient les amplitudes de réflexion et de transmission. Pour une lame 50/50, cette matrice vaut ²

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (4.38)$$

Dans leur propagation le long du bras α , les ondes monochromatiques de longueur d'onde λ accumulent une phase $2\pi l_\alpha/\lambda$ qui s'exprime en termes de la pulsation ω et du temps de vol τ_α comme $\omega\tau_\alpha$. Un calcul identique à celui effectué précédemment montre alors que la matrice donnant l'amplitude de transition entre les entrées et sorties de l'interféromètre est

$$S = e^{i\omega\bar{\tau}} \begin{pmatrix} \cos(\omega\delta\tau/2) & -\sin(\omega\delta\tau/2) \\ \sin(\omega\delta\tau/2) & \cos(\omega\delta\tau/2) \end{pmatrix} \quad (4.39)$$

où $\bar{\tau}$ désigne la moyenne du temps de vol sur les deux bras et $\delta\tau$ la différence entre ces temps de vol. En pratique, on peut bien sûr modifier les phases accumulées le long des deux bras, et donc les temps de vol effectifs, en y disposant des éléments optiques transparents.

L'interféromètre de Mach-Zehnder est donc un interféromètre à deux chemins où chaque chemin n'est parcouru qu'une fois, à la différence de l'interféromètre de Michelson. De tels interféromètres ont été réalisés pour des atomes et même des molécules [111], mais aussi pour des électrons se propageant

2. Le facteur i est en fait lié à la réflexion d'une onde classique à l'interface entre l'air et le diélectrique de la lame.

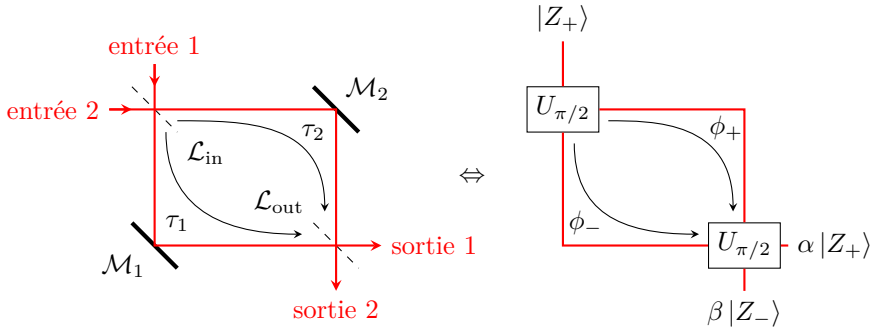


FIGURE 4.6 – Interféromètre de Mach-Zehnder. À *gauche* : un faisceau lumineux est séparé en deux par une lame semi-réfléchissante $\mathcal{L}_{\text{in}}$ en deux trajets 1 et 2 passant par les miroirs $\mathcal{M}_{1,2}$ avant d'être recombinaés par une lame semi-réfléchissante $\mathcal{L}_{\text{out}}$. Le temps de vol sur les deux branches est différent. À *droite* : équivalent optique de l'interféromètre de Ramsey en physique quantique. Une impulsion $\pi/2$ est utilisée pour transformer l'état $|Z_+\rangle$ en $|X_+\rangle = (|Z_+\rangle + |Z_-\rangle)/\sqrt{2}$. Les deux états de cette superposition évoluent alors librement et accumulent une phase différente, avant d'être recombinaés par une deuxième impulsion $\pi/2$.

dans des circuits électriques [219]. Dans le premier cas, la phase relative entre les deux bras peut être modifiée par la mise en rotation du dispositif (effet Sagnac) et dans le cas électronique, par la présence d'un flux magnétique à travers la surface délimitée par les deux trajectoires (effet Aharonov-Bohm, voir section 1.2.4). À l'échelle de la particule unique (photon, atome ou électron), on peut voir l'interféromètre de Mach-Zehnder comme une réalisation contrôlée d'un schéma d'interférences entre deux trajectoires comme les trous d'Young utilisés par Feynman pour introduire et discuter les interférences quantiques [156, tome Mécanique quantique]. Finalement, l'interféromètre de Ramsey [308] présenté précédemment apparaît lui aussi comme interféromètre de Mach-Zehnder. Les chemins ne sont pas associés à des trajectoires différentes dans l'espace réel mais à des trajectoires dans un espace de configurations internes à l'objet considéré comme dans la discussion de la section 1.2.1.

4.2 Oscillateur harmonique quantique

Après cette étude détaillée des états du qubit, il est temps de nous intéresser au deuxième élément fondamental de l'information quantique : l'oscillateur harmonique quantique. Un oscillateur harmonique correspond, dans sa forme la plus simple, au mouvement d'une particule de masse m dans un potentiel quadratique. Cependant, l'importance de ce système est bien plus grande que cela puisqu'on peut s'en servir pour décrire le comportement d'un système

physique autour d'une position d'équilibre stable. Le traitement de ce modèle dans un formalisme quantique paraît donc naturellement important. De plus, il s'agit d'un système pour lequel de nombreuses réalisations expérimentales existent, telles que des particules piégées dans des puits ou les excitations du champ électromagnétique.

Le but de cette section va donc être de mener l'étude d'un oscillateur harmonique dans un cadre quantique, afin de déterminer tout d'abord le spectre qui est associé à ce modèle et les états quantiques correspondants. Puis nous nous intéresserons à une classe d'états bien particuliers, qui présentent un ensemble de propriétés les plaçant à la frontière des mondes quantique et classique, les états cohérents.

4.2.1 États propres et énergies propres

Dans le monde classique, le hamiltonien de l'oscillateur harmonique à une dimension s'écrit

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2, \quad (4.40)$$

où ω est la pulsation de l'oscillateur, p l'impulsion de la particule et q sa position. Bien qu'il s'agisse d'un des rares cas où tout est calculable par l'intégrale de chemins [155], nous retracerons ici la procédure de quantification canonique qui permet d'accéder rapidement aux résultats dont nous aurons besoin pour la suite de cet ouvrage.

Par la procédure de quantification canonique (section 1.3.3), on dispose des opérateurs position q et impulsion p tels que

$$[q, p] = i\hbar \mathbb{1}. \quad (4.41)$$

Ces opérateurs sont associés à deux bases distinctes de l'espace des états pour une particule massive : les états propres de l'opérateur position sont, comme son nom l'indique, les états $|q\rangle$ localisés en un point q de l'espace, et ceux de l'opérateur impulsion sont les états localisés dans l'espace des impulsions.

Commençons par chercher les états propres du hamiltonien du système et ses énergies propres : ce sont les états qui restent stables dans l'évolution temporelle de l'oscillateur. Les opérateurs adimensionnés $P = p/(m\hbar\omega)^{1/2}$ et $Q = q(m\omega/\hbar)^{1/2}$ vérifient la relation de commutation $[Q, P] = i\mathbb{1}$ et permettent d'exprimer le hamiltonien (4.40) sous la forme $H = \frac{\hbar\omega}{2} (P^2 + Q^2)$. Pour avancer, remarquons que l'on peut voir le terme $\frac{1}{2} (P^2 + Q^2)$ comme résultant du calcul d'un module à ceci près qu'on raisonne ici avec des opérateurs qui ne commutent pas. Néanmoins, en définissant les opérateurs dits de *création*

et d'annihilation par

$$a^\dagger = \frac{Q - iP}{\sqrt{2}}, \quad (4.42a)$$

$$a = \frac{Q + iP}{\sqrt{2}}, \quad (4.42b)$$

qui vérifient la relation de commutation

$$[a, a^\dagger] = \mathbb{1}, \quad (4.43)$$

le hamiltonien se réécrit

$$H = \hbar\omega \left(a^\dagger a + \frac{1}{2} \right) = \hbar\omega \left(N + \frac{1}{2} \right), \quad (4.44)$$

où par définition

$$N = a^\dagger a. \quad (4.45)$$

Toute l'analyse se réduit à comprendre l'opérateur N , dit opérateur nombre. D'abord, son spectre est nécessairement positif (c'est en somme un carré). De plus, comme $[N, a] = -a$ et $[N, a^\dagger] = a^\dagger$, ces deux opérateurs permettent respectivement de descendre et de monter dans le spectre car ils diminuent (resp. augmentent) la valeur propre d'une unité. La seule manière de ne pas tomber en descendant sur une valeur propre négative est qu'elles soient toutes entières et que $a|0\rangle = 0$. Par conséquent $\text{Sp}(N) = \mathbb{N}$. La plus petite représentation de l'algèbre $[a, a^\dagger] = \mathbb{1}$ est donc un espace construit à partir d'un état noté $|0\rangle$, qui vérifie $N|0\rangle = 0$, sur lequel on agit en appliquant successivement l'opérateur $a^\dagger$. L'espace propre associé à la valeur propre $n \in \mathbb{N}^*$ est alors de dimension un, engendré par l'état normé $|n\rangle = (a^\dagger)^n |0\rangle / \sqrt{n!}$ de sorte que $a^\dagger |n\rangle = \sqrt{n+1} |n+1\rangle$ pour $n \geq 0$ et $a |n\rangle = \sqrt{n} |n-1\rangle$ pour $n \geq 1$. Ceci assure que les opérateurs a et $a^\dagger$ sont bien adjoints l'un de l'autre. Les énergies propres de l'oscillateur harmonique uni-dimensionnel sont donc quantifiées et même équidistantes les unes des autres :

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right). \quad (4.46)$$

4.2.2 États quasi classiques

Définition et propriétés

Une question naturelle à se poser à partir de la description quantique de l'oscillateur harmonique est celle des états dits quasi classiques qui s'approchent le plus d'une description classique d'un oscillateur harmonique. Ce ne sont

pas les états $|n\rangle$ discutés précédemment car, pour ces derniers, les opérateurs position et impulsion ont des valeurs moyennes nulles :

$$\langle n|Q|n\rangle = \left\langle n \left| \frac{a + a^\dagger}{\sqrt{2}} \right| n \right\rangle = 0. \quad (4.47)$$

Quels sont alors les états qui vont permettre, à tout instant, de redonner des valeurs moyennes du champ non nulles ? Il faut pour cela des états qui assurent une valeur moyenne non nulle à a et $a^\dagger$. Imaginons que l'on parte d'un état $|\psi\rangle$ tel que $\langle a \rangle_{|\psi(0)\rangle} = \alpha \neq 0$. L'équation de Schrödinger montre alors que

$$i\hbar \frac{d\langle a \rangle_{|\psi(t)\rangle}}{dt} = \langle [a, H] \rangle_{|\psi(t)\rangle}. \quad (4.48)$$

En utilisant $[a, H] = \hbar\omega a$, le commutateur se calcule directement et on obtient par intégration

$$\langle a \rangle_{|\psi(t)\rangle} = \langle a \rangle_{|\psi(0)\rangle} e^{-i\omega t} \quad (4.49)$$

et donc que, dans le plan qui représente (aux facteurs d'adimensionnement près) l'espace des phases, le point défini par les valeurs moyennes de P et Q suit une trajectoire circulaire à la vitesse angulaire ω tout comme le ferait l'amplitude d'un oscillateur harmonique classique dont l'énergie serait

$$E_{cl} = \hbar\omega|\alpha|^2. \quad (4.50)$$

Le fait que, pour un oscillateur harmonique quantique, les valeurs moyennes de P et Q évoluent selon les équations du mouvement classique est un résultat dû à Ehrenfest qui ne nous apprend rien sur l'état quantique considéré. Pour sélectionner certains états, nous devons pousser la correspondance plus loin et imposer que l'énergie E_{cl} corresponde à l'énergie quantique nécessaire pour créer un tel état, c'est-à-dire que $\langle \alpha|H|\alpha\rangle - \langle 0|H|0\rangle = \hbar\omega|\alpha|^2$. Cela est équivalent à la condition $\langle \alpha|a^\dagger a|\alpha\rangle = |\alpha|^2$. Or cela implique qu'un tel état est un état propre de l'opérateur annihilation³. En effet :

$$\|(a - \alpha)|\alpha\rangle\|^2 = \langle \alpha|(a^\dagger - \alpha^*)(a - \alpha)|\alpha\rangle = 2|\alpha|^2 - 2|\alpha|^2 = 0. \quad (4.51)$$

On a donc bien démontré que l'état $|\alpha\rangle$ est un état propre de l'opérateur annihilation, avec la valeur propre α :

$$a|\alpha\rangle = \alpha|\alpha\rangle. \quad (4.52)$$

À partir de cette simple propriété, on va pouvoir trouver la représentation de cet état dans la base des états nombres. En effet, en décomposant sur cette base l'état $|\alpha\rangle$, $|\alpha\rangle = \sum_{n=0}^{\infty} c_n(\alpha)|n\rangle$ et en faisant agir l'opérateur annihilation sur cet état, on obtient une relation de récurrence sur les coefficients $c_n(\alpha)$:

3. Il s'agit de la façon dont ces états sont définis habituellement.

$c_{n+1}(\alpha) = \alpha c_n(\alpha) / \sqrt{n+1}$, ce qui montre que $c_n(\alpha) = c_0(\alpha) \alpha^n / \sqrt{n!}$. La condition de normalisation permet finalement d'écrire

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (4.53)$$

Nous voyons que la phase de l'amplitude devant $|n\rangle$ est linéaire en n . C'est la même phase pour chaque excitation élémentaire et c'est pour cette raison que l'on appelle de tels états les *états cohérents*. Remarquons qu'un état cohérent le reste au cours du temps. L'état $|n\rangle$ prenant juste la phase $e^{-i\omega n t}$ sur l'intervalle temporel $[0, t]$, l'équation (4.53) montre directement que, partant de l'état $|\alpha\rangle$, l'état à l'instant t est aussi un état cohérent de paramètre $\alpha(t) = \alpha e^{-i\omega t}$.

Une dernière notion importante dans l'étude des états cohérents est celle d'*opérateur déplacement* de paramètre α défini par

$$D(\alpha) = e^{\alpha a^\dagger - \alpha^* a}. \quad (4.54)$$

En utilisant le fait que $e^{A+B} = e^A e^B e^{-[A,B]/2}$ si A et B commutent avec $[A, B]$, on montre les propriétés suivantes :

$$D(\alpha) = e^{-|\alpha|^2/2} e^{\alpha a^\dagger} e^{-\alpha^* a} \text{ et } D^\dagger(\alpha) = D^{-1}(\alpha) = D(-\alpha). \quad (4.55)$$

Ceci entraîne immédiatement que

$$|\alpha\rangle = D(\alpha) |0\rangle. \quad (4.56)$$

Cet opérateur est unitaire et possède la propriété suivante, qui nous servira dans la prochaine section

$$D^{-1}(\alpha) a D(\alpha) = a + \alpha. \quad (4.57)$$

Cette équation justifie le nom d'opérateur déplacement : Q et P sont respectivement translatés de $\sqrt{2}\Re(\alpha)$ et $\sqrt{2}\Im(\alpha)$. Les états cohérents sont donc obtenus à partir du vide $|0\rangle$ par application d'un opérateur qui les translate dans l'espace des phases. Enfin, nous pouvons calculer le produit de deux opérateurs déplacements, ce qui donne

$$D(\alpha) D(\beta) = e^{i\Im(\alpha \beta^*)} D(\alpha + \beta). \quad (4.58)$$

Cela montre que l'on peut passer d'un état cohérent à un autre à l'aide des opérateurs de déplacement à une phase près. Celle-ci provient du fait que les opérateurs $D(\alpha)$ représentent le groupe des translations à une phase près⁴. C'est une propriété centrale de la théorie quantique liée au fait que la phase globale n'est pas pertinente pour la définition d'un état quantique.

4. On parle de représentation projective.

Incertitudes et fonction d'onde associée

Une propriété intéressante des états cohérents concerne leur lien avec la relation d'indétermination de Heisenberg. En effet, si on cherche à calculer la quantité $\Delta q \Delta p$ pour un état cohérent, on va trouver le résultat minimal autorisé par la théorie quantique, et ce quelle que soit la valeur du paramètre α . Pour le démontrer, il suffit d'évaluer $(\Delta q)^2 = \left(\langle \alpha | q^2 | \alpha \rangle - \langle \alpha | q | \alpha \rangle^2 \right) = \hbar/2m\omega$. De même, on a $(\Delta p)^2 = \hbar m\omega/2$. Mises ensemble, ces équations montrent bien que l'inégalité de Heisenberg (1.54) est saturée pour les états cohérents

$$\Delta q \Delta p = \frac{\hbar}{2}. \quad (4.59)$$

Ce résultat montre que les états cohérents sont donc les états les plus classiques que l'on peut créer.

Une dernière question d'intérêt autour de ces états est celle de la fonction d'onde qui leur est associée. En effet, puisque nous avons trouvé les états cohérents en cherchant des états dont la position moyenne était non nulle, il peut être intéressant de regarder comment ces états se comportent en représentation position, c'est-à-dire de trouver leur fonction d'onde. Par définition, on a

$$a |\alpha\rangle = \alpha |\alpha\rangle \quad \leftrightarrow \quad \sqrt{\frac{m\omega}{2\hbar}} \left(q + \frac{\hbar}{m\omega} \frac{\partial}{\partial q} \right) \psi_\alpha(q) = \alpha \psi_\alpha(q). \quad (4.60)$$

La solution de cette équation différentielle est une fonction d'onde gaussienne, donnée par

$$\psi_\alpha(q) = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}} \exp \left(-\frac{m\omega}{2\hbar} \left(q - \sqrt{\frac{2\hbar}{m\omega}} \alpha \right)^2 \right). \quad (4.61)$$

À l'aide de cette équation, on voit facilement qu'un état cohérent est similaire à l'état du vide, mais qui est déplacé de telle sorte que son centre soit désormais en α au lieu de 0. C'est cette vision d'états gaussiens qui va nous permettre de représenter les états cohérents facilement dans l'espace des phases, comme sur la figure 4.7.

4.2.3 Photons dans une cavité

L'une des grandes découvertes de la physique quantique du début du xx^e siècle est celle du photon en tant que particule associée au champ électromagnétique. Dans une conception plus moderne, cette particule est une excitation du champ électromagnétique quantique. À impulsion et polarisation bien définies, ces excitations sont celles d'un oscillateur harmonique quantique. Cela provient du fait qu'il est possible de décrire la dynamique du champ électromagnétique classique en termes d'oscillateurs harmoniques *via* les modes

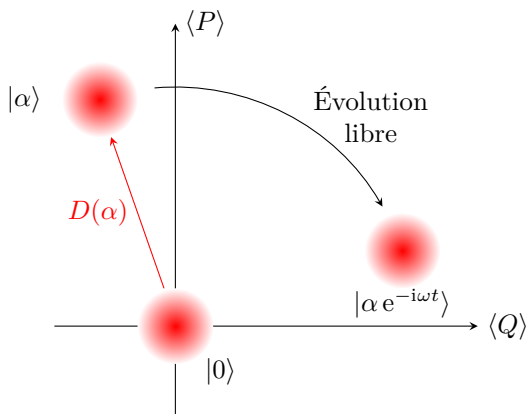


FIGURE 4.7 – Représentation de différents états cohérents dans le plan $(\langle Q \rangle, \langle P \rangle)$. Dans ce plan, un état cohérent α est représenté par une gaussienne d'écart type $1/2$, centrée sur $(\Re(\alpha), \Im(\alpha))$. Pour passer du vide à l'état cohérent $|\alpha\rangle$, on applique l'opérateur $D(\alpha)$, qui correspond dans ce plan à une translation. Au cours du temps, un état cohérent évolue en suivant une rotation dans le sens direct avec pulsation ω .

normaux du champ [105]. Lorsque l'on considère un champ dans une cavité parfaitement conductrice, ces modes sont bien espacés et la structure spatiale des champs électromagnétiques est déterminée par les conditions aux limites [365].

Dans le cas d'une cavité fermée et parfaite, les modes étant bien isolés, il est possible de considérer un seul mode du champ dont la pulsation est égale à la pulsation de résonance de la cavité. Après quantification, ce mode du champ est caractérisé par un état fondamental $|0\rangle$, aussi nommé vide, correspondant à l'absence de photon dans ce mode. Les autres états correspondent à des états excités obtenus grâce aux opérateurs de création et d'annihilation $a^\dagger$ et a du mode. Parmi eux, les états propres du hamiltonien (1.49) correspondent à des états avec un nombre fixé de photons. Le champ électrique et le potentiel vecteur en un point donné (comme le centre de la cavité)⁵ sont alors équivalents (à des facteurs multiplicatifs près) aux opérateurs q et p du cas classique⁶,

$$\mathbf{E} \propto (a + a^\dagger) \quad \mathbf{A} \propto i(a - a^\dagger). \quad (4.62)$$

L'étude et la manipulation des états quantiques d'un mode du champ électromagnétique en cavité constitue l'électrodynamique quantique en cavité. Ce domaine de recherche, exploré entres autres par l'équipe de S. Haroche [194],

5. En toute généralité, ces champs possèdent une structure spatiale et une polarisation. Ceux-ci n'interviennent pas lorsqu'on s'intéresse aux champs en un point donné.

6. On s'est ici placé en jauge de Coulomb. Dans le formalisme hamiltonien, les champs $\mathbf{E}$ et $\mathbf{A}$ transverses sont alors conjugués l'un de l'autre.

constitue un véritable laboratoire pour l'ingénierie quantique et la manipulation de l'information quantique que nous discuterons au chapitre 10.

Ayant identifié la notion de photon comme une excitation élémentaire du champ quantifié, essayons d'identifier les configurations classiques du champ qui sont solutions des équations de Maxwell. L'identification entre modes du champ et oscillateurs montre, au travers de la discussion de la section 4.2.2, que les états cohérents correspondent aux configurations classiques du champ en donnant des valeurs moyennes non nulles à $\mathbf{A}$ et $\mathbf{E}$. Par ailleurs, en l'absence de matière, elles obéissent aux équations classiques du mouvement.

Regardons comment générer ces états cohérents. Pour cela, partons du vide quantique $|0\rangle$ et cherchons quel est l'état quantique engendré par un courant classique $\mathbf{j}$ dépendant du temps. Le couplage entre le champ et le courant classique dépendant du temps que l'on impose est un terme linéaire de la forme $-\mathbf{j}(t) \cdot \mathbf{A}$. Pour un seul mode, le courant agit au travers de sa projection sur la structure spatiale du mode considéré, et apparaît donc au travers d'un scalaire dépendant du temps $j(t)$. Le hamiltonien qui décrit le mode couplé au courant classique est alors :

$$H(t) = \hbar\omega \left(a^\dagger a + \frac{1}{2} \right) - \gamma j(t)(a - a^\dagger), \quad (4.63)$$

où γ est le facteur multiplicatif liant A à a et $a^\dagger$. Cet hamiltonien est en fait un cas particulier d'un problème plus large dans lequel l'oscillateur est couplé linéairement à un forçage extérieur décrit par un paramètre complexe dépendant du temps $\lambda(t)$:

$$H(t) = \frac{\hbar\omega}{2} (a^\dagger a + a a^\dagger) - \hbar(\lambda(t) a^\dagger + \lambda^*(t) a). \quad (4.64)$$

En nous inspirant de ce que nous avons fait pour trouver les états quasi classiques de l'oscillateur harmonique, regardons alors l'évolution de $a(t) = U[0, t]^\dagger a U[0, t]$ où $U[0, t]$ désigne l'opérateur évolution associé au hamiltonien $H(t)$. Cet opérateur vérifie

$$i\hbar \frac{da(t)}{dt} = [a, H] = \hbar(\omega a(t) - \lambda(t)). \quad (4.65)$$

La solution de cette équation est donnée par

$$a(t) = a(0) e^{-i\omega t} + i \int_0^t e^{-i\omega(t-s)} \lambda(s) ds. \quad (4.66)$$

Ceci montre directement que, en partant d'un état cohérent de paramètre α , l'évolution sous l'action du hamiltonien $H(t)$ produit un état cohérent avec comme paramètre

$$\alpha(t) = \alpha e^{-i\omega t} + i \int_0^t e^{-i\omega(t-s)} \lambda(s) ds. \quad (4.67)$$

Cependant, ce n'est pas tout : nous n'avons déterminé l'état qu'à une phase près. Cette dernière est bien présente car l'opérateur évolution $U[0, t]$ peut être obtenu en composant des opérateurs déplacements. Un calcul direct que nous laissons en exercice au lecteur montre que

$$|\psi(t)\rangle = e^{i\Theta(t)} |\alpha(t)\rangle \quad (4.68a)$$

$$\Theta(t) = \Im \left(\int_0^t ds \int_0^s e^{i\omega(s-s')} \lambda(s) \lambda^*(s') ds' \right). \quad (4.68b)$$

Cette discussion montre comment l'application d'un courant électrique classique entraîne la création d'un état cohérent du champ électromagnétique quantique. On pourrait penser que l'on a retrouvé l'électrodynamique classique au sein de l'électrodynamique quantique mais il reste encore un point à éclaircir : pourquoi n'observe-t-on pas des superpositions de tels états quasi classiques dans notre vie quotidienne ? Ceci ne peut être compris que grâce au processus de décohérence, exposé au chapitre 9.

4.3 Tomographie quantique du qubit

En médecine, la *tomographie* par rayons X (scanner) permet de reconstruire la densité des tissus en trois dimensions d'un patient au moyen de multiples mesures d'absorption. Le problème de la tomographie quantique est analogue et cherche à savoir comment reconstruire un état quantique à partir de mesures effectuées sur des réalisations physiques de cet état. Une réalisation est tout simplement un système physique préparé dans l'état $|\psi\rangle$. Par rapport au cas classique où le fait de multiplier les mesures sur une réalisation ne va pas (en première approximation) altérer les propriétés du système, ici, on aura besoin d'un nombre N de réalisations pour reconstruire l'état $|\psi\rangle$ *a priori* inconnu.

Le problème de l'estimation consiste à quantifier la performance du processus en fonction du nombre de réalisations disponibles. Y a-t-il un optimum ? Comment varie-t-il en fonction de N ?

4.3.1 Reconstruire $|\psi\rangle$ avec une infinité de réalisations

On dispose d'une source capable de produire autant de réalisations d'un état quantique $|\psi\rangle$ d'un qubit que nécessaire. Cela peut être une source de photons uniques avec tous le même état de polarisation. Que faut-il pour reconstruire $|\psi\rangle$ ou, de manière équivalente, le vecteur de Bloch $\mathbf{n}[|\psi\rangle]$?

Avoir un nombre potentiellement infini de mesures permet de reconstruire la valeur moyenne d'une observable. Ainsi, pour $n_z = \langle |\uparrow\rangle\langle\downarrow| - |\leftrightarrow\rangle\langle\leftrightarrow| \rangle_{|\psi\rangle}$, on peut reconstruire l'état à une symétrie de rotation autour de l'axe (Oz) près, ce qui correspond à la symétrie laissant l'observable invariante. Avec cette mesure, on a un ensemble d'états qui, génériquement, forme un cercle sur la sphère de Bloch comme illustré sur la partie gauche de la figure 4.8. Grâce à

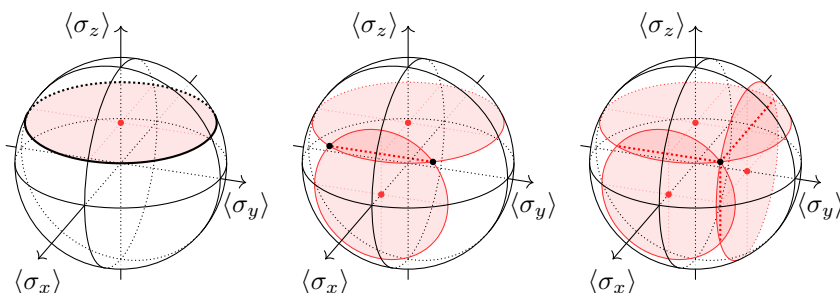


FIGURE 4.8 – Reconstruction tomographique à partir d’une source parfaite. À partir de la mesure d’une observable, on obtient un cercle sur la sphère de Bloch, correspondant à la symétrie qui laisse l’observable mesurée invariante. Au milieu, après la mesure de deux observables, on obtient deux états possibles pour le qubit. Après la mesure de trois observables, on peut déterminer l’état du qubit.

la mesure de deux observables distinctes, on récupère l’intersection de deux cercles, c’est-à-dire deux états comme le montre la partie droite de la figure 4.8. On comprend en fin de compte que trois observables sont nécessaires pour une reconstruction de l’état d’un qubit par une mesure de valeur moyenne.

4.3.2 Le cas d’une source imparfaite

En pratique, une expérience de physique ne produit pas forcément toujours le même état pur mais fabrique une famille aléatoire d’états que l’on note $(|\psi_\alpha\rangle)$, chacun ayant la probabilité p_α d’être obtenu. Une telle source est dite imparfaite ou encore statistique puisqu’elle ne nous fournit pas un unique état pur. On reviendra sur cette dénomination dans le prochain chapitre.

Le vecteur de Bloch fut construit pour un état pur, mais la construction de la sphère de Bloch montre que pour un ensemble d’états $(|\psi_\alpha\rangle, p_\alpha)$, on peut définir un vecteur de Bloch en prenant le barycentre des états $\mathbf{n}[|\psi_\alpha\rangle]$ avec les poids p_α

$$\mathbf{n}[|\psi_\alpha\rangle, p_\alpha] = \sum_{\alpha} p_{\alpha} \mathbf{n}[|\psi_{\alpha}\rangle] . \quad (4.69)$$

Nous avons donc la vision géométrique suivante pour les sources réalistes d’états d’un qubit : la *boule de Bloch* correspond aux états *mélanges* créés par une source imparfaite. La sphère de Bloch correspond toujours bien sûr aux états purs⁷. En pratique, cette boule caractérise toutes les valeurs moyennes de toutes les observables à un qubit.

7. Notons que cette vision géométrique montre clairement qu’il nous faut bien trois observables pour reconstruire complètement ce qui sort de la source. Une observable nous réduit à un disque (associé à des états mélanges), deux observables à un segment (toujours associé à des mélanges) et trois observables nous réduisent à un unique état mélange.

Cependant, remarquons que les éléments de cette boule de Bloch peuvent être obtenus de nombreuses façons. Il n'y a pas unicité de la décomposition d'un état mélange comme mélange statistique d'états quantiques. En revanche, si on considère la décomposition sous forme d'un mélange de deux états orthogonaux, on a en général unicité. En effet, deux états orthogonaux constituent deux points diamétralement opposés de la sphère de Bloch. Les états mélanges sont alors situés sur le segment qui rejoint ces deux points. Mais justement, prenons un point au hasard à l'intérieur de la boule de Bloch à l'exception du centre de celle-ci. Il existe alors un unique couple d'états orthogonaux dont c'est un mélange statistique ; il suffit de prendre la droite qui joint le centre de la boule au mélange que l'on considère. Les deux points d'intersection de cette droite avec la sphère constituent le couple d'états purs orthogonaux recherchés. Le centre de la boule de Bloch constitue une exception puisque c'est le mélange 50/50 de n'importe quel paire d'états orthogonaux.

4.3.3 Nombre fini de réalisations

Nous allons maintenant traiter le cas plus réaliste où le nombre de réalisations est fini. C'est un cas important pour bien comprendre les subtilités du problème de l'estimation d'un état quantique.

Nous allons considérer $N \geq 1$ réalisations physiques d'un état d'un qubit (état pur ou état mélange en toute généralité) et nous voulons formuler le problème de l'estimation de cet état. Commençons par le cas d'un état pur $|\psi\rangle$ dont on reçoit N réalisations. À l'issue de mesures sur ces N réalisations, on va produire un *estimateur* sous la forme d'un état quantique $|\psi_{\text{est}}[R]\rangle$ qui dépend des résultats de mesure R . On définit la *qualité de l'estimateur* comme le carré du module du recouvrement entre la fonction d'onde $|\psi_{\text{est}}[R]\rangle$ estimée et $|\psi\rangle$. Ceci représente la probabilité que l'état inconnu soit « présent » dans l'estimateur

$$|\langle\psi_{\text{est}}[R]|\psi\rangle|^2 = \text{Qualité de l'estimateur} . \quad (4.70)$$

Évidemment, l'algorithme d'estimation donne l'estimateur en fonction des résultats des mesures effectuées. La qualité de l'estimation de l'état $|\psi\rangle$ est donnée en moyennant sur l'ensemble des résultats de mesure

$$Q(\text{Estimation de } |\psi\rangle) = \mathbb{E}_R \left(|\langle\psi_{\text{est}}[R]|\psi\rangle|^2 \right) . \quad (4.71)$$

La notation $\mathbb{E}_R(\cdot)$ signifie que l'on prend une moyenne sur les résultats de mesure R dont la distribution de probabilité dépend bien évidemment de l'état $|\psi\rangle$. En conséquence, la valeur de cet indicateur de qualité dépend aussi de cet état.

Enfin, comme on veut une estimation de la qualité de tout le processus d'estimation (mesures et algorithme qui détermine l'estimateur à partir des résultats des mesures), on moyenne sur tous les états possibles du qubit,

c'est-à-dire sur la sphère de Bloch. L'estimateur de la qualité de l'algorithme d'estimation que nous noterons Est_N est alors donné par

$$\begin{aligned} Q(\text{Est}_N) &= \mathbb{E}_{|\psi\rangle} (Q(\text{Estimation de } |\psi\rangle)) \\ &= \mathbb{E}_{|\psi\rangle} \left(\mathbb{E}_R (|\langle \psi_{\text{est}}[R] | \psi \rangle|^2) \right), \end{aligned} \quad (4.72)$$

où cette fois $\mathbb{E}_{|\psi\rangle}$ signifie que l'on moyenne sur tous les états physiques, ce qui dans le cas d'un qubit revient à moyenner sur la sphère de Bloch.

Les questions qui se posent maintenant sont les suivantes. Pour un N fixé, la qualité de l'estimation est-elle bornée par $\phi_N \leq 1$ et que vaut cette borne ϕ_N ? Si elle existe, comment atteindre cette borne, par quelle stratégie de mesure et quel algorithme d'estimation?

Cas $N = 1$

Imaginons que l'on ait un seul qubit isolé dans un état $|\psi\rangle$ inconnu. La seule chose que l'on puisse faire est de faire une mesure selon un axe. Ceci donne un bit classique d'information.

Prenons l'exemple d'un photon passant à travers un polariseur vertical. Si le photon passe, le choix naturel (à définir) consiste à dire que le qubit est polarisé dans la direction du polariseur. S'il ne passe pas, on prend l'état qui est orthogonal. Sur la sphère de Bloch, la direction de mesure est donnée par le vecteur $\mathbf{n}_M$. Si le qubit est dans un état $|\psi\rangle$ pur, associé au vecteur de Bloch $\mathbf{n}$, les probabilités des deux résultats de mesure sont alors données par

$$p_M(\text{Passe}|\mathbf{n}) = |\langle \mathbf{n} | \mathbf{n}_M \rangle|^2 = \frac{1}{2} (1 + \mathbf{n} \cdot \mathbf{n}_M) \quad (4.73a)$$

$$p_M(\text{Ne passe pas}|\mathbf{n}) = |\langle \mathbf{n} | -\mathbf{n}_M \rangle|^2 = \frac{1}{2} (1 - \mathbf{n} \cdot \mathbf{n}_M). \quad (4.73b)$$

Regardons maintenant la qualité d'estimation dans l'état $|\mathbf{n}\rangle$. Nous sommes dans un cas où l'algorithme d'estimation peut générer deux états possibles. En moyennant sur les deux possibilités de mesure (le photon passe ou ne passe pas à travers le polariseur), on obtient donc un mélange statistique de deux états : $|\mathbf{n}_M\rangle$ avec la probabilité $p_M(\text{Passe}|\mathbf{n})$ et $|\mathbf{n}_M^\perp\rangle$ avec la probabilité $p_M(\text{Ne passe pas}|\mathbf{n})$. La qualité de l'estimateur dans l'état $|\psi\rangle$ s'écrit en fonction des vecteurs de Bloch associés à l'état inconnu et au choix de la polarisation mesurée

$$Q(\text{Est}_1, |\psi\rangle) = \frac{1}{2} + \frac{(\mathbf{n} \cdot \mathbf{n}_M)^2}{2}. \quad (4.74)$$

Notons que cet estimateur dépend du choix de $\mathbf{n}$ et de $\mathbf{n}_M$ et vaut un pour les deux états $|\pm \mathbf{n}_M\rangle$, ce qui revient à dire que l'estimateur est parfaitement adapté à les reconnaître. Tous les autres états sont moins bien reconnus. Nous pouvons alors obtenir l'estimateur de qualité moyenné sur tous les états

en moyennant sur toute la sphère de Bloch. Le résultat que l'on obtient est optimal pour $N = 1$ (mais nous n'allons pas le montrer)

$$Q(\text{Est}_1) = \mathbb{E}_{\mathbf{n}} \left(\frac{1}{2} + \frac{(\mathbf{n} \cdot \mathbf{n}_M)^2}{2} \right) = \frac{1}{2} + \frac{1}{2} \int_{S^2} \cos^2 \theta \sin \theta \frac{d\theta d\phi}{4\pi} = \frac{2}{3} \quad (4.75)$$

On voit donc ici très concrètement que la donnée d'une seule réalisation d'un état quantique inconnu d'un qubit ne permet pas de reconstruire celui-ci parfaitement.

Il est intéressant de tester d'autres protocoles. Le protocole le plus mauvais consisterait à prendre le vecteur opposé à celui donné par le protocole optimal. En clair, quand on reçoit un photon d'une polarisation donnée, verticale par exemple, on dit qu'il est de la polarisation opposée, c'est-à-dire horizontale dans cet exemple. Le lecteur pourra reprendre le calcul et montrera que le pire protocole a une qualité égale à $1/3$. Enfin, on peut choisir un protocole complètement aléatoire qui consiste à tirer aléatoirement l'état $\mathbf{n}_M$ ou $-\mathbf{n}_M$. La qualité de l'estimation est alors donnée par $1/2$. Comme quoi, même dans le monde quantique, il y a une justice : le protocole intelligent est plus efficace que le protocole aléatoire qui est lui même plus efficace que le protocole le plus stupide.

Cas $N > 1$

Le résultat principal de l'estimation d'un état à partir d'un nombre fini de mesures est dû à Massard et Popescu [266]⁸ et s'énonce ainsi :

Théorème 4.1 (Massard & Popescu). *Pour la tomographie de l'état d'un qubit à partir de N réalisations, la qualité de l'estimateur est bornée par*

$$Q(\text{Est}_N) \leq \frac{N+1}{N+2}. \quad (4.76)$$

Un point très important est que cette borne n'est pas atteinte pour des mesures indépendantes effectuées sur les différentes réalisations mais par une mesure collective. C'est assez typique de ce qui se passe en quantique mais on ne peut donner le protocole ici, n'ayant pas encore discuté de la notion d'intrication.

Pour aller plus loin sans entrer dans le protocole collectif, nous allons nous concentrer sur la limite $N \rightarrow \infty$ avec des mesures effectuées indépendamment sur les N réalisations.

Limite $N \rightarrow \infty$: tomographie

Notre protocole repose sur l'idée que nous avons exposée plus haut qui nécessitait d'avoir trois observables distinctes. Pour $N \gg 1$, nous allons donc

8. Le problème de l'estimation à partir de N réalisations de l'état d'un qubit est discuté dans le cours de Serge Haroche au collège de France en 2010.

effectuer $n = \frac{N}{3}$ mesures relatives à une observable dans une direction donnée, n dans une direction orthogonale et n dans la troisième direction sur la sphère de Bloch. On appellera naturellement ces trois directions les directions de polarisation. Les mesures avec le polariseur $\mathbf{n}_z$ vont nous donner une estimation n_z de la composante du vecteur de Bloch selon (Oz) . Mais pour un jeu de résultats donné (qui est une succession de « passe » et de « ne passe pas »), on aura un estimateur. Sur l'ensemble des résultats possibles, ces estimateurs seront dispersés et nous devons calculer cette dispersion afin de caractériser le mélange statistique d'états estimés ainsi produits. Celui-ci nous permettra de trouver la qualité du protocole relative à un état inconnu $|\psi\rangle$.

Pour l'état $|\psi\rangle$ décrit par le vecteur de Bloch $\mathbf{n} = n_x \mathbf{e}_x + n_y \mathbf{e}_y + n_z \mathbf{e}_z$, les probabilités que le photon passe ou ne passe pas au travers d'un polariseur selon Z sont

$$p_z(\text{Passe}|\psi) = \frac{1 + n_z}{2} \text{ et } p_z(\text{Ne passe pas}|\psi) = \frac{1 - n_z}{2}. \quad (4.77)$$

Pour n résultats, la probabilité d'avoir n_+ résultats « passe » est donnée par une loi binomiale

$$\binom{n}{n_+} \left(\frac{1 + n_z}{2}\right)^{n_+} \left(\frac{1 - n_z}{2}\right)^{n - n_+}. \quad (4.78)$$

Qu'en est-il maintenant de l'estimateur ? Avec un nombre n_+ de « passe » et $n - n_+$ de « ne passe pas », on estimerait les probabilités à partir des fréquences statistiques de ces résultats de mesure

$$p_{\text{Est}}(\text{Passe}|\psi) = \frac{n_+}{n} \text{ et } p_{\text{Est}}(\text{Ne passe pas}|\psi) = \frac{n - n_+}{n}. \quad (4.79)$$

Cela conduit naturellement à définir l'estimation de n_z par

$$n_z^{\text{Est}} = \frac{2n_+}{n} - 1 = \frac{n_+ - (n - n_+)}{n}. \quad (4.80)$$

Connaissant la loi de probabilité de n_+ à n fixé, on connaît donc la moyenne et la fluctuation de l'estimateur n_z^{Est} . Reste à l'estimer dans la limite où n est grand. Notons $p = \frac{1+n_z}{2}$; nous avons donc

$$p(n_+) = \frac{n!}{n_+! (n - n_+)!} p^{n_+} (1 - p)^{n - n_+}. \quad (4.81)$$

Posons $n_+ = xn$, $x \in [0, 1]$ et cherchons à estimer la loi de x , en utilisant le fait que n et n_+ sont grands. La formule de Stirling, $\ln n! \simeq n \ln n - n$, permet d'obtenir

$$\ln p(n_+) \simeq -nx \ln \frac{x}{p} - n(1 - x) \ln \frac{1 - x}{1 - p}. \quad (4.82)$$

Nous obtenons donc

$$p(n_+) \approx e^{-nF_p(x)} \text{ avec } F_p(x) = x \ln \frac{x}{p} + (1 - x) \ln \frac{1 - x}{1 - p}, \quad (4.83)$$

où $F_p(x)$ est appelée *fonctionnelle de grande déviation*. Typiquement, cette fonction atteint un minimum pour $x = p$, ce qui correspond à la valeur la plus probable de x . Le développement de Taylor en $x - p$ nous permet alors d'accéder aux fluctuations de x

$$F_p(x) = \frac{(x-p)^2}{2p(1-p)} + \mathcal{O}((x-p)^3). \quad (4.84)$$

De là, on tire une approximation gaussienne de la loi de x mais aussi et surtout, les valeurs de la moyenne et de l'écart type de x

$$\mathbb{E}(x) = \frac{1+n_z}{2} \text{ et } \mathbb{E}((x - \mathbb{E}(x))^2) = \frac{1-n_z^2}{4n}. \quad (4.85)$$

On en déduit alors les fluctuations et la valeur moyenne des estimations $n_z^{\text{Est}} = 2x - 1$

$$\mathbb{E}(n_z^{\text{Est}}) = n_z \text{ et } \mathbb{E}((n_z^{\text{Est}} - n_z)^2) = \frac{1-n_z^2}{n}. \quad (4.86)$$

Notons que la fluctuation est d'autant plus petite que le nombre de mesures $n = N/3$ est grand. Remarquons aussi que la fluctuation est minimale lorsque l'état $|\mathbf{n}\rangle$ est proche des états $|Z_{\pm}\rangle$. Cela n'est pas surprenant et reflète que les états situés à l'équateur de la sphère de Bloch sont ceux pour lesquels l'ambiguïté à les distinguer en mesurant σ_z est la plus grande alors que cette observable discerne à coup sûr entre les deux états $|Z_{\pm}\rangle$. Nous allons maintenant utiliser ce résultat pour effectuer le calcul de l'efficacité de l'estimation tomographique.

Rappelons que le protocole d'estimation nous fournit des états purs décrits par des vecteurs de Bloch dont on connaît la dispersion. Plus précisément

$$\mathbf{n}[R] = \mathbf{n} + \delta\mathbf{n}[R], \quad (4.87)$$

où la valeur moyenne des vecteurs de Bloch estimés $\mathbf{n}[R]$ est bien le vecteur de Bloch de l'état inconnu $\mathbf{n}$. De plus, les fluctuations $\delta\mathbf{n}[R]$ sont telles que $\mathbb{E}_R(\delta n_i \delta n_j) = 0$ pour $i \neq j$ car les observables choisies sont mesurées sur des sous-échantillons différents et les fluctuations sont donc décorréliées. Le calcul qui précède nous donne pour terminer $\mathbb{E}_R(\delta n_i^2) = 3[1 - n_z^2]/N$. En utilisant l'expression du module au carré du produit scalaire de deux états associés à deux vecteurs de Bloch $\mathbf{n}_{1,2}$,

$$|\langle\psi_1|\psi_2\rangle|^2 = \frac{1 + \mathbf{n}_1 \cdot \mathbf{n}_2}{2} \quad (4.88)$$

pour les vecteurs de Bloch $\mathbf{n}$ et $\mathbf{n}[R] = \frac{\mathbf{n} + \delta\mathbf{n}[R]}{\|\mathbf{n} + \delta\mathbf{n}[R]\|}$. Nous avons donc

$$|\langle\psi[R]|\psi\rangle|^2 = 1 + \frac{1}{2} \left(\frac{1 + \mathbf{n} \cdot \delta\mathbf{n}[R]}{\|\mathbf{n} + \delta\mathbf{n}[R]\|} - 1 \right). \quad (4.89)$$

On va calculer cela en utilisant les fluctuations à l'ordre deux de $\delta \mathbf{n}[R]$. La seule difficulté technique consiste à calculer l'inverse de la norme à l'ordre deux

$$\frac{1}{\|\mathbf{n} + \delta \mathbf{n}\|} = 1 - \delta \mathbf{n} \cdot \mathbf{n} + \frac{3}{2} (\delta \mathbf{n}' \cdot \mathbf{n})^2 - \frac{1}{2} (\delta \mathbf{n})^2 + \dots \quad (4.90)$$

Ainsi, en remontant à l'expression apparaissant dans le produit scalaire, nous avons, à l'ordre deux

$$\frac{1 + \mathbf{n} \cdot \delta \mathbf{n}}{\|\mathbf{n} + \delta \mathbf{n}\|} \approx 1 - \frac{1}{2} (\delta \mathbf{n})^2 + \frac{1}{2} (\delta \mathbf{n} \cdot \mathbf{n})^2. \quad (4.91)$$

Donc

$$|\langle \psi[R] | \psi \rangle|^2 \simeq 1 + \frac{1}{4} \left[(\delta \mathbf{n} \cdot \mathbf{n})^2 - (\delta \mathbf{n})^2 \right]. \quad (4.92)$$

Puisque

$$(\delta \mathbf{n} \cdot \mathbf{n})^2 = \sum_{i \neq j} n_i n_j \delta n_i \delta n_j + \sum_i (n_i)^2 (\delta n_i)^2, \quad (4.93)$$

nous avons, en se rappelant que les fluctuations sont décorréliées,

$$\mathbb{E}_R \left(|\langle \psi[R] | \psi \rangle|^2 \right) \approx 1 - \frac{1}{4} \sum_i \left\langle (\delta n_i)^2 \right\rangle \left(1 - (n_i)^2 \right) \quad (4.94a)$$

$$\approx 1 - \frac{1}{4} \sum_i \frac{3}{N} \left(1 - (n_i)^2 \right)^2, \quad (4.94b)$$

soit finalement

$$\mathbb{E}_R \left(|\langle \psi[R] | \psi \rangle|^2 \right) \approx 1 - \frac{3}{4N} \sum_i [1 - n_i^2]^2. \quad (4.95)$$

Moyennons maintenant sur la sphère de Bloch

$$\mathbb{E}_{\mathbf{n}} \left(\mathbb{E}_R \left[|\langle \psi[R] | \psi \rangle|^2 \right] \right) \approx 1 - \frac{9}{4N} \mathbb{E}_{\mathbf{n}} \left([1 - n_z^2]^2 \right) \quad (4.96)$$

et nous avons

$$\mathbb{E}_{\mathbf{n}} \left([1 - n_z^2]^2 \right) = \frac{1}{4\pi} 2\pi \int_0^\pi (1 - \cos^2 \theta)^2 \sin \theta d\theta = \frac{8}{15}. \quad (4.97)$$

D'où finalement

$$\mathbb{E}_{\mathbf{n}} \left(\mathbb{E}_R \left[|\langle \psi[R] | \psi \rangle|^2 \right] \right) \underset{N \rightarrow \infty}{\simeq} 1 - \frac{6}{5N} + \mathcal{O} \left(\frac{1}{N^2} \right) \quad (4.98)$$

pour le protocole de tomographie basé sur la mesure de trois observables.

Comme on s'y attendait, la qualité tend vers 1 pour un nombre de réalisation tendant vers l'infini. En revanche on remarquera que le résultat obtenu est strictement moins bon que la borne optimale

$$\frac{N+1}{N+2} \underset{N \rightarrow \infty}{\simeq} 1 - \frac{1}{N} + \mathcal{O}\left(\frac{1}{N^2}\right). \quad (4.99)$$

Cela reflète le fait, déjà mentionné, qu'effectuer des mesures sur les réalisations de manière indépendante n'est pas la meilleure façon de faire. En effectuant des mesures collectives, c'est-à-dire portant sur l'ensemble des N réalisations de l'état $|\psi\rangle$, nous verrons que l'on peut atteindre un meilleur résultat, un peu comme si, en physique quantique, le tout contenait fondamentalement plus d'information que les parties. Mais cette discussion nécessite d'avoir étudié les états d'un système composé de plusieurs parties. C'est précisément en explorant les conséquences du principe de superposition pour de tels systèmes dès le prochain chapitre que nous allons rentrer dans le vif du sujet de cet ouvrage.

Chapitre 5

États intriqués : définition et propriétés générales

La manifestation la plus spectaculaire de la superposition des amplitudes apparaît lorsque l'on considère des systèmes quantiques composites préparés dans une classe particulière d'états physiques appelés états intriqués. Ces états ont des propriétés complètement différentes de celles que notre intuition classique nous ferait imaginer. Ce phénomène, appelé intrication quantique, est une spécificité de la théorie quantique, et n'a donc pas d'équivalent classique.

L'intrication quantique entraîne qu'il est génériquement impossible d'affecter un état quantique bien déterminé à chacune des parties du système composé préparé dans un état quantique bien défini, sauf si l'on conditionne les résultats des expériences effectuées sur une partie à des mesures effectuées sur l'autre. En l'absence de communication entre les parties, chaque sous-système apparaît dans un mélange statistique d'états quantiques dont la forme reflète l'état intriqué du système composé complet. L'intrication entraîne donc que, dans un monde quantique, le tout peut être plus ordonné que les parties.

De manière encore plus surprenante, nous verrons que plus qu'une exception, l'intrication est en fait la règle : nous montrerons que toute source imparfaite d'état quantiques, c'est-à-dire qui ne produit pas toujours le même état peut être vue comme provenant d'un état intriqué entre le système et un environnement auxiliaire au moyen de mesures projectives effectuées dans ce dernier. Ce résultat important dû à Gisin, Hughston, Jozsa et Wootters montre que, dans le cas quantique, on peut gagner de l'information en étendant le système considéré, alors que dans un cadre classique ce n'est jamais le cas. Inspiré par ce résultat, nous verrons comment généraliser la notion de mesure présentée au chapitre 1 en utilisant l'intrication et nous verrons que cela définit le processus de mesure le plus général dans le cadre quantique.

5.1 Systèmes composés

5.1.1 Qu'est-ce qu'un système composé ?

La notion de *système composé* suppose que l'on peut bien séparer le système complet en deux sous-systèmes sur lesquels on peut, en principe, faire des mesures sans que les appareils puissent mutuellement s'influencer¹. On demandera ainsi que l'on puisse construire un dispositif expérimental réalisant des mesures dans deux régions de l'espace suffisamment séparées pour qu'aucun signal classique ne puisse les connecter. Ces mesures seront effectuées sur des systèmes qui véhiculent des degrés de liberté quantiques.

Une condition préliminaire est de pouvoir amener ces systèmes dans deux régions bien séparées de l'espace. Parfois, c'est le dispositif source qui permet de le faire, comme dans les sources de photons jumeaux émis par un atome immobile dans deux directions opposées. Cependant, il faut parfois intervenir pour les séparer. Par exemple, en optique, une lame semi-réfléchissante sépare en deux voies distinctes un faisceau laser. Dans les systèmes d'atomes froids, des impulsions lumineuses adaptées agissent sur l'état interne des atomes et permettent donc de les séparer selon leur état atomique interne. Enfin, certains circuits électroniques quantiques comportent des composants conçus pour aiguiller les électrons vers divers contacts de sortie.

Ensuite, on doit pouvoir, au moins en principe, assurer une séparation causale des mesures : cela signifie que les systèmes doivent être suffisamment séparés pour qu'aucune communication ne puisse avoir lieu entre les mesures. Puisqu'aucun signal ne peut aller plus vite que la lumière, cela est en principe possible².

Dans une représentation spatio-temporelle, comme celle de la figure 5.1, l'ensemble des événements influençables par un événement donné correspond à l'intérieur d'un cône appelé cône de lumière. Cette même figure illustre une réalisation de la situation décrite plus haut : aucun signal provenant de la phase de mesure de l'un des appareils n'atteint l'autre appareil pendant sa phase de mesure. En général, il faut donc que la zone d'influence d'une mesure effectuée par l'un des appareils soit en dehors de celle de tout événement associé à l'autre mesure. Il est clair que, plus le processus de mesure demande du temps, plus on doit éloigner les appareils de mesure.

Ainsi, lorsque Alain Aspect effectua ses premières expériences sur les paires de photons intriquées à l'université d'Orsay, il veilla à ce que les mesures effectuées sur un photon ne puissent pas influencer celles effectuées sur l'autre [18, 20]. La relative lenteur du système de détection des photons due

1. Cela ne signifie pas que cela doive absolument être le cas mais plutôt qu'en principe, il est possible de le faire.

2. Dans le langage relativiste, cela veut dire que les mesures doivent être effectuées dans un intervalle de genre espace. Pour les lecteurs qui ne seraient pas familiers avec la physique relativiste, nous y reviendrons plus en détail au chapitre 13, volume II.

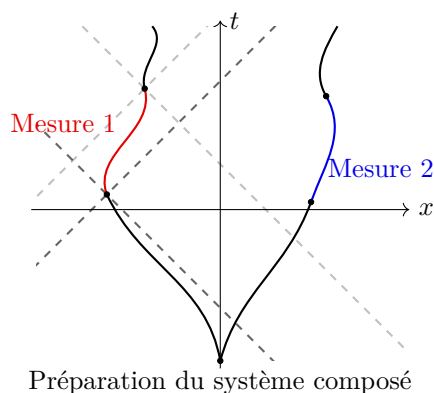


FIGURE 5.1 — Diagramme d'espace-temps pour un système composé, séparable en deux parties. Pour être sûr que les mesures sur les sous-parties ne puissent s'influencer, les sous-systèmes doivent être suffisamment séparés. La zone d'influence correspondant à l'une des mesures ne doit pas recouvrir celle de l'autre. Ici, on a représenté ces zones issues des événements correspondant au début et à la fin de la mesure 1. La mesure 2 est suffisamment éloignée pour ne pas être influencée par la première.

au choix aléatoire d'une polarisation l'obligea à séparer les deux dispositifs de mesure de 13 mètres³.

Nous venons donc de donner des conditions de principe pour pouvoir parler d'un système composé. Notons que toute cette approche repose sur le fait que les processus d'interaction sous-jacents aux mesures sont locaux. Elle repose aussi sur le fait que, dans un très grand nombre de situations, il est possible d'ignorer la complexité interne des objets physiques que l'on manipule. Il est donc nécessaire de pouvoir isoler clairement les degrés de liberté auxquels on s'intéresse. De la sorte, nous n'avons ni à descendre à l'échelle des objets les plus élémentaires de l'Univers, ni à considérer l'Univers et son état quantique dans leur globalité.

5.1.2 Les états dans un système composé

On va maintenant considérer le cas d'un système composé de deux parties A et B et discuter des possibilités qu'offre le principe de superposition. Par tradition, les expériences seront menées par deux observateurs : Alice qui effectue des mesures sur le sous-système A et Bob qui effectue des mesures sur le sous-système B .

3. Ce qui a nécessité la mise au point d'un système de mesures opérant en moins de 43 nanosecondes, ce qui était en soi un exploit.

Les états du système complet

En physique classique, les choses sont simples : les configurations d'un système composé sont de la forme (a, b) où a et b sont des configurations de chacun des sous-systèmes. Mathématiquement, l'espace des configurations du système composé est le produit cartésien des espaces de configurations de chacun des sous-systèmes composés. Ainsi, si on considère N bits classiques, l'espace des configurations classiques est un registre de N chiffres binaires. Il comprend donc 2^N configurations.

En physique quantique, chacune des parties est décrite par un espace de Hilbert, noté respectivement $\mathcal{H}_A$ et $\mathcal{H}_B$, et dont des bases sont données respectivement par $(|\phi_i^A\rangle)_{i \in I_A}$ et $(|\phi_j^B\rangle)_{j \in I_B}$. Les états d'une base correspondent à des configurations parfaitement discernables (chapitre 1) et leur indice peut être relié à des configurations « classiques » comme la position ou à des nombres quantiques comme le moment angulaire le long d'un axe de quantification. Pour décrire les états du système composé, les « configurations » pertinentes sont de la forme $(i, j) \in I_A \times I_B$. Or, en physique quantique, on associe à chaque configuration une amplitude. La donnée de ces amplitudes définit un vecteur d'état défini sur le *produit tensoriel* des espaces des états des deux parties A et B , $\mathcal{H}_A \otimes \mathcal{H}_B$ dont une base est

$$(|\phi_i^A\rangle \otimes |\phi_j^B\rangle)_{(i,j) \in I_A \times I_B}. \quad (5.1)$$

Ainsi, un état quelconque $|\Psi\rangle$ peut être décomposé sur cette base de $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ comme

$$|\Psi\rangle = \sum_{(i,j) \in I_A \times I_B} a_{ij} |\phi_i^A\rangle \otimes |\phi_j^B\rangle, \quad (5.2)$$

où les amplitudes a_{ij} sont des nombres complexes. Lorsqu'une unique amplitude est non nulle, on dit que l'état global $|\Psi\rangle$ est *factorisé*. Dans ce cas, et seulement dans ce cas, il est possible d'associer à chaque sous-système un état quantique qui lui est propre. Dans le cas général où plusieurs amplitudes sont non nulles, l'état est dit *intriqué*. Les phénomènes associés à cette structure particulière de l'état du système correspondent aux phénomènes *d'intrication*. La différence essentielle avec le cas factorisé est que, de manière générique, même si le système composé est préparé dans un état pur $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, il n'est pas possible d'affecter un état quantique bien défini aux sous-systèmes.

Les états relatifs

Cette étrangeté de l'intrication amène à se demander ce que l'on perçoit en regardant une seule partie du système composé.

Reprenons l'état défini par l'équation (5.2). Supposons maintenant que Bob fasse des mesures idéales dont les résultats possibles correspondent aux vecteurs de la base orthonormée $|\phi_j^B\rangle$ de $\mathcal{H}_B$. Si on considère le système complet, il

s'agit d'une mesure projective dont le sous-espace correspondant à ce résultat est dégénéré. En effet, les vecteurs d'états pour lesquels on est dans l'état $|\phi_j^B\rangle$ côté Bob forment le sous-espace $\mathcal{H}_A \otimes (\mathbb{C} |\phi_j^B\rangle)$ de $\mathcal{H}_A \otimes \mathcal{H}_B$. Après une telle mesure, l'état est donc, avant normalisation, la projection orthogonale de $|\Psi\rangle$ sur le sous-espace $\mathcal{H}_A \otimes (\mathbb{C} |\phi_j^B\rangle)$ de $\mathcal{H}_A \otimes \mathcal{H}_B$. La règle de Born nous donne alors la probabilité $p_B(j)$ de trouver le résultat j :

$$p_B(j) = \sum_{i \in I_A} |a_{ij}|^2. \quad (5.3)$$

L'état post-mesure est alors, après normalisation, $|\psi(A|j)\rangle \otimes |\phi_j^B\rangle$ où

$$|\psi(A|j)\rangle = \frac{1}{\sqrt{p_B(j)}} \sum_{i \in I_A} a_{ij} |\phi_i^A\rangle. \quad (5.4)$$

Un calcul direct montre que ce vecteur est indépendant de la base $(|\phi_i^A\rangle)_i$ de $\mathcal{H}_A$. L'état $|\psi(A|j)\rangle$, qui ne dépend donc que de $|\Psi\rangle$ et du résultat de la mesure effectuée par Bob, est appelé *état relatif* de A par rapport au résultat de mesure j effectué par B . Pour comprendre ce que représente cet état quantique, cherchons la probabilité conditionnelle qu'Alice identifie un état donné $|\varphi\rangle \in \mathcal{H}_A$ sachant que Bob a fait une mesure et obtenu le résultat j . Cette probabilité conditionnelle $p_A(|\varphi\rangle | j)$ est le quotient de la probabilité $|(\langle\varphi| \otimes \langle\phi_j^B|) |\Psi\rangle|^2$ qu'Alice et Bob aient détecté l'état $|\varphi\rangle \otimes |\chi_j^B\rangle$ dans $|\Psi\rangle$ par $p_B(j)$ donnée par (5.3). On peut alors exprimer le résultat final comme une règle de Born appliquée par Alice à son état relatif $|\psi(A|j)\rangle$:

$$p_A(|\varphi\rangle | j) = |\langle\varphi|\psi(A|j)\rangle|^2. \quad (5.5)$$

Ceci montre que $|\psi(A|j)\rangle$ est l'analogue quantique d'une probabilité conditionnelle (chapitre 2) au sens suivant : là où classiquement une probabilité conditionnelle représente l'ignorance qui subsiste lorsque l'on a obtenu le résultat de mesure j côté Bob, l'état quantique relatif $|\psi(A|j)\rangle$ représente l'incertitude quantique qui subsiste lorsque l'on a obtenu le résultat de mesure j côté Bob.

Enfin, l'état intriqué complet peut être réécrit sous la forme suivante

$$|\Psi\rangle = \sum_{j \in I_B} \sqrt{p_B(j)} |\psi(A|j)\rangle \otimes |\phi_j^B\rangle. \quad (5.6)$$

Cette équation fondamentale exprime l'état intriqué complet d'une manière qui est adaptée à la mesure côté Bob. Elle fait apparaître les états relatifs pour Alice qui, soulignons-le, ne sont pas forcément deux à deux orthogonaux. En revanche les états côté Bob le sont parce qu'ils sont associés à des résultats d'une mesure quantique projective côté Bob. Cette situation est représentée sur la figure 5.2. Bien entendu, on peut aussi écrire une décomposition adaptée

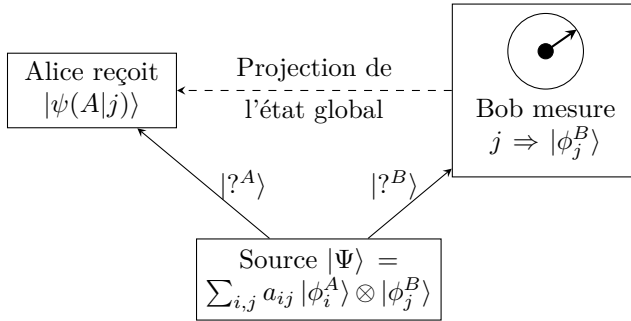


FIGURE 5.2 – Une source émet un état intriqué de deux systèmes dont l'un est envoyé à Alice et l'autre à Bob. Si on considère les expériences dans lesquelles Bob effectue une mesure et qu'il trouve le résultat correspondant à l'état $|\phi_j^B\rangle$, alors le système d'Alice est dans l'état relatif $|\psi(A|j)\rangle$. Cet état conditionné au résultat de la mesure par Bob est un état pur.

à une mesure côté Alice :

$$|\Psi\rangle = \sum_{i \in I_A} \sqrt{p_A(i)} |\phi_i^A\rangle \otimes |\psi(B|i)\rangle, \quad (5.7)$$

où, cette fois, $p_A(i)$ est la probabilité qu'Alice obtienne le résultat de mesure correspondant au vecteur $|\phi_i^A\rangle$ et $|\psi(B|i)\rangle \in \mathcal{H}_B$ désigne l'état de Bob relatif à ce résultat de mesure obtenu par Alice.

Mélange statistique et matrice densité

Regardons maintenant ce qui se passe lorsque Alice ne dispose pas de l'information sur le résultat obtenu par Bob. Dans cette situation, comme les résultats des mesures effectuées par Bob sont parfaitement discernables, on peut dire qu'elle reçoit chaque état $|\psi(A|j)\rangle$ avec la probabilité $p_B(j)$. Comment Alice va-t-elle décrire les résultats des expériences qu'elle peut mener ? Commençons par regarder quels sont les résultats qu'Alice va obtenir si elle cherche à mesurer une quantité A , dont les valeurs possibles a_i correspondent à des vecteurs $|a_i\rangle \in \mathcal{H}_A$. La probabilité d'observer le résultat a_i est

$$p_A(a_i) = \sum_{j \in I_B} p_B(j) |\langle a_i | \psi(A|j) \rangle|^2, \quad (5.8)$$

c'est-à-dire la somme sur les résultats de mesure de Bob des probabilités de mesurer a_i dans chaque état relatif $|\psi(A|j)\rangle$ pondérés par la probabilité que Bob ait mesuré le résultat j .

En général, on ne peut pas décrire cette probabilité par une formule de type règle de Born $|\langle a_i | \psi_A \rangle|^2$ avec un état pur $|\psi_A\rangle \in \mathcal{H}_A$. Un candidat

naturel serait de la forme $\sum_j \sqrt{p_B(j)} |\psi(A|j)\rangle$ mais on voit bien que les termes d'interférences qui apparaîtraient entre différents résultats de mesure de Bob ne sont pas présents dans la formule (5.8). Il est donc nécessaire d'introduire un nouveau moyen de décrire les états quantiques qui contienne à la fois les états purs que nous avons vus jusqu'ici, mais également cette notion, que l'on appelle un *mélange statistique d'états purs*⁴. On introduit alors l'opérateur densité

$$\rho_A = \sum_{j \in I_B} p_B(j) |\psi(A|j)\rangle \langle \psi(A|j)|. \quad (5.9)$$

Comment peut-on réécrire la probabilité d'obtenir le résultat de mesure a_i avec cette quantité ? On a vu au chapitre 1 que si $(|\alpha\rangle)$ est une base de $\mathcal{H}_A$ alors $\langle \phi | \psi \rangle = \sum_{\alpha} \langle \phi | \alpha \rangle \langle \alpha | \psi \rangle$. On peut donc réécrire

$$\begin{aligned} p_A(a_i) &= \sum_{j \in I_B} p_B(j) \langle a_i | \psi(A|j) \rangle \sum_{\alpha} \langle \psi(A|j) | \alpha \rangle \langle \alpha | a_i \rangle \\ &= \sum_{\alpha} \langle \alpha | a_i \rangle \left\langle a_i \left| \sum_{j \in I_B} p_B(j) |\psi(A|j)\rangle \langle \psi(A|j)| \right| \alpha \right\rangle \\ &= \text{tr}_{\mathcal{H}_A} (|a_i\rangle \langle a_i| \rho_A). \end{aligned} \quad (5.10)$$

L'opérateur densité ρ_A est donc l'objet qui nous permet de calculer toutes les prédictions pour les expériences qu'Alice voudrait réaliser.

De manière complètement générale, un opérateur densité est fait pour décrire un mélange statistique d'états quantiques : à un mélange d'état $|\psi_j\rangle$ avec des probabilités p_j , on associe l'opérateur densité $\rho = \sum_j p_j |\psi_j\rangle \langle \psi_j|$. La probabilité d'identifier un état quantique $|a\rangle$ au sein de ce mélange est alors $P(a|\rho) = \langle a | \rho | a \rangle = \text{tr}(|a\rangle \langle a| \rho)$.

De façon plus formelle encore, un *opérateur densité* est défini comme un opérateur linéaire ρ vérifiant les propriétés suivantes :

- ρ est hermitien : $\rho = \rho^\dagger$;
- ρ est normalisé : $\text{tr} \rho = 1$;
- ρ est positif : pour tout état $|\phi\rangle$, $\text{tr}(\rho |\phi\rangle \langle \phi|) \geq 0$.

Tout opérateur densité défini de la sorte peut être diagonalisé dans une base orthonormée avec des valeurs propres réelles, positives et sommant à l'unité. Il correspond donc au mélange statistique de ses vecteurs propres, avec des probabilités données par les valeurs propres correspondantes, et nous retrouvons bien la première caractérisation que nous avons présentée. Les éléments diagonaux de cet opérateur, appelés *populations*, représentent les probabilités d'obtenir les états de la base considérée tandis que les termes hors diagonaux, appelés *cohérences*, caractérisent les effets d'interférences quantiques.

La valeur moyenne de n'importe quelle observable O se calcule directement à partir de l'opérateur densité comme $\langle O \rangle = \text{tr}(O\rho)$. Enfin, l'équivalent de

4. Ou tout simplement un état mélange.

la quantité $|\langle\phi|\psi\rangle|^2$ pour des mélanges décrits par des opérateurs densités ρ_1 et ρ_2 est alors le recouvrement $\text{tr}_{\mathcal{H}}(\rho_1\rho_2)$, comme on vient de le voir pour le cas particulier de la probabilité de se trouver dans l'état pur $|a_i\rangle$, décrit par l'opérateur densité $|a_i\rangle\langle a_i|$.

Dans le cas qui nous intéresse, y a-t-il un moyen de donner l'opérateur densité représentant l'état mélange reçu par Alice ? En utilisant l'expression (5.6) de l'état du système complet $|\psi\rangle$ et l'orthogonalité des états $|\phi_j^B\rangle$, nous voyons que l'opérateur densité ρ_A défini par (5.9) se réécrit sous la forme compacte

$$\rho_A = \text{tr}_{\mathcal{H}_B}(|\Psi\rangle\langle\Psi|). \quad (5.11)$$

De manière tout à fait remarquable, cette expression étant une trace (qualifiée de partielle puisque l'on trace juste sur un sous-espace), elle ne fait plus du tout intervenir les mesures effectuées par Bob. Celui-ci aurait fait des mesures d'une autre quantité physique correspondant à une autre base orthonormée de $\mathcal{H}_B$, nous aurions obtenu le même ρ_A . Cela montre que ce que reçoit Alice est indépendant de ce qui se déroule chez Bob. Ce nouvel opérateur, appelé *opérateur densité réduit*, caractérise complètement la statistique des résultats de mesures locales faites par Alice.

Cependant, nous avons décrit les résultats des expériences que pourrait faire Alice en termes d'un mélange d'états relatifs qui eux dépendent bien des mesures effectuées par Bob. N'y a-t-il pas là un paradoxe ? En fait, il n'en est rien. Pour qu'Alice puisse décrire les expériences qu'elle fait dans son laboratoire en termes des états relatifs aux mesures effectuées par Bob, elle a besoin des résultats expérimentaux de Bob. C'est cela qui va lui permettre de classer ses propres résultats d'expériences en fonction des résultats des mesures effectuées par Bob. En prenant l'ensemble de ses résultats correspondant au résultat j de Bob, et en effectuant une tomographie (voir chapitre 4), Alice va reconstruire l'état relatif $|\psi(A|j)\rangle$. Mais en l'absence d'information fournie par Bob, Alice serait incapable de reconstruire autre chose que la matrice densité ρ_A . Elle ne pourrait pas reconstruire son écriture sous la forme (5.9). Si ce qui se passe côté Bob reste caché à Alice, celle-ci voit arriver dans son laboratoire des systèmes quantiques préparés de manière imparfaite, c'est-à-dire pas tous dans le même état pur. Elle doit donc utiliser l'opérateur densité ρ_A comme cela est résumé sur la figure 5.3.

Un point intéressant est que ρ_A possède de nombreuses décompositions sous forme d'un mélange statistique d'états qui ne sont pas deux à deux orthogonaux : nous avons vu que différents choix de quantités mesurées par Bob conduisaient à différentes décompositions (5.9). Nous reviendrons sur ce point dans la section 5.3 de ce chapitre.

L'essentiel à retenir de cette discussion est que, dans le cas d'un système composé qui est préparé dans un état intriqué, il n'est pas possible d'associer un état quantique pur à chacune des deux parties du système. Pour calculer des probabilités qui concernent Alice ou Bob séparément (probabilités marginales), on doit utiliser un mélange statistique. En revanche, il est possible de définir

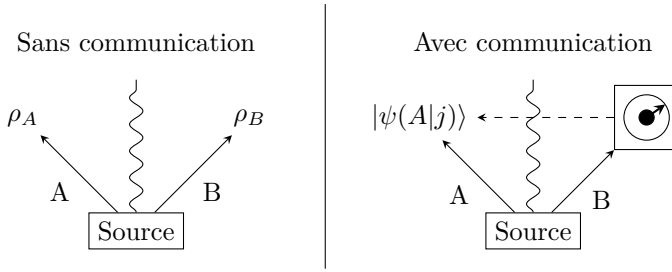


FIGURE 5.3 — Une source émet un état intriqué de deux systèmes dont l'un est envoyé à Alice et l'autre à Bob. Lorsque Bob ne communique aucune information à Alice, celle-ci voit arriver un mélange statistique d'états décrit par l'opérateur densité ρ_A . Elle ne peut prédire quel état elle va recevoir car elle n'a aucun moyen de savoir ce que Bob fait dans son laboratoire, séparé du sien par un « rideau » représenté ici par une ligne ondulée. Néanmoins, pour chaque résultat de mesure possible côté Bob, on sait qu'Alice reçoit l'état pur relatif à ce résultat $|\psi(A|j)\rangle$. Pour qu'elle puisse tomographier un tel état, il faut qu'elle sache ce que Bob a mesuré. Il faut donc que Bob lui communique cette information par un canal de communication classique représenté ici par une ligne pointillée.

un état pur pour chacune des parties en conditionnant l'état au résultat d'une mesure effectuée sur l'autre partie. Ces états relatifs permettent de calculer les probabilités conditionnelles pour Alice ou Bob *via* une formule type règle de Born. Sous une forme anodine, il s'agit d'une différence fondamentale entre le classique et le quantique sur laquelle nous reviendrons plus loin dans cet ouvrage.

Cas des états factorisés

Avant d'aller plus loin dans l'étude des états intriqués, revenons rapidement sur le cas où la source émet des états factorisés, c'est-à-dire qui sont de la forme

$$|\Psi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle. \quad (5.12)$$

Dans ce cas, les opérateurs densités réduits vus par Alice et Bob s'écrivent sous la forme très simple suivante :

$$\rho_A = \text{tr}_{\mathcal{H}_B} |\Psi\rangle\langle\Psi| = |\phi_A\rangle\langle\phi_A|, \quad (5.13a)$$

$$\rho_B = \text{tr}_{\mathcal{H}_A} |\Psi\rangle\langle\Psi| = |\phi_B\rangle\langle\phi_B|. \quad (5.13b)$$

Alice et Bob voient donc chacun un état pur comme s'ils avaient accès chacun à une source produisant exactement cet état. Dans ce cas, on note qu'il n'y a pas de corrélation entre Alice et Bob : si $\hat{A}$ et $\hat{B}$ sont deux observables, chacune

relative à Alice et Bob respectivement, alors la probabilité pour qu'Alice mesure a_i et Bob b_j est

$$p(a_i, b_j) = |\langle a_i | \phi_A \rangle|^2 \times |\langle b_j | \phi_B \rangle|^2 = p_A(a_i) p_B(b_j). \quad (5.14)$$

Il y a donc indépendance statistique entre ce qui se passe du côté d'Alice et ce qui se passe du côté de Bob.

En revanche, dans le cas où l'état est intriqué, il y a des corrélations entre Alice et Bob que nous allons discuter dans la prochaine section. Nous verrons dans le chapitre 11 comment les étudier expérimentalement et en quoi elles sont fondamentalement différentes des corrélations que l'on peut trouver dans le monde classique.

Remarquons enfin que la propriété de factorisation est préservée par n'importe quelle opération unitaire ou de mesure quantique effectuée sur A ou B seul. Par conséquent, pour détruire la factorisation et donc générer de l'intrication, il faut effectuer une opération qui couple les systèmes A et B . C'est ce qui arrive dès qu'il y a des interactions entre les deux sous-systèmes. Comme on le verra dans les chapitres consacrés à la dynamique des systèmes quantiques ouverts, les interactions vont souvent entraîner une perte du caractère quantique de l'état du système (atténuation des termes d'interférences), ce qui pose problème dans la perspective de la manipulation contrôlée d'états quantiques.

5.2 Les états intriqués et leurs symétries

Dans la section précédente, nous avons montré que tout état intriqué peut être écrit d'une manière adaptée à une mesure effectuée côté Bob ou côté Alice. Cette notion permet de faire apparaître les états relatifs aux résultats de mesure. Mais les états relatifs à des résultats de mesure distincts ne sont en général pas orthogonaux.

Cependant, nous allons voir que, parmi toutes les décompositions en états relatifs, il en existe une pour laquelle les états relatifs sont deux à deux orthogonaux et donc parfaitement discernables. Cette décomposition est appelée la *décomposition de Schmidt* de l'état intriqué. Nous allons voir que, dans un certain nombre de cas, cette écriture n'est pas unique. Il existe plusieurs choix de bases côté Alice et côté Bob qui fournissent une décomposition de Schmidt. Comme souvent en physique, une telle ambiguïté est la signature d'une symétrie. Dans le cas présent, c'est la conséquence d'une symétrie de l'état intriqué introduite par Zurek [392]. Cette symétrie est au cœur de l'argumentation de Zurek visant à justifier la règle de Born et l'émergence des probabilités dans le formalisme quantique [393], ce dont nous reparlerons dans le chapitre 16, volume II.

5.2.1 La décomposition de Schmidt des états intriqués

Considérons un état intriqué que nous écrivons en utilisant des états relatifs à une mesure effectuée par Bob :

$$|\Psi\rangle = \sum_j \sqrt{p_B(j)} |\psi(A|j)\rangle \otimes |\phi_j^B\rangle, \quad (5.15)$$

où les $|\phi_j^B\rangle$ forment une base orthonormée. Les états relatifs quant à eux sont bien normés mais pas forcément deux à deux orthogonaux.

D'un point de vue informationnel, ces différentes décompositions qui diffèrent par le choix des mesures effectuées par Bob (qui détermine la base $|\phi_j^B\rangle$) ne se valent pas. Le nombre de bits dont Bob a besoin pour encoder ses résultats de mesure en vue de les transmettre à Alice est donné par :

$$S[p_B] = - \sum_{j \in I_B} p_B(j) \log p_B(j). \quad (5.16)$$

Ayant reçu ces informations classiques en provenance de Bob, Alice pourra alors trier ses résultats expérimentaux en sous-ensembles de données qui correspondront chacun à un état relatif bien défini. Si un grand nombre de résultats de mesure $j \in I_B$ correspond à des états relatifs pour Alice qui sont peu discernables, on se doute que ce n'est pas optimal : Bob devra utiliser beaucoup de bits pour encoder des résultats de mesure différents mais qui correspondent pour Alice à des états relatifs très peu différents. Si, en revanche, Bob choisit de mesurer une quantité pour laquelle les résultats de mesure correspondent à des états relatifs orthogonaux côté Alice, notre intuition nous dit que cela devrait minimiser l'information que Bob a besoin de transmettre à Alice.

Nous verrons plus loin, par un calcul de théorie de l'information, que cette intuition s'avère exacte. Mais pour le moment, elle attire notre attention sur le fait qu'un choix de mesure de Bob qui conduit à des états relatifs orthogonaux pour Alice est potentiellement spécial. Regardons en premier lieu s'il en existe un. Pour cela, choisissons côté Bob une base orthonormée $|\phi_j^B\rangle$ qui diagonalise l'opérateur densité réduit ρ_B . L'hermiticité de l'opérateur densité assure qu'il est toujours possible de trouver une telle base. Comme dans le paragraphe précédent, notons $|\psi(A|j)\rangle$ les états relatifs d'Alice correspondant à ce choix de mesure de Bob. L'opérateur densité ρ_B est alors obtenu en prenant la trace partielle sur l'espace des états pour Alice :

$$\begin{aligned} \rho_B &= \text{tr}_{\mathcal{H}_A} (|\Psi\rangle\langle\Psi|) \\ &= \sum_{k,l} \sqrt{p_B(k)p_B(l)} \text{tr}_{\mathcal{H}_A} (|\psi(A|k)\rangle\langle\psi(A|l)|) |\phi_k^B\rangle\langle\phi_l^B| \\ &= \sum_{k,l} \sqrt{p_B(k)p_B(l)} \langle\psi(A|l)|\psi(A|k)\rangle |\phi_k^B\rangle\langle\phi_l^B|. \end{aligned} \quad (5.17)$$

Comme on a pris une base dans laquelle l'opérateur densité vu par Bob est diagonal, on en déduit que $\langle \psi(A|l) | \psi(A|k) \rangle = 0$ pour $k \neq l$. Pour ce choix précis de base orthonormée côté Bob, la famille des états relatifs $|\psi(A|j)\rangle$ est donc également orthogonale. Finalement nous obtenons une forme particulièrement simple et importante de l'état intriqué :

$$|\Psi\rangle = \sum_j \sqrt{p(j)} |\psi(A|j)\rangle \otimes |\psi(B|j)\rangle \quad (5.18)$$

où les états utilisés du côté d'Alice et du côté de Bob forment des familles orthonormées :

$$\forall(i, j), \quad \langle \psi(A|i) | \psi(A|j) \rangle = \langle \chi(B|i) | \chi(B|j) \rangle = \delta_{i,j}. \quad (5.19)$$

Notons que grâce à cela, ils peuvent donc être vus comme des états relatifs réciproques les uns aux autres.

Cette décomposition particulière est ce qu'on appelle la décomposition de Schmidt de l'état intriqué. Dans une telle base, les deux parties A et B ont des matrices densités réduites qui ont le même spectre de valeurs propres non nulles. Ces valeurs propres ne sont autres que les probabilités qui caractérisent les mélanges d'état pour Alice et Bob :

$$\rho_A = \sum_j p_j |\psi(A|j)\rangle \langle \psi(A|j)|, \quad (5.20a)$$

$$\rho_B = \sum_j p_j |\psi(B|j)\rangle \langle \psi(B|j)|. \quad (5.20b)$$

Les états intriqués correspondent au cas où le spectre en question présente plus d'une valeur propre non nulle (s'il n'y en a qu'une, on a un état factorisé).

Remarquons que d'un point de vue informationnel, la décomposition de Schmidt est optimale : les mesures effectuées par Bob permettent d'associer deux états discernables à coup sûr côté Alice. La décomposition de Schmidt est particulièrement intéressante parce qu'elle met en évidence une corrélation parfaite entre des mesures effectuées par Alice et par Bob. En effet, si Alice mesure une observable dont la base propre associée contient la famille orthonormale des $|\psi(A|j)\rangle$ de la décomposition de Schmidt, alors on aura une corrélation parfaite avec les mesures effectuées par Bob sur une observable dont la base propre associée contient les états $|\psi(B|j)\rangle$.

5.2.2 La décomposition de Schmidt est-elle unique ?

L'existence de la décomposition de Schmidt dans le cas quantique nous dit qu'il existe une observable pour chaque partie qui donne des corrélations parfaites des résultats de mesure. Notons que l'existence de corrélations parfaites entre deux systèmes ayant interagi par le passé n'est pas spécifique au quantique. En classique, c'est ce qui se passe pour certaines expériences bien

choisies. Par exemple, si on prépare deux boules de billard de masses identiques et qu'on les lance en collision frontale à la même vitesse, leurs directions de sortie seront parfaitement opposées. Cependant, la physique quantique permet parfois l'apparition de corrélations parfaites sur beaucoup plus de paires d'observables que dans le cas classique. Nous allons donc chercher s'il y a d'autres couples d'observables qui présentent des corrélations aussi fortes.

Examinons pour commencer le cas où les valeurs propres non nulles sont non dégénérées. On montre alors que la décomposition de Schmidt est unique à des phases près, c'est-à-dire que l'on peut tout aussi bien choisir les vecteurs de base suivants sans pour autant changer la description de l'état :

$$|\phi_j'^A\rangle = e^{i\theta_j} |\phi_j^A\rangle, \quad (5.21a)$$

$$|\phi_j'^B\rangle = e^{-i\theta_j} |\phi_j^B\rangle. \quad (5.21b)$$

Cette symétrie de l'état intriqué, introduite par Zurek [391], porte le nom d'*invariance*⁵. Dans le cas où les valeurs propres non nulles sont non dégénérées, la décomposition de Schmidt n'a pas vraiment d'ambiguïté. Les espaces propres sont unidimensionnels et l'invariance est donnée par les facteurs de phase que nous venons de discuter. Dans le cas où l'on a n valeurs propres non nulles, elle est donnée par un groupe $U(1)^n$ où $U(1)$ est le groupe des nombres complexes de module unité.

Dans le cas où des valeurs propres non nulles sont dégénérées, on a une ambiguïté supplémentaire sur la décomposition de Schmidt. En effet, prenons l'exemple de l'état intriqué à deux qubits suivant⁶ :

$$|B_{00}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \quad (5.22)$$

Alice et Bob voient chacun de leur côté un opérateur densité, représenté par une matrice dans la base $(|0\rangle, |1\rangle)$ de leur sous-espace, qui correspond à un qubit totalement dépolarisé :

$$\rho_A = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (5.23)$$

Mais nous pouvons tout aussi bien changer de base. Considérons les états $|X_{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ qui forment aussi une base orthonormée. L'état $|B_{00}\rangle$ se réécrit alors sous la forme :

$$|B_{00}\rangle = \frac{|X_+X_+\rangle + |X_-X_-\rangle}{\sqrt{2}}. \quad (5.24)$$

5. Ce nom est l'abréviation de *environment assisted invariance*, c'est-à-dire « symétrie assistée par l'environnement ». Il se justifie par le fait que l'évolution d'un système quantique ouvert conduit naturellement à l'émergence d'un état intriqué entre le système et l'ensemble des degrés de liberté avec lesquels il interagit et qu'on appelle l'environnement, ce dont nous reparlerons longuement dans le chapitre 9, ainsi que dans le chapitre 14, volume II. La symétrie introduite par Zurek n'est pas à proprement parler une symétrie de l'état du système seul mais une symétrie de l'état du système et de son environnement.

6. La raison de cette notation sera explicitée dans le chapitre 7.

De manière plus générale, considérons les deux états orthogonaux correspondant à deux vecteurs opposés sur la sphère de Bloch. Considérons $\mathbf{n}(\theta, \varphi)$ le vecteur de Bloch repéré par les coordonnées sphériques θ et φ et introduisons l'état associé $|\mathbf{n}(\theta, \varphi)\rangle$ défini par l'équation (4.12). Son état orthogonal s'obtient en remplaçant θ par $\pi + \theta$ et correspond à $|\mathbf{-n}(\theta, \varphi)\rangle$. On vérifie alors directement que

$$|B_{00}\rangle = \frac{1}{\sqrt{2}} (|\mathbf{n}(\theta, \varphi), \mathbf{n}(\theta, -\varphi)\rangle + |\mathbf{-n}(\theta, \varphi), \mathbf{-n}(\theta, -\varphi)\rangle) . \quad (5.25)$$

Cette ambiguïté dans la décomposition de Schmidt n'apparaît que parce que des valeurs propres des opérateurs densités réduits sont dégénérées. La symétrie dont nous avons parlé précédemment s'élargit alors. Une transformation générale de $SU(2)$ peut être décrite de manière générale par une matrice

$$\begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \quad (5.26)$$

où α et β sont deux nombres complexes dont le module somme à l'unité. On peut donc toujours les paramétrer au moyen de trois angles $\theta, \varphi_{1,2}$ de telle sorte que

$$\alpha = e^{i\varphi_1} \cos \frac{\theta}{2} \quad \text{et} \quad \beta = e^{i\varphi_2} \sin \frac{\theta}{2} . \quad (5.27)$$

Nous voyons alors que cette transformation spéciale unitaire générale $U(\theta, \varphi_1, \varphi_2)$ envoie la base canonique $(|0\rangle, |1\rangle)$ sur la base considérée plus haut avec une phase relative entre les deux vecteurs :

$$U(\theta, \varphi_1, \varphi_2)|0\rangle = e^{i(\varphi_1+\varphi_2)/2} |\mathbf{n}(\theta, \varphi_1 - \varphi_2)\rangle , \quad (5.28a)$$

$$U(\theta, \varphi_1, \varphi_2)|1\rangle = e^{-i(\varphi_1+\varphi_2)/2} |\mathbf{-n}(\theta, \varphi_1 - \varphi_2)\rangle . \quad (5.28b)$$

En conséquence, l'équation (5.24) signifie tout simplement que

$$(U(\theta, \varphi_1, \varphi_2) \otimes U(\theta, -\varphi_1, -\varphi_2)) |B_{00}\rangle = |B_{00}\rangle . \quad (5.29)$$

Par rapport au cas d'un état intriqué de deux qubits dont le spectre est non dégénéré, l'invariance de $|B_{00}\rangle$ est étendue à un groupe $U(2) = U(1) \times SU(2)$. Le groupe d'invariance est donc de dimension totale 4 (3 pour $SU(2)$ et 1 pour $U(1)$) alors que dans le cas générique, il est de dimension 2.

Notons que l'invariance se traduit par une symétrie de l'état réduit pour chacune des deux parties. En effet, nous avons vu que $\rho_A = \text{tr}_{\mathcal{H}_B} (|\Psi\rangle\langle\Psi|)$. En conséquence, si l'état du système composé $|\Psi\rangle$ est invariant sous $U_A \otimes U_B$ produit d'opérateurs unitaires agissant respectivement sur $\mathcal{H}_A$ et $\mathcal{H}_B$, alors

$$\begin{aligned} \rho_A &= \text{tr}_{\mathcal{H}_B} \left((U_A \otimes U_B) |\Psi\rangle\langle\Psi| (U_A^\dagger \otimes U_B^\dagger) \right) \\ &= \text{tr}_{\mathcal{H}_B} \left((U_A \otimes \mathbb{1}) |\Psi\rangle\langle\Psi| (U_A^{-1} \otimes \mathbb{1}) \right) \\ &= U_A \rho_A U_A^{-1} \end{aligned} \quad (5.30)$$

et donc l'opérateur densité ρ_A est laissé invariant par la transformation unitaire U_A . De même ρ_B est laissé invariant par U_B . Réciproquement, nous allons montrer que si $|\Psi\rangle$ est un état intriqué qui donne comme état réduit ρ_A pour le système A , alors

$$\mathcal{C}(\rho_A) = \{U \in \mathcal{U}(\mathcal{H}_A) / [U, \rho] = 0\} \quad (5.31)$$

est un groupe isomorphe au groupe d'envariance de $|\Psi\rangle$.

Pour le démontrer, considérons un état intriqué $|\Psi\rangle$ qui donne par trace partielle l'opérateur densité ρ_A pour Alice et supposons que nous ayons une transformation U_A unitaire telle que $[U_A, \rho_A] = 0$. En conséquence, cette transformation laisse invariant chaque sous-espace propre de ρ_A . Considérons un tel sous-espace propre, associé à la valeur propre $p \in [0, 1]$ et introduisons la décomposition de Schmidt de l'état intriqué $|\Psi\rangle$ sous la forme suivante :

$$|\Psi\rangle = \sum_{p \in \text{Sp}(\rho_A)} \sqrt{p} \sum_{j \in I_p} |\phi_j^A\rangle \otimes |\phi_j^B\rangle, \quad (5.32)$$

où les $|\phi_j^A\rangle$ (resp. les $|\phi_j^B\rangle$) pour $j \in I_p$ forment une base orthonormée du sous-espace propre de ρ_A (resp. ρ_B) associé à la valeur propre p . En fait, le problème consiste à chercher des transformations U_B dans chaque sous-espace propre qui vérifient :

$$(U_A \otimes U_B) \left(\sum_{j \in I_p} |\phi_j^A\rangle \otimes |\phi_j^B\rangle \right) = \sum_{j \in I_p} |\phi_j^A\rangle \otimes |\phi_j^B\rangle. \quad (5.33)$$

En termes de matrices, cela s'écrit :

$$\sum_{(i, i', j) \in I_p^3} U_{i,j}^A U_{i',j}^B |\phi_i^A\rangle \otimes |\phi_{i'}^B\rangle = \sum_{j \in I_p} |\phi_j^A\rangle \otimes |\phi_j^B\rangle. \quad (5.34)$$

Pour que cette condition soit satisfaite, il faut et il suffit d'imposer que :

$$\sum_j U_{i,j}^A U_{i',j}^B = \delta_{i,i'}. \quad (5.35)$$

Or, comme U_A est unitaire, il suffit de prendre comme matrice représentant U_B dans la base $|\phi_j^B\rangle$ la matrice $U_{i',j}^B = (U_{i,j}^A)^*$ de sorte que l'unitarité de U_A assure la validité de (5.35). Cette construction fournit un opérateur U_B qui agit sur l'espace des états côté Bob de telle sorte que $(U_A \otimes U_B)|\Psi\rangle = |\Psi\rangle$. Nous avons montré que chaque élément du groupe d'envariance de l'état intriqué $|\Psi\rangle$ est totalement déterminé par U_A ou, de manière équivalente, par U_B . En conséquence, le groupe d'envariance est isomorphe au commutant de ρ_A (ou de ρ_B). Soulignons une fois encore que cela montre qu'une envariance non triviale provient nécessairement de valeurs propres non nulles dégénérées pour ρ_A ou ρ_B .

5.2.3 États intriqués et corrélations de mesure

Comme nous l'avons vu, lorsque Alice et Bob partagent un état intriqué, il apparaît une corrélation parfaite entre les mesures d'une observable d'Alice et d'une observable de Bob dès lors que ces deux observables ont pour bases propres les familles orthogonales qui apparaissent dans la décomposition de Schmidt de l'état intriqué.

Cependant, une envariance non triviale signifie que la décomposition de Schmidt n'est pas unique. On s'attend donc à avoir plus de corrélations parfaites entre Alice et Bob que dans le cas générique. En effet, considérons qu'Alice et Bob partagent une paire de qubits dans l'état

$$|B_{00}\rangle = \frac{|Z_+Z_+\rangle + |Z_-Z_-\rangle}{\sqrt{2}}. \quad (5.36)$$

Cet état intriqué peut tout aussi bien s'écrire

$$|B_{00}\rangle = \frac{|X_+X_+\rangle + |X_-X_-\rangle}{\sqrt{2}}. \quad (5.37)$$

La première écriture est bien adaptée aux mesures de spin dans la direction z via l'observable σ_z . Pour une telle mesure, si Alice trouve la valeur $+1$ ou -1 , Bob trouvera exactement le même résultat. La seconde écriture quant à elle est bien adaptée pour des mesures dans la direction x et là aussi il y a corrélation parfaite entre les mesures des deux observateurs. Plus généralement, l'envariance étendue au sens discuté dans l'équation (5.29) pour l'état $|B_0\rangle$ assure une corrélation parfaite pour les mesures côté Alice et côté Bob dans n'importe quelle direction de la sphère de Bloch.

Pour souligner le rôle crucial joué par l'envariance étendue dans ce raisonnement, considérons le cas d'un état intriqué possédant une envariance générique (non étendue). Pour cela, introduisons :

$$|B_{00}^{(X)}\rangle = \frac{|X_+X_+\rangle + |X_-X_-\rangle}{\sqrt{2}}, \quad (5.38a)$$

$$|B_{01}^{(X)}\rangle = \frac{|X_-X_+\rangle + |X_+X_-\rangle}{\sqrt{2}} \quad (5.38b)$$

et pour α et β deux nombres réels positifs de carré somme unité, introduisons

$$|B_z(\alpha, \beta)\rangle = \alpha|Z_+Z_+\rangle + \beta|Z_-Z_-\rangle. \quad (5.39)$$

Lorsque $\alpha = \beta = 1/\sqrt{2}$, on retrouve l'état $|B_0\rangle$ considéré dans le paragraphe précédent. Pour α et β non nuls, l'état $|B_z(\alpha, \beta)\rangle$ est intriqué et il y a corrélation parfaite entre σ_z^A et σ_z^B . De même, pour les états $|B_{00}^{(X)}\rangle = |B_{00}\rangle$ et $|B_{01}^{(X)}\rangle$, nous avons respectivement corrélation et anticorrélation parfaite entre σ_x^A et σ_x^B . Cependant, l'état $|B_z(\alpha, \beta)\rangle$ se réécrit sous la forme :

$$|B_z(\alpha, \beta)\rangle = \frac{\alpha + \beta}{\sqrt{2}} |B_{00}^{(X)}\rangle + \frac{\alpha - \beta}{\sqrt{2}} |B_{01}^{(X)}\rangle. \quad (5.40)$$

Dès lors que $\alpha + \beta \neq 0$ ou $\alpha - \beta \neq 0$, on voit que l'on aura un mélange probabiliste de corrélations et d'anticorrélations parfaites entre Alice et Bob pour σ_x . Dans le cas d'une envariance générique, les corrélations parfaites entre Alice et Bob n'ont donc lieu que pour un seul choix de mesure.

5.3 Intrication et mélange

Jusqu'à maintenant, nous avons étudié ce qui se passe pour chaque sous-système lorsque l'on connaît l'état du système composé. La question naturelle à se poser désormais concerne le raisonnement inverse : est-il possible, à partir de la connaissance uniquement de ce qui se passe pour un des sous-systèmes, de déduire des informations sur un hypothétique état pur décrivant un système plus important ?

5.3.1 Sources imparfaites et purification

Dans cette section, nous allons voir que la réponse à cette question est la clé qui permet de comprendre et de modéliser des sources d'états quantiques réalistes, c'est-à-dire imparfaites. En effet, dans une expérience de laboratoire, on dispose rarement d'une source qui produit des objets quantiques préparés dans le même état. Même si certaines expériences atteignent un degré de contrôle très poussé, il y a toujours des « fluctuations » dans le processus de préparation qui font qu'une source réelle est imparfaite et produit un mélange statistique d'états au lieu d'un état pur.

Pour commencer, convoquons nos expérimentateurs habituels, Alice et Bob. On peut très facilement dire qu'une source qui produit un mélange pour Bob provient d'une source d'états purs intriqués pour Alice et Bob. En effet, imaginons que la source à laquelle Bob a accès émet un mélange statistique d'états décrit par la matrice densité

$$\rho_B = \sum_i p_i |\phi_i^B\rangle\langle\phi_i^B|, \quad (5.41)$$

où les états $|\phi_i^B\rangle$ ne sont pas forcément deux à deux orthogonaux. Posons alors

$$|\Psi\rangle = \sum_i \sqrt{p_i} |\phi_i^A\rangle |\phi_i^B\rangle, \quad (5.42)$$

où les états $|\phi_i^A\rangle$ sont des états orthogonaux et normés dans un espace des états $\mathcal{H}_A$ de dimension au moins égale à celle du nombre d'états apparaissant dans l'équation (5.41). Nous pouvons alors à partir de cet état retrouver l'opérateur densité ρ_B comme étant l'opérateur densité réduit produit à partir de l'état $|\Psi\rangle$ pour le sous-système B . Nous venons ainsi de montrer comment on peut fabriquer toute source dite imparfaite, c'est-à-dire qui produit un mélange d'états, à partir d'une source parfaite d'états intriqués et d'un processus de

mesure projective effectué par Alice. Un tel processus est appelé une *purification* du mélange. À ce stade, c'est juste une écriture formelle. Rappelons encore une fois que si Alice et Bob ne communiquent pas entre eux et n'effectuent que des mesures dans leur laboratoire respectif, ils n'ont aucun moyen de remonter à autre chose qu'à leur opérateur densité réduit respectif.

Néanmoins, essayons de voir si cette idée formelle peut être poussée plus loin. Sachant que la description d'un état mélange n'est pas unique dès lors que l'on s'autorise à utiliser un ensemble d'états non orthogonaux (par exemple plus de deux états pour un qubit), nous allons nous demander si ces multiples descriptions d'une même source imparfaite peuvent être obtenues à partir d'un même état intriqué.

5.3.2 Théorème de Gisin, Hughston, Jozsa et Wootters

Comment deux écritures différentes d'un état mélange sont-elles reliées ? Pour le comprendre, prenons deux décompositions de Schmidt qui donnent le même opérateur densité réduit :

$$|\Psi_I\rangle = \sum_i \sqrt{p_i} |\phi_i^A\rangle |\phi_i^B\rangle, \quad (5.43a)$$

$$|\Psi_{II}\rangle = \sum_i \sqrt{p_i} |\phi_i'^A\rangle |\phi_i'^B\rangle. \quad (5.43b)$$

Nous avons déjà utilisé le fait que l'ensemble des coefficients de chaque décomposition donne le spectre de ρ_B . Par transformation unitaire sur B , nous pouvons nous ramener au cas où $|\phi_i^B\rangle = |\phi_i'^B\rangle$. Il existe alors une transformation unitaire U_A qui envoie chaque $|\phi_i^A\rangle$ sur $|\phi_i'^A\rangle$. On a alors

$$|\Psi_{II}\rangle = (U_A \otimes \mathbb{I}) |\Psi_I\rangle. \quad (5.44)$$

Pour conclure, on part maintenant de deux états intriqués dans un même espace de Hilbert qui, par trace partielle sur l'un des sous-espaces, donnent le même opérateur densité. On peut alors écrire les décompositions de Schmidt de ces deux états intriqués et le raisonnement ci-dessus s'applique. On peut donc relier ces deux états intriqués par une transformation unitaire qui agit uniquement sur le sous-espace sur lequel on trace.

Nous venons de montrer le résultat suivant dû à Gisin, Hughston, Jozsa et Wootters [212] : les différentes manières d'écrire un opérateur densité comme un mélange statistique d'états orthogonaux peuvent toutes être obtenues à partir d'un état pur intriqué entre le système et un système auxiliaire suffisamment vaste et elles diffèrent par le choix d'une mesure projective effectuée sur le système auxiliaire.

Ce résultat donne une tout autre dimension aux purifications formelles introduites dans la section précédente : il montre que la purification est essentiellement unique. Ce n'est donc pas une simple représentation formelle

et cela permet de réinterpréter les résultats précédents en termes physiques. Ainsi, nous pouvons dire que toutes les descriptions d'un mélange statistique reçu par Bob proviennent d'un processus de préparation physique qui consiste à partir d'un état intriqué partagé avec Alice et à faire une mesure côté Alice. Changer la mesure change alors la description de l'état reçu par Bob comme mélange explicite d'états quantiques.

Illustrons ce dernier point par un exemple : on prépare un qubit dans un état mélange équiprobable de $N \geq 2$ états $|\varphi_j\rangle$ ($1 \leq j \leq N$) normés mais pas forcément orthogonaux. Ce mélange est obtenu, par exemple, en partant de l'état intriqué

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^N |A_j\rangle |\varphi_j\rangle, \quad (5.45)$$

où les états $|A_j\rangle$ forment une famille orthonormée d'un espace des états pour Alice. Si Alice effectue une mesure projective d'une observable dont les $|A_j\rangle$ sont des vecteurs propres, cela génère pour Bob le mélange équiprobable des états $|\varphi_j\rangle$. Mais Alice peut aussi changer d'observable en effectuant une transformation unitaire qui laisse le sous-espace engendré par les $|A_j\rangle$ invariant. Introduisons la matrice unitaire $N \times N$ qui relie les vecteurs propres $|A'_k\rangle$ de cette nouvelle observable dans ce sous-espace à ceux de l'observable précédente

$$|A_j\rangle = \sum_{k=1}^N U_{j,k} |A'_k\rangle. \quad (5.46)$$

L'état intriqué $|\Psi\rangle$ s'écrit alors :

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^N \sum_{k=1}^N |A'_k\rangle \otimes U_{j,k} |\varphi_j\rangle \quad (5.47a)$$

$$= \frac{1}{\sqrt{N}} \sum_{k=1}^N |A'_k\rangle \otimes \sum_{j=1}^N U_{j,k} |\varphi_j\rangle. \quad (5.47b)$$

En conséquence, ce choix d'une nouvelle observable engendre cette fois-ci un mélange équiprobable d'états $|\phi'_k\rangle$ obtenus à partir des $|\varphi_j\rangle$ par transformation unitaire

$$|\phi'_k\rangle = \sum_{j=1}^N U_{k,j} |\varphi_j\rangle. \quad (5.48)$$

Cette nouvelle mesure côté Alice va donc permettre de décrire le mélange côté Bob comme provenant d'une nouvelle famille d'états ayant le même nombre d'états N .

La question naturelle est alors celle de la forme optimale de l'état intriqué à choisir pour exprimer tout ce qu'il y a à connaître sur l'état côté Bob. En effet,

le nombre N de l'exemple précédent n'est pas forcément le nombre minimal d'états nécessaires pour représenter l'opérateur densité vu par Bob, puisqu'il peut même être supérieur à la dimension de l'espace des états de Bob. L'idée consiste à utiliser la décomposition de Schmidt de l'état intriqué global, car on sait qu'elle va fournir une famille orthogonale qui diagonalise l'opérateur densité de Bob. Sous cette forme, le nombre minimal de vecteurs nécessaires pour décrire cet opérateur densité apparaît naturellement. Pour fixer les idées, considérons un mélange équiprobable d'états qui ne sont pas deux à deux orthogonaux :

$$\rho_B = \frac{1}{N} \sum_{j=1}^N |\varphi_j\rangle\langle\varphi_j|. \quad (5.49)$$

Dans ce cas, on va pouvoir écrire les vecteurs $|\varphi_j\rangle$ comme combinaisons linéaires d'un plus petit nombre $P < N$ de vecteurs $|\phi_p\rangle$ qui forment la base orthonormée diagonalisant ρ_B

$$|\varphi_j\rangle = \sum_{p=1}^P M_{j,p} |\phi_p\rangle, \quad (5.50)$$

où la matrice $M_{j,p}$ est une matrice rectangulaire $N \times P$. L'état intriqué $|\Psi\rangle$ se réécrit alors

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^N \sum_{p=1}^P M_{j,p} |A_j\rangle \otimes |\phi_p\rangle \quad (5.51)$$

$$= \sum_{p=1}^P \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N M_{j,p} |A_j\rangle \right) \otimes |\phi_p\rangle. \quad (5.52)$$

Comme les $|\phi_p\rangle$ constituent une famille orthonormée dans laquelle l'opérateur densité de Bob est diagonal, les vecteurs (non normés)

$$|\varphi_A(p)\rangle = \frac{1}{\sqrt{N}} \sum_j M_{j,p} |\varphi_j\rangle \quad (5.53)$$

forment une famille orthogonale d'états pour Alice. Ce sont ces états qu'il faut prendre pour mesurer côté Alice si on veut retomber sur la décomposition de Schmidt de ρ_B . On voit donc que si on part d'une représentation de ρ_B comprenant un très grand nombre d'états, Alice peut effectivement trouver une observable avec un spectre beaucoup plus petit, qui donnera une nouvelle description de ρ_A , cette fois avec un nombre beaucoup moins grand de vecteurs impliqués côté Bob.

Le théorème GHJW exprime donc une correspondance entre le point de vue de Bob seul et le point de vue « état intriqué partagé puis mesure » pour

décrire une source imparfaite d'états quantiques. Ce résultat fait émerger la source imparfaite comme résultant d'un processus physique dont l'ambiguïté provient du fait que nous ne pouvons pas voir ce qui est fait côté Alice. L'état intriqué $|\Psi\rangle$ partagé entre Alice et Bob n'est pas accessible par un seul des deux observateurs alors qu'il contient une information sur le processus de préparation du système global : seule une mesure globale ou des mesures de corrélations permettraient de le révéler.

Cette vision des choses est très utile car, comme on le verra au fil de ce livre, on peut toujours en physique quantique décrire les choses de manière unitaire pour peu que l'on étende convenablement le système (et donc l'espace des états). Alors que dans le monde classique, étendre un système ne le rend pas plus ordonné, dans le monde quantique c'est l'inverse et on peut toujours se ramener à des états purs en élargissant l'espace des états.

5.3.3 Mélanges et théorie de l'information

Nous sommes maintenant en position de revenir sur la notion de mélange statistique d'états à la lumière de la théorie de l'information. En effet, nous avons vu qu'un mélange statistique d'états quantiques pour Bob est décrit par un opérateur densité ρ_B . Celui-ci peut certes s'écrire comme un mélange statistique d'états de différentes manières qui proviennent d'un choix différent de mesure projective effectuée par Alice sur un état intriqué $|\Psi\rangle$ partagé avec Bob et tel que $\rho_B = \text{tr}_{\mathcal{H}_A}(|\Psi\rangle\langle\Psi|)$. Chaque choix de mesure par Alice donne alors naissance à une expression de ρ_B comme mélange des états relatifs de Bob. À chacune de ces décompositions est associée une entropie de Shannon

$$S[|\Psi\rangle, A] = - \sum_j p_A(j) \ln p_A(j), \quad (5.54)$$

qui représente le nombre de bits nécessaires pour encoder les résultats des mesures effectuées par Alice (voir chapitre 2). Ce nombre ne dépend que de l'état intriqué $|\Psi\rangle$ et de l'observable A choisie par Alice. Parmi toutes ces observables se trouvent celles qui correspondent à la décomposition de Schmidt de $|\Psi\rangle$,

$$|\Psi\rangle = \sum_a \sqrt{p_a} |\phi_a^A\rangle \otimes |\phi_a^B\rangle, \quad (5.55)$$

que l'on étend de manière à disposer de bases orthonormées $|\phi_a^A\rangle$ de $\mathcal{H}_A$ et $|\phi_a^B\rangle$ de $\mathcal{H}_B$. Certains des coefficients p_a sont alors nuls. L'entropie $S[|\Psi\rangle, A_S]$ où A_S diagonalise ρ_A peut alors s'exprimer en termes de l'opérateur densité ρ_B sans aucune référence à ce qui se passe chez Alice :

$$- \sum_a p_a \ln p_a = - \text{tr}_{\mathcal{H}_B}(\rho_B \ln \rho_B), \quad (5.56)$$

où $\ln \rho_B$ est défini par son expression dans sa base propre. Cette quantité est appelée l'entropie de von Neumann de l'opérateur densité ρ_B et sera notée $S_{\text{vn}}[\rho_A]$.

Si Alice prend une autre observable qui correspond à une autre base orthonormée $|\xi_j^A\rangle$, alors il existe une matrice unitaire U reliant ces deux bases $|\phi_a^A\rangle = \sum U_{a,j} |\xi_j^A\rangle$ et elle nous permet de trouver les états relatifs côté Bob :

$$|\Psi\rangle = \sum_j \left(\sum_a \sqrt{p_a} U_{a,j} |\psi_a^B\rangle \right) \otimes |\xi_j^A\rangle. \quad (5.57)$$

Les probabilités des résultats de mesure d'Alice dans cette base sont alors exprimées en termes des probabilités correspondantes dans la base de Schmidt par

$$p_A(j) = \sum_a |U_{a,j}|^2 p_a. \quad (5.58)$$

Mais la matrice $U_{a,j}$ étant unitaire, la matrice formée par les probabilités associées $p(b \rightarrow j) = |U_{b,j}|^2$ possède la propriété d'être bistochastique, c'est-à-dire que

$$\sum_j p(a \rightarrow j) = \sum_a p(a \rightarrow j) = 1. \quad (5.59)$$

Or nous avons montré à la section 2.3.5 que, sous ces hypothèses, l'entropie de Shannon de la distribution de probabilité $j \mapsto p_A(j)$ est supérieure à celle de la distribution donnée par les p_a . En conséquence, nous venons de montrer que

$$S[\rho_B] \leq - \sum_j p(j) \ln p(j) \quad (5.60)$$

pour toute écriture de ρ_A sous la forme d'un mélange statistique d'états avec les probabilités $p(j)$.

Ceci montre une première interprétation de l'entropie de von Neumann : c'est la plus petite dispersion statistique d'un mélange d'états qui est décrit par ρ_B . Nous venons ainsi de justifier notre intuition de la section 5.2.1 : la décomposition de Schmidt fournit bien le choix le plus économique en termes de volume de données à faire passer d'Alice vers Bob (resp. de Bob vers Alice) pour que Bob (resp. Alice) puisse identifier ses états relatifs aux mesures effectuées par l'autre.

5.4 Mesures et intrication

Dans le chapitre 1, nous avons introduit la notion de mesure idéale (projective) et nous avons mentionné le fait qu'en pratique la plupart des mesures

n'étaient pas de ce type. Cependant, nous n'avions pas discuté comment aller au-delà de la mesure projective idéale. Le but de cette section est de montrer comment, grâce à l'intrication, il est possible de dépasser la seule notion de la mesure projective d'une observable du système pour définir celle de mesure généralisée. L'idée, qui n'est pas surprenante dans le contexte du présent chapitre, consiste à effectuer des mesures projectives non pas sur le système lui-même mais dans son environnement. Après avoir construit ainsi les mesures généralisées, nous examinerons leurs caractéristiques à la lumière de l'information qu'elles permettent de retrouver sur l'état quantique du système observé.

5.4.1 Qu'est-ce qu'une mesure généralisée ?

Définition

Une *mesure généralisée* consiste à effectuer une mesure projective dans l'environnement d'un système après avoir intriqué le système avec son environnement. On commence donc par partir d'un système préparé dans un état quantique $|\psi\rangle \in \mathcal{H}_S$ et d'un environnement initialement dans un état $|E_0\rangle$. On procède ensuite en deux temps. D'abord, on intrique le système et son environnement au moyen d'un opérateur unitaire U_M qui agit sur l'espace des états total $\mathcal{H}_S \otimes \mathcal{H}_E$. On effectue alors une mesure projective dans l'environnement, ce qui revient à chercher la décomposition de $U_M(|\psi\rangle|E_0\rangle)$ en états relatifs par rapport à une base orthonormée $|m_i\rangle$ de $\mathcal{H}_E$. Cette décomposition s'obtient en insérant $\mathbb{1}_E = \sum_i |m_i\rangle\langle m_i|$:

$$U_M(|\psi\rangle|E_0\rangle) = \sum_i \langle m_i|U_M(|\psi\rangle|E_0\rangle) |m_i\rangle \quad (5.61a)$$

$$= \sum_i (M_i|\psi\rangle) \otimes |m_i\rangle, \quad (5.61b)$$

où l'opérateur M_i agissant sur $\mathcal{H}_S$ est donné par :

$$M_i|\psi\rangle = \langle m_i|U_M(|\psi\rangle|E_0\rangle). \quad (5.62)$$

Ces opérateurs caractérisent complètement le processus d'intrication puis mesure projective dans l'environnement. La probabilité de trouver le résultat m_i est alors égale à $\langle M_i^\dagger M_i \rangle_{|\psi\rangle}$ et l'état normé après mesure n'est autre que l'état relatif à m_i :

$$|\psi(S|m_i)\rangle = \frac{M_i|\psi\rangle}{\sqrt{\langle M_i^\dagger M_i \rangle_{|\psi\rangle}}}. \quad (5.63)$$

En utilisant de nouveau la relation de fermeture, nous voyons que

$$\sum_i M_i^\dagger M_i = \mathbb{1}_{\mathcal{H}_S}. \quad (5.64)$$

Cette relation fondamentale assure que la somme des probabilités des différents résultats de mesure est bien l'unité. Cependant, compte tenu de la définition (5.62), il est clair qu'il existe beaucoup plus de résultats de mesure différents que la dimension de l'espace des états du système.

On voit donc apparaître une nouvelle notion de mesure qui, comme la mesure projective, donne lieu à une évolution stochastique du vecteur d'état du système. Cependant, et c'est là le point crucial, les différents états correspondant à différents résultats de mesure ne sont pas deux à deux orthogonaux.

Avant de discuter ce point plus précisément, remarquons que, si on ne s'intéresse pas aux états post-mesure mais seulement aux probabilités d'obtention des différents résultats, il peut être pertinent de considérer les opérateurs hermitiens et positifs $E_i = M_i^\dagger M_i$ qui vérifient $\sum_i E_i = \mathbb{1}$. Un tel jeu d'opérateurs est appelé en anglais un *positive operator valued measurement* (POVM) et permet de définir des probabilités sur l'ensemble des résultats de mesure par $P(m_i|\psi) = \langle E_i \rangle_{|\psi\rangle}$.

Notons que si on part d'un processus de mesure non idéale donné par un ensemble d'opérateurs M_i satisfaisant (5.64), alors on construit le POVM par $E_i = M_i^\dagger M_i$. On peut montrer que dans ce cas, il existe des opérateurs unitaires U_i tels que $M_i = U_i \sqrt{E_i}$ où la racine carrée $\sqrt{E_i}$ est bien définie du fait de la positivité des opérateurs E_i . L'intérêt de ce formalisme réside dans le fait qu'il permet de discuter en termes simples toutes les situations où la seule chose qui importe est la statistique des résultats des mesures.

Les mesures généralisées sur un qubit

Pour rendre cette discussion plus concrète, considérons les mesures généralisées sur un qubit. Tout d'abord, les mesures généralisées d'un qubit avec deux valeurs possibles sont forcément des mesures projectives. En effet, notons M_0 et M_1 les deux opérateurs qui définissent cette mesure, nous avons :

$$M_0^\dagger M_0 + M_1^\dagger M_1 = \mathbb{1}. \quad (5.65)$$

Les opérateurs $E_j = M_j^\dagger M_j$ étant hermitiens, positifs et de rang un, on peut les écrire sous la forme

$$M_j^\dagger M_j = \frac{\alpha_j}{2} (\mathbb{1} + \mathbf{n}_j \cdot \boldsymbol{\sigma}), \quad (5.66)$$

où α_0 et α_1 sont deux nombres réels positifs et les vecteurs $\mathbf{n}_k$ normés. De plus la condition ci-dessus entraîne que $\alpha_0 + \alpha_1 = 1$ et que les vecteurs $\mathbf{n}_0$ et $\mathbf{n}_1$ vérifient $\alpha_0 \mathbf{n}_0 + \alpha_1 \mathbf{n}_1 = 0$. En conséquence, les deux vecteurs sont opposés et les α_j sont égaux à l'unité. Ceci montre que la mesure généralisée avec deux résultats est projective. Le premier exemple non trivial de mesure généralisée sur un qubit comporte donc trois résultats.

Considérons une mesure généralisée avec trois résultats correspondant aux trois opérateurs hermitiens $E_{0,1,-1}$, positifs et de rang 1. En utilisant

l'équation (5.64) pour les opérateurs $E_k = M_k^\dagger M_k$ de la forme donnée par (5.66) où les α_k ($k \in \{0, 1, -1\}$) sont des réels positifs et les $\mathbf{n}_k$ sont des vecteurs normés, nous obtenons :

$$\alpha_0 + \alpha_1 + \alpha_{-1} = 2, \quad (5.67a)$$

$$\alpha_0 \mathbf{n}_0 + \alpha_1 \mathbf{n}_1 + \alpha_{-1} \mathbf{n}_{-1} = \mathbf{0}. \quad (5.67b)$$

Prenons un exemple symétrique pour lequel les trois coefficients $\alpha_{0,1,-1}$ sont égaux à $\alpha = 2/3$. Dans ce cas, les trois vecteurs $\mathbf{n}_{0,1,-1}$ de la sphère de Bloch somment au vecteur nul $\mathbf{0}$. Un choix possible est :

$$|\mathbf{n}_0\rangle = |+\rangle, \quad (5.68a)$$

$$|\mathbf{n}_{\pm}\rangle = \pm \frac{1}{2} |+\rangle + \frac{\sqrt{3}}{2} |-\rangle, \quad (5.68b)$$

de sorte que les trois vecteurs définissent un triangle équilatéral dans le plan $(\mathbf{e}_z, \mathbf{e}_x)$ et dont $\mathbf{e}_z$ est un des sommets. Il est bien sûr possible de réaliser cette mesure généralisée au moyen d'une mesure projective dans un environnement formé d'un système à trois états, ce que nous laissons comme exercice pour le lecteur.

Ceci étant, il est possible de faire cette mesure en considérant un système de deux qubits et en effectuant des mesures projectives sur ce système global. Pour cela, il faut trouver le moyen de relever les trois états $|\mathbf{n}_{0,\pm 1}\rangle$ en trois états orthogonaux dans l'espace des états à deux qubits. Pour les trois états précédents, nous avons $|\langle \mathbf{n}_a | \mathbf{n}_b \rangle|^2 = 1/2$. Une possibilité consiste à prendre des états $|\chi_a\rangle = \alpha |\mathbf{n}_a\rangle \otimes |+\rangle + \beta |+\rangle \otimes |-\rangle$ avec $\alpha^2 + \beta^2 = 1$. La condition d'orthogonalité des états $|\chi_a\rangle$ nous donne $\alpha = \sqrt{2/3}$ et $\beta = 1/\sqrt{3}$. Ces trois états sont cependant tous orthogonaux à l'état $|\chi_2\rangle = |-\rangle \otimes |-\rangle$. On a donc obtenu une base orthonormée de l'espace des états à deux qubits : $(|\chi_2\rangle, |\chi_0\rangle, |\chi_{\pm}\rangle)$. Les opérateurs décrivant la mesure généralisée sont alors définis par $M_a = \langle +_B | \Pi_{|\chi_a\rangle} | +_B \rangle$ pour $a = 0, \pm 1, 2$ sont alors :

$$M_a^\dagger M_a = \frac{2}{3} |\mathbf{n}_a\rangle \langle \mathbf{n}_a| \quad (a = 0, \pm 1), \quad (5.69a)$$

$$M_2^\dagger M_2 = 0. \quad (5.69b)$$

Cela correspond bien à la mesure généralisée envisagée précédemment même si quatre résultats sont en principe possibles. Mais ici, le résultat supplémentaire associé à l'état $|\chi_2\rangle$ a une probabilité nulle.

Une idée similaire est utilisée pour discuter du protocole d'estimation d'un qubit à partir de deux réalisations. Dans cet exemple, une mesure projective sur le système des deux qubits définit pour chacun des deux une mesure projective dont les résultats sont associés à quatre vecteurs qui définissent un tétraèdre régulier sur la sphère de Bloch. Le protocole d'estimation qui consiste à choisir comme état estimé celui correspondant au résultat de cette mesure permet d'atteindre la borne de Massard et Popescu discutée à la section 4.3.3.

Enfin, nous verrons dans le chapitre 6 qu'une mesure généralisée à trois résultats sur un qubit est ce qui permet d'obtenir une discrimination sans erreur la plus efficace entre deux états non orthogonaux. Là aussi, cela illustre la puissance de ce type de mesure qui ne peut être obtenu qu'en élargissant le système. Dans le domaine quantique, cela permet de faire mieux alors que dans le domaine classique, effectuer des mesures sur un système élargi est synonyme de rajouter du bruit et donc ne permet pas d'espérer un gain.

La photodétection comme mesure généralisée

La détection des photons par un photo-détecteur est un exemple concret de mesure généralisée. Considérons un mode du champ électromagnétique qui est idéalement décrit par un oscillateur harmonique quantifié (section 4.2). On note b et $b^\dagger$ les opérateurs de destruction et de création d'un photon dans ce mode et $|n\rangle$ l'état normé comprenant n photons. En pratique, une manière de compter les photons consiste à les détecter en les absorbant. Un tel détecteur donne deux résultats : 0 s'il ne détecte aucun photon et 1 s'il en absorbe un. Si notre détecteur possède une efficacité de 100 % durant sa fenêtre de fonctionnement et que l'on sait juste qu'il a absorbé à coup sûr un seul photon, le résultat 1 correspond à l'opérateur⁷ :

$$M_1 = \sum_{n=1}^{+\infty} |n-1\rangle\langle n|. \quad (5.70)$$

On voit alors que $M_1^\dagger M_1 = \mathbb{1} - |0\rangle\langle 0|$ et la relation (5.64) conduit alors naturellement à poser

$$M_0 = |0\rangle\langle 0|. \quad (5.71)$$

Le seul état laissé invariant par cette mesure non projective est donc l'état vide $|0\rangle$: tous les autres sont affectés par l'appareil de mesure. En particulier, notons que sur le sous-espace des états à 0 ou 1 photon, l'état final après mesure est toujours l'état vide $|0\rangle$. Ceci montre bien qu'une telle mesure ne peut être répétée ce qui serait possible pour une mesure projective : en effet, une mesure projective du nombre de photons ne changerait pas ce nombre ; il les compterait sans les absorber. Ce n'est pas le cas des détecteurs que nous rencontrons dans la vie courante, à commencer par les cellules photosensibles de notre œil.

5.4.2 Mesures faibles *versus* mesures projectives

Revenons maintenant sur la question générale suivante : à quel point une mesure généralisée engendre-t-elle des états deux à deux orthogonaux ? Tout

7. C'est évidemment un modèle très simplifié et pas vraiment réaliste.

dépend de l'intrication engendrée entre le système et son environnement par l'action de l'opérateur unitaire U_M .

Un cas limite évident est celui où cet opérateur n'engendre pas d'intrication du tout : $U_M = U_S \otimes U_E$ où $U_{E/S}$ sont des opérateurs unitaires agissant respectivement sur $\mathcal{H}_S$ et $\mathcal{H}_E$. Dans ce cas, nous avons $M_i = \langle m_i | E_0 \rangle U_S$. Nous obtenons certes différents résultats de mesure, mais avec des probabilités indépendantes de l'état du système avant-mesure $|\psi\rangle$ et surtout, un état post-mesure qui, à une phase près, se trouve être $U_S|\psi\rangle$. En fait, la mesure ne nous apprend rien et elle ne change pas l'état du système. Cela n'est pas surprenant car nous n'avons pas intriqué le système et l'environnement dans lequel nous effectuons les mesures. Ce dernier ne contient donc aucune empreinte sur l'état $|\psi\rangle$.

L'autre cas extrême est celui où l'opérateur U_M produit directement une décomposition de Schmidt pour l'état intriqué du système et de son environnement. Dans ce cas, le processus de mesure engendre bien des états post-mesure deux à deux orthogonaux. Il suffit pour cela de choisir une transformation unitaire U_M qui envoie chaque vecteur $|s_i\rangle$ d'une base orthonormée de $\mathcal{H}_S$ sur $|s_i\rangle \otimes |m_i\rangle \in \mathcal{H}_S \otimes \mathcal{H}_E$ où les $|m_i\rangle$ sont des vecteurs normés deux à deux orthogonaux dans $\mathcal{H}_E$. Dans ce cas, les opérateurs M_i sont les projecteurs orthogonaux sur les différents vecteurs $|s_i\rangle$ et on retrouve une mesure projective par rapport à cette base orthonormée de $\mathcal{H}_S$.

D'une manière générale, les mesures généralisées interpolent entre ces deux cas limites. Par exemple, on peut chercher à s'éloigner des mesures projectives en imaginant un processus d'intrication entre le système et son environnement qui laisse une grande ambiguïté au niveau des états relatifs associés à un même résultat de mesure. C'est le cas si l'opérateur U_M envoie un certain nombre d'états deux à deux orthogonaux $|s_j\rangle$ du système sur $|s'_j\rangle \otimes |m\rangle$ où les $|s'_j\rangle$ sont également deux à deux orthogonaux : l'orthogonalité est préservée mais différents états du système correspondent au même résultat de mesure associé à l'état $|m\rangle$ dans l'environnement. Nous allons voir que la détection usuelle de photons appartient à cette classe.

À l'opposé, on peut chercher à perturber le système à minima. Cela nécessite des expériences dans lesquelles le couplage entre le système et son environnement est assez faible et bien contrôlé. C'est le domaine des *mesures faibles* qui a fait l'objet d'un intérêt considérable ces dernières décennies.

Regardons comment cela fonctionne. Supposons que l'opérateur U_M intrique faiblement le système et son environnement. Par exemple, lorsque

$$U_M = \exp(i\varepsilon S \otimes E), \quad (5.72)$$

où S et E sont deux opérateurs hermitiens respectivement sur le système et son environnement et où ε est un paramètre réel sans dimension qui caractérise la force du couplage. Une telle forme apparaît lorsque l'on considère une interaction durant un temps fini : ε peut alors être vu comme un temps convenablement adimensionné et $S \otimes E$ comme le hamiltonien d'interaction

convenablement adimensionné. Un exemple concret de mesure de ce type pour un qubit sera discuté dans la section 10.2.5 (page 441).

Diagonalisons $S = \sum_i s_i |s_i\rangle\langle s_i|$ dans une base orthogonale et partons d'un état $|\psi\rangle = \sum_i \alpha_i |s_i\rangle$ pour le système et de $|E_0\rangle$ pour l'environnement. Alors, après la mesure, l'état est

$$|\Psi\rangle = \sum_i \alpha_i |s_i\rangle \otimes e^{i\varepsilon s_i E} |E_0\rangle. \quad (5.73)$$

Introduisons les états $|E_i\rangle = e^{i\varepsilon s_i E} |E_0\rangle$. Si ces différents états étaient deux à deux orthogonaux, nous aurions une décomposition de Schmidt de l'état intriqué $|\Psi\rangle$ et donc une mesure projective. Le cas qui nous intéresse ici est plutôt celui où le couplage est suffisamment faible pour que le système et son environnement soient peu intriqués. On procède alors comme précédemment en considérant les mesures relativement à une base $|x\rangle$ de $\mathcal{H}_E$ et les opérateurs M_x correspondants sont alors donnés par :

$$M_x = \langle x | e^{i\varepsilon S \otimes E} | E_0 \rangle. \quad (5.74)$$

Nous pouvons alors calculer la distribution de probabilité des résultats de mesure. Considérons pour commencer le cas où le système a été préparé initialement dans un état propre de S , soit $|\psi\rangle = |s_i\rangle$. La distribution de probabilité p_i des résultats de mesure est alors donnée par

$$p_i(x) = \left| \langle x | e^{i\varepsilon s_i E} | E_0 \rangle \right|^2. \quad (5.75)$$

Lorsque ε est petit, ce qui est le cas ici, les différentes distributions de probabilités p_i ont un recouvrement important. Tout l'enjeu dans une mesure faible consiste donc à savoir discerner entre deux distributions de probabilités différentes. Pour cela, il faut travailler avec un échantillon statistique de plusieurs résultats de mesure car, en examinant un unique résultat de mesure, on ne peut pas savoir de quelle loi il est issu. Imaginons que l'on dispose de deux sources, l'une associée à l'état $|s_i\rangle$ et l'autre à $|s_j\rangle$ et que l'on doive discerner entre p_i et p_j . Intuitivement, chaque résultat de mesure contribue à nous informer sur le fait que l'on a affaire à une distribution ou à l'autre et on s'attend à ce qu'il existe un nombre minimal de mesures N_{mes} nécessaire pour discerner entre ces deux distributions avec une probabilité élevée. Comme expliqué sur la figure 5.4, ce nombre est d'autant plus élevé que les distributions de probabilités p_i et p_j sont similaires.

D'autre part, le processus de mesure faible intrique le système et l'ensemble des degrés de liberté de son environnement. L'information sur la préparation du système dans tel ou tel état pur est donc maintenant partagée avec lui. Mesurer la quantité X associée à la base $|x\rangle$ donne accès à une partie de cette information mais le choix de X n'est peut-être pas optimal. Cela suggère qu'il existe une borne inférieure de N_{mes} qui ne dépend que du couplage entre le système et l'environnement : cette valeur constitue une limite quantique au processus de mesure faible.

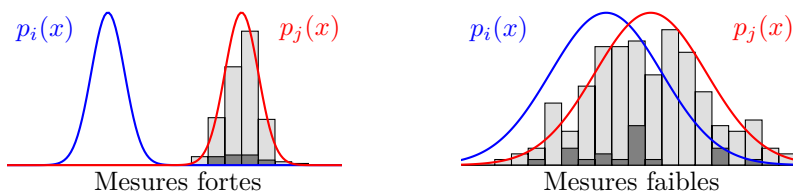


FIGURE 5.4 – Histogrammes représentant 20 (gris foncé) et 200 (gris clair) tirages d’une variable aléatoire distribuée selon la loi de probabilité p_i . Pour cette illustration, les deux lois p_i et p_j sont des gaussiennes d’écart types égaux mais de moyennes différentes. À gauche : cas où les lois de probabilités sont bien séparées (mesures projectives). À droite : cas où les lois de probabilité se recouvrent (mesures faibles). Dans le cas des mesures projectives, 20 tirages permettent de discriminer sans ambiguïté entre p_i et p_j alors que pour le cas des mesures faibles il en faut 200.

5.4.3 Quelle information est extraite du système ?

Pour mieux comprendre l’origine de cette limite quantique dans les mesures faibles, nous allons maintenant analyser le processus de mesure généralisé en termes de théorie de l’information.

Information gagnée *via* la mesure

Commençons par déterminer quelle est l’information gagnée par le processus de mesure. Plus précisément, quelle est l’information récupérable sur le choix de l’état $|s_{i,j}\rangle$ à partir de la statistique des résultats de mesures de l’observable X dans l’environnement ?

C’est exactement un problème de théorie de la communication dans lequel nous identifions l’émission E au choix, pour l’état initial du système, de l’un ou l’autre des états propres de l’observable S et la réception R aux résultats de mesures de X . La quantité d’information transmise est alors l’information mutuelle $I[E, R] = S[E] + S[R] - S[E, R]$ (section 2.5.2). Considérons pour simplifier un exemple dans lequel l’émetteur choisit de façon équiprobable entre deux états $|s_i\rangle$ et $|s_j\rangle$. L’information associée à l’émetteur est $S[E] = 1$ bit par message envoyé. Le récepteur reçoit des messages x dont la distribution de probabilité est

$$p_R(x) = \frac{p_i(x) + p_j(x)}{2}. \quad (5.76)$$

Enfin, la donnée de l’émission et de la réception nous donne une source de messages (s, x) où $s = i, j$ avec $p(i, x) = p_i(x)/2$ et $p(j, x) = p_j(x)/2$. L’information mutuelle est alors donnée par

$$I_{\text{mes}} = I[E, R] = \sum_x \left[p_i(x) \log \left(\frac{p_i(x)}{p_R(x)} \right) + p_j(x) \log \left(\frac{p_j(x)}{p_R(x)} \right) \right]. \quad (5.77)$$

Cette quantité représente l'information gagnée en bits lors du processus de mesure. Elle dépend non seulement du couplage entre le système et son environnement mais aussi du choix de l'observable X définissant le processus de mesure généralisée. Remarquons que si les deux distributions de probabilités sont totalement séparées, c'est-à-dire que $\forall x, p_i(x)p_j(x) = 0$, alors $p_R(x) = p_j(x)/2$ (resp. $p_i(x)/2$) dès lors que $p_i(x) = 0$ (resp. $p_j(x) = 0$) et donc $I[E, R] = 1$ bit. Dit autrement, lorsque les deux distributions sont bien séparées, on gagne bien un bit d'information à chaque mesure.

Dans le cas des mesures faibles, nous pouvons développer cette expression en la différence $\Delta p(x) = (p_j - p_i)(x)$ et le premier ordre non trivial est

$$I[E, R] \approx \sum_x \frac{(\Delta p(x))^2}{4p_R(x)} \quad (5.78)$$

et donc, lors de chaque mesure, on gagne bien moins qu'un bit d'information et même aucun dès lors que $p_i = p_j$. Dans ce régime, le nombre minimal de mesures N_{mes} nécessaire pour récupérer le bit d'information correspondant à la préparation de $|s_i\rangle$ ou de $|s_j\rangle$ est donc $N_{\text{mes}} = 1/I[E, R]$.

Information perdue dans l'environnement

Pour quantifier l'efficacité du processus de mesure, il nous faut comparer la quantité (5.77) avec la quantité d'information maximum que l'on pourrait récupérer au moyen de mesures dans l'environnement. En effet, dans notre exemple, si on trace sur le système, l'opérateur densité pour l'environnement est un mélange statistique équiprobable des deux états $|E_i\rangle$ et $|E_j\rangle$ qui ne sont *a priori* pas orthogonaux. Si on effectue alors une mesure projective, voire généralisée, M sur cet environnement, nous pouvons donc récupérer comme information $I[E, M]$ où M correspond à la mesure réalisée sur l'environnement. La quantité d'information maximale que l'on peut alors extraire de l'environnement est $I_{\text{env}} = \max_M (I[E, M])$, qui est clairement supérieure à la quantité que nous avons calculée précédemment car R correspond à un des choix de mesures possibles dans l'environnement.

Tout le problème consiste donc à calculer I_{env} , ce qui est loin d'être évident car il faut prendre la borne supérieure sur toutes les mesures possibles. Nous allons, pour le moment, en donner une estimation dans l'exemple qui nous intéresse ici. Lorsqu'ils sont différents, les deux états $|E_i\rangle$ et $|E_j\rangle$ engendrent un espace des états $\mathbb{C}^2$ de sorte qu'on peut les voir comme deux états *a priori* non orthogonaux d'un qubit, c'est-à-dire deux vecteurs de Bloch distincts $\mathbf{n}_i$ et $\mathbf{n}_j$ sur la sphère de Bloch. Notre problème consiste donc à déterminer quelle mesure nous fournit l'information maximale sur un mélange de deux états non orthogonaux d'un qubit. Pour commencer, considérons une mesure projective qui possède donc deux résultats correspondant à deux vecteurs opposés $\pm \mathbf{m}$ sur la sphère de Bloch. Nous devons donc chercher quel choix de $\pm \mathbf{m}$ nous apprend le maximum de choses sur les deux états $\mathbf{n}_i$ et $\mathbf{n}_j$.

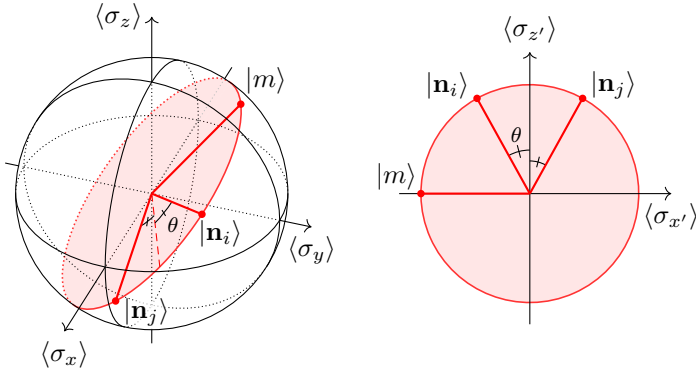


FIGURE 5.5 – Une représentation schématique de la procédure permettant d’acquérir le plus d’information possible sur les états $|E_i\rangle$ et $|E_j\rangle$. Ces états peuvent être représentés sur la sphère de Bloch comme deux vecteurs $\mathbf{n}_{i,j}$, et le choix de mesure fournissant la plus grande information est donné par le vecteur $\mathbf{m}$ correspondant à la direction (Ox') dans la base $(Ox'y'z')$ obtenue par l’équation (5.81).

Rappelons-nous que les probabilités des résultats de mesures $\pm \mathbf{m}$ sont données par :

$$p(\mathbf{m}|\mathbf{n}_i) = \frac{1 + \mathbf{m} \cdot \mathbf{n}_i}{2} \text{ et } p(-\mathbf{m}|\mathbf{n}_i) = \frac{1 - \mathbf{m} \cdot \mathbf{n}_i}{2}, \quad (5.79a)$$

$$p(\mathbf{m}|\mathbf{n}_j) = \frac{1 + \mathbf{m} \cdot \mathbf{n}_j}{2} \text{ et } p(-\mathbf{m}|\mathbf{n}_j) = \frac{1 - \mathbf{m} \cdot \mathbf{n}_j}{2}. \quad (5.79b)$$

Intuitivement, nous voulons choisir $\pm \mathbf{m}$ pour discriminer au maximum les vecteurs $\mathbf{n}_{i,j}$, c’est-à-dire que nous cherchons à rendre les distributions de probabilités $p(\cdot|\mathbf{n}_i)$ et $p(\cdot|\mathbf{n}_j)$ les plus différentes possibles. Cela revient donc à maximiser $\mathbf{m} \cdot (\mathbf{n}_i - \mathbf{n}_j)$. Le choix naturel est alors

$$\mathbf{m} = \frac{\mathbf{n}_i - \mathbf{n}_j}{\|\mathbf{n}_i - \mathbf{n}_j\|}. \quad (5.80)$$

Étant donné deux vecteurs distincts sur la sphère de Bloch, nous pouvons toujours choisir une nouvelle base pour nous ramener au cas

$$|\mathbf{n}_{i,j}\rangle = \cos \theta/2 |Z'_+\rangle \pm \sin \theta/2 |Z'_-\rangle, \quad (5.81)$$

où $(|Z'_+\rangle, |Z'_-\rangle)$ est une base orthonormée correspondant aux vecteurs $\mathbf{e}_{z'}$ et $-\mathbf{e}_{z'}$ sur la sphère de Bloch, et $\theta \in]0, \pi/2]$. Les deux vecteurs $\mathbf{n}_{i,j}$ sont donc situés dans le plan $(\mathbf{e}_{x'}, \mathbf{e}_{z'})$, symétriques par rapport à $\mathbf{e}_{z'}$ et séparés par un angle θ de celui-ci, de sorte que $\mathbf{m} = \mathbf{e}_{x'}$ et $|\langle \mathbf{n}_i | \mathbf{n}_j \rangle|^2 = \cos^2 \theta$. Cette configuration est représentée sur la figure 5.5.

Un calcul direct donne alors l'information mutuelle associée $I[E, M]$ pour ce processus de mesure :

$$I[E, M] = \frac{1 + \sin \theta}{2} \log(1 + \sin \theta) + \frac{1 - \sin \theta}{2} \log(1 - \sin \theta). \quad (5.82)$$

Le résultat remarquable que nous admettrons pour le moment [255] est que le membre de droite de (5.82) est bien égal à I_{env} . Cette fonction croît de 0, lorsque $\theta = 0$, à 1 pour $\theta = \pi/2$. En utilisant les approximations $I[E, M] \approx \theta^2$ et $\sin \theta = \sqrt{1 - |\langle \mathbf{n}_i | \mathbf{n}_j \rangle|^2}$, valables quand θ est petit, nous obtenons une formule simple de l'information maximale extractible en fonction du recouvrement des états $|E_{i,j}\rangle$.

La limite quantique aux mesures généralisées

La discussion ci-dessus nous a permis d'introduire deux quantités informationnelles I_{mes} et I_{env} qui représentent respectivement la quantité d'information récupérée lors du processus de mesure et la quantité d'information maximale récupérable dans l'environnement. La limitation fondamentale sur le processus de mesure quantique est tout simplement que

$$I_{\text{mes}} \leq I_{\text{env}}. \quad (5.83)$$

Dit autrement, si un processus de mesure généralisé n'est pas optimal, c'est uniquement parce que l'on a pas capturé toute l'information qui est disponible dans l'environnement. L'inégalité ci-dessus nous donne donc une borne inférieure au nombre de mesures qu'il faut faire pour discerner entre la préparation de $|s_i\rangle$ et celle de $|s_j\rangle$: $N_{\text{mes}} \geq 1/I_{\text{env}}$. On voit donc naturellement émerger la limite quantique que nous avons qualitativement discutée précédemment mais ici, c'est un véritable critère quantitatif qui est obtenu.

Dans le cas du modèle de mesure faible discuté dans la section 5.4.2, nous pouvons obtenir une expression explicite du nombre de mesures nécessaires pour discerner entre les deux états orthogonaux $|s_i\rangle$ et $|s_j\rangle$. En effet, dans le processus de mesure faible, les états $|E_i\rangle$ et $|E_j\rangle$ sont proches : leur recouvrement est proche de l'unité et l'information maximale extractible de l'environnement est alors donnée par $I_{\text{env}} \approx \theta^2$ avec $|\langle E_i | E_j \rangle| \approx 1 - \theta^2/2$. Un calcul perturbatif à l'ordre deux de ce recouvrement (détaillé dans la section suivante) nous montre alors que

$$I_{\text{env}} \approx \theta^2 \approx \varepsilon^2 S_E (s_i - s_j)^2, \quad (5.84)$$

où $S_E = \langle E^2 \rangle_{|E_0\rangle} - \langle E \rangle_{|E_0\rangle}^2$ représente la fluctuation de quantique de E dans l'état $|E_0\rangle$. On obtient donc une borne inférieure d'origine quantique au nombre minimal de mesures qu'il faut effectuer pour discerner entre la préparation de $|s_i\rangle$ et $|s_j\rangle$:

$$N_{\text{mes}} \geq \frac{1}{\varepsilon^2 S_E (s_i - s_j)^2}. \quad (5.85)$$

Cette limite quantique est d'autant plus élevée que ε est petit, c'est-à-dire que la mesure est faible. Elle est également d'autant plus petite que les valeurs s_i et s_j sont proches. Cette limite pour les mesures généralisées trouve bien son origine dans la théorie de l'information : elle provient du fait que l'on ne récupère pas toute l'information potentiellement accessible dans l'environnement.

5.4.4 Approche perturbative

Pour finir, nous allons présenter une approche du problème de la limite quantique des mesures faibles en discutant les deux points mentionnés précédemment au travers d'un développement perturbatif en ε . Comme nous le verrons, il est possible de calculer les deux informations I_{mes} et I_{env} et de vérifier explicitement que l'inégalité (5.83) est vérifiée perturbativement en ε .

Nombre de mesures minimales

Pour cela, regardons comment les différentes distributions p_j diffèrent par leur valeur moyenne et par leur écart type à l'ordre le plus bas en ε . Au premier ordre, nous avons

$$p_j(x) \approx p_0(x) + s_j \varepsilon \langle i[\Pi_x, E] \rangle_{|E_0\rangle} + \mathcal{O}((\varepsilon s_j)^2), \quad (5.86)$$

où nous avons introduit le projecteur orthogonal $\Pi_x = |x\rangle\langle x|$. Or, dès lors que l'observable $X = \sum_x x \Pi_x$ ne commute pas avec E (ce qui est le cas intéressant ici), la valeur moyenne des résultats expérimentaux x obtenus dépend de la valeur propre de S . Perturbativement en ε , nous avons

$$\mathbb{E}_{p_j}(x) = \mathbb{E}_{p_0}(x) + i\varepsilon s_j \langle [X, E] \rangle_{|E_0\rangle} + \mathcal{O}((\varepsilon s_j)^2). \quad (5.87)$$

Le coefficient

$$\chi = \left(\frac{\partial \mathbb{E}_{p_j}(x)}{\partial (s_j \varepsilon)} \right)_{(\varepsilon s_j) \approx 0} = i \langle [X, E] \rangle_{|E_0\rangle} \quad (5.88)$$

caractérise donc la réponse au premier ordre en ε de cette valeur moyenne à la présence du système dans un état propre $|s_j\rangle$. Si on fait une expérience de contrôle dans laquelle on envoie deux états propres $|s_i\rangle$ et $|s_j\rangle$ distincts, on peut espérer les discerner pratiquement dès la première mesure à la condition que les valeurs moyennes de X correspondant à $s_i \neq s_j$ soient bien séparées, c'est-à-dire par plus que les fluctuations statistiques des distributions p_i et p_j . À l'ordre zéro, ces fluctuations sont données par

$$S_X \approx \langle X^2 \rangle_{|E_0\rangle} - \langle X \rangle_{|E_0\rangle}^2. \quad (5.89)$$

Nous voyons donc que nous obtenons des distributions de même écart type mais dont les moyennes diffèrent linéairement en ε . Analysons maintenant comment discerner deux distributions de probabilités p_j différentes au moyen

de N mesures effectuées sur des réalisations identiques. Le théorème central limite nous dit alors que la variable aléatoire $X_N = \sum_{j=1}^N x_j / \sqrt{N}$, où les x_j sont les résultats des N mesures, converge en loi vers une gaussienne dont la valeur moyenne est $\sqrt{N} \mathbb{E}_{p_j}(x)$ et la fluctuation est égale à celle de x pour la loi p_j . À la limite $\varepsilon \ll 1$, la séparation des valeurs moyennes de X_N pour les mesures faites avec $|s_i\rangle$ et $|s_j\rangle$ se comporte en $\varepsilon |s_i - s_j| \chi \sqrt{N}$ alors que les écarts types de ces variables aléatoires sont $\sqrt{S_X}$ donné par (5.89). On peut donc estimer le nombre minimal de mesures nécessaires pour discerner entre $|s_i\rangle$ et $|s_j\rangle$ en écrivant que $|\mathbb{E}_{p_i}(X_N) - \mathbb{E}_{p_j}(X_N)| \gtrsim 2\sqrt{S_X}$, ce qui donne⁸ :

$$N_{\text{mes}} \approx \frac{4S_X}{(\varepsilon |s_i - s_j| \chi)^2}. \quad (5.90)$$

Dans la limite $\varepsilon \rightarrow 0$, N_{mes} devient effectivement grand devant un.

✱ Effet des mesures sur les cohérences

Maintenant que nous avons vu quelle était la taille minimale de l'échantillon typique nécessaire pour discerner entre p_i et p_j pour $i \neq j$, analysons l'effet de ces mesures sur le système. L'état du système ne peut être décrit que comme un mélange statistique qui est obtenu en prenant la trace sur l'environnement du projecteur sur l'état intriqué $|\Psi\rangle$. Il est donc décrit par l'opérateur densité réduit

$$\rho_{\text{post}} = \sum_{i,j} \alpha_i \alpha_j^* \langle E_j | E_i \rangle |s_i\rangle \langle s_j|. \quad (5.91)$$

Dans la base orthonormée $|s_i\rangle$ diagonalisant l'observable S , nous voyons ainsi apparaître une matrice dont les éléments diagonaux ne sont pas modifiés mais dont les éléments hors diagonaux sont altérés par le coefficient

$$\mathcal{D}(i, j) = \langle E_j | E_i \rangle = \left\langle e^{i\varepsilon(s_i - s_j)E} \right\rangle_{|E_0\rangle}. \quad (5.92)$$

Ce coefficient encode un phénomène de décohérence, c'est-à-dire d'atténuation des interférences quantiques, et sera étudié en détail à partir du chapitre 9. Comme précédemment, analysons cela perturbativement en ε . Le module de $\mathcal{D}(i, j)$ est, au premier ordre non trivial en ε , donné par

$$|\mathcal{D}(i, j)| \approx 1 - \frac{\varepsilon^2 (s_i - s_j)^2}{2} S_E, \quad (5.93)$$

où $S_E = \langle E^2 \rangle_{|E_0\rangle} - \langle E \rangle_{|E_0\rangle}^2$ désigne la fluctuation de la grandeur E dans l'état $|E_0\rangle$. C'est là qu'est caché le prix à payer associé pour la mesure faible :

8. Le facteur 2 vient du fait que l'on doit séparer nettement deux distributions de probabilité qui ont chacune une largeur de $\sqrt{S_X}$: il faut prendre une séparation double.

un état $|E_0\rangle$ qui serait état propre de E n'enregistrerait aucune information sur l'état $|s_i\rangle$. Les distributions de probabilités p_i et p_0 seraient identiques et la mesure n'apporterait donc rien. C'est parce que nous utilisons un état $|E_0\rangle$ qui est altéré lors de l'interaction avec le système que la mesure nous apporte quelque chose mais cela se paye par rétroaction sur l'état du système : $|\mathcal{D}(i, j)| < 1$ pour $i \neq j$. Pour N systèmes préparés respectivement dans $|s_i\rangle^{\otimes N}$ et $|s_j\rangle^{\otimes N}$, les cohérences entre ces deux états de N réalisations sont atténuées par $|\mathcal{D}(i, j)|^N$. On voit donc que la cohérence entre ces deux états quantiques pour N réalisations disparaît dès lors que $N \gtrsim N_{\text{dec}}$ qui est donné par

$$N_{\text{dec}} \approx \frac{1}{(\varepsilon|s_i - s_j|)^2 S_E}. \quad (5.94)$$

✂ La limite quantique

Les équations (5.90) et (5.94), bien que perturbatives, montrent qu'il y a des contraintes sur l'acquisition d'information dans un tel processus de mesure. En effet, l'équation (5.90) nous donne une borne inférieure N_{mes} pour bien séparer les distributions de probabilités $p_i(x)$ et $p_j(x)$ en N mesures, alors que (5.94) nous donne une borne supérieure N_{dec} pour ne pas trop détruire les cohérences entre $|s_i\rangle^{\otimes N}$ et $|s_j\rangle^{\otimes N}$:

$$N_{\text{mes}} \approx \frac{4S_X}{(\chi\varepsilon|s_i - s_j|)^2} \quad \text{et} \quad N_{\text{dec}} \approx \frac{1}{(\varepsilon|s_i - s_j|)^2 S_E}. \quad (5.95)$$

Leur rapport ne dépend plus de la séparation $|s_i - s_j|$ ni de la constante de couplage ε et est donné par

$$\frac{N_{\text{dec}}}{N_{\text{mes}}} \approx \frac{\chi^2}{4S_X S_E}. \quad (5.96)$$

Définissons $S_{XE} = \langle XE \rangle_{|E_0\rangle} - \langle X \rangle_{|E_0\rangle} \langle E \rangle_{|E_0\rangle}$. Nous pouvons considérer le produit hermitien positif sur l'espace des opérateurs défini par $(A, E) \mapsto \langle A^\dagger E \rangle_{|E_0\rangle}$. L'inégalité de Cauchy-Schwarz $|\langle A^\dagger E \rangle_{|E_0\rangle}|^2 \leq \langle A^\dagger A \rangle_{|E_0\rangle} \langle E^\dagger E \rangle_{|E_0\rangle}$ appliquée à $X - \langle X \rangle_{|E_0\rangle}$ et $E - \langle E \rangle_{|E_0\rangle}$ montre que

$$|S_{XE}|^2 \leq S_X S_E. \quad (5.97)$$

Or, comme $S_{XE}^* = S_{EX}$ et que $\chi = i\langle [X, E] \rangle_{|E_0\rangle} = -2\Im(S_{XE})$, nous voyons que $|S_{XE}|^2 = \chi^2/4 + \Re(S_{XE})^2 \leq S_X S_E$. Ceci entraîne donc que

$$\frac{N_{\text{dec}}}{N_{\text{mes}}} \lesssim \frac{\chi^2}{\chi^2 + 4\Re(S_{XE})^2} \leq 1. \quad (5.98)$$

Cette inégalité est la limite quantique sur l'efficacité du processus de mesure faible. Si nous imaginons que le processus de mesure est répété au cours

du temps sur une même réalisation et qu'à chaque fois, l'environnement est réinitialisé de manière à éliminer tous les effets de mémoire, nous pouvons voir N comme une mesure du temps d'interaction entre le système et le dispositif utilisé pour le sonder⁹. La limite quantique $N_{\text{dec}} \lesssim N_{\text{mes}}$ signifie alors que le temps nécessaire pour discriminer entre deux valeurs propres de l'observable S du système est supérieur au temps de décohérence dans la base propre $(|s_i\rangle)_i$. Bien qu'obtenu dans le cadre d'un modèle très simplifié, ce raisonnement est exactement celui qui conduit aux limites quantiques sur les temps de mesure et de décohérence établies pour des systèmes plus réalistes [100].

✂ Comparaison avec l'approche informationnelle

Pour clore cette section, réexaminons les quantités I_{mes} et I_{env} dans l'exemple considéré en utilisant une approche perturbative valable dans le régime de mesures faibles.

Ayant vu l'expression de l'information stockée dans l'environnement dans la section précédente, considérons maintenant l'information gagnée par le processus de mesure. Dans la limite des mesures faibles, les distributions de probabilité p_i et p_j sont très proches et $I_{\text{mes}} \approx \sum_x (\Delta p(x))^2 / 4p_R(x)$. Si on suppose que les distributions p_i et p_j sont gaussiennes, qu'elles ont pour fluctuations S_X et que leurs valeurs moyennes sont décalées de $\varepsilon(s_i - s_j)\chi$ où $\chi = i\langle [X, E] \rangle_{|E_0\rangle}$ est la fonction de réponse de X au couplage à E , alors nous obtenons au premier ordre non trivial en ε :

$$I_{\text{mes}} \approx \frac{(\varepsilon\chi)^2}{4S_X} (s_i - s_j)^2. \quad (5.99)$$

Le ratio $I_{\text{mes}}/I_{\text{env}}$ qui mesure l'efficacité du processus de mesure est alors donné par

$$\frac{I_{\text{mes}}}{I_{\text{env}}} \approx \frac{\chi^2}{4S_E S_X}. \quad (5.100)$$

Nous voyons donc apparaître le même ratio que dans le calcul de $N_{\text{dec}}/N_{\text{mes}}$ dont nous avons montré qu'il est inférieur à l'unité. Cette analyse remet en perspective la limite quantique (5.98) : nous avons interprété le rapport $N_{\text{dec}}/N_{\text{mes}}$ comme le rapport du temps de décohérence par le temps nécessaire pour faire la mesure. Nous venons ici de le faire apparaître comme le ratio entre le flux d'information extrait par la mesure et le flux d'information potentiellement récupérable dans l'environnement, ce qui correspond à la limite quantique définie en termes informationnels (5.83).

9. La limite $\varepsilon \ll 1$ signifie que l'on prend un pas de temps élémentaire très court et un couplage assez faible.

Chapitre 6

Impossibilités quantiques

À divers moments, nous avons signalé que les états quantiques diffèrent largement de leur contrepartie classique, notamment du fait de leur inaccessibilité *via* une unique mesure. Le but du présent chapitre est de donner une vision globale de ces différences essentielles en présentant plusieurs résultats, appelés théorèmes d'impossibilité, qui résultent des principes fondamentaux de la théorie quantique et qui mettent en lumière les différences essentielles entre états classiques et états quantiques.

Le résultat le plus important est sans doute le théorème de non-clonage d'un état quantique quelconque qui stipule qu'il est impossible pour un observateur recevant un état quantique de le cloner à moins qu'il ne dispose de sa recette de préparation. Ce résultat est équivalent à l'impossibilité de discerner à coup sûr deux états quantiques non orthogonaux. Ces deux théorèmes montrent clairement que l'information contenue dans un état quantique est très différente de celle contenue dans un état classique puisque ces deux opérations ne posent aucun problème dans notre monde classique. Ils expriment aussi l'inaccessibilité d'un état quantique à partir d'une seule réalisation physique de ce dernier. De façon tout à fait similaire, il est impossible d'effacer l'information contenue dans un état quantique ou de les cacher dans les corrélations entre deux systèmes.

Ces résultats suggèrent une permanence de l'information associée aux états quantiques qui ne peut donc être ni dupliquée ni détruite. Il est seulement possible de mettre un état quantique sous le tapis en le transférant du système que l'on étudie vers des degrés de liberté environnementaux.

6.1 Indiscernabilité des états non orthogonaux

Dans notre description de la théorie quantique, nous avons pour l'instant distingué deux aspects. Le premier nous donne la description de l'état d'un système physique et de son évolution *via* une équation du mouvement. Le second décrit le processus de mesure lui-même. Dans ce cadre de pensée, on

est à même de se poser la question de savoir comment le processus de mesure nous renseigne sur l'état d'un système physique qui nous est inconnu. Ainsi, la première question que l'on peut se poser est de savoir si grâce à une et une seule mesure nous pouvons discerner parfaitement deux états quantiques de notre système. Nous allons voir que cela n'est pas toujours possible.

Les mesures généralisées fournissent un cadre naturel pour considérer des états non orthogonaux (section 5.4) et il est donc naturel de poser cette question en ces termes. En notant M_j l'opérateur associé au résultat de la mesure j , la probabilité d'obtenir le résultat j est $p_j(\psi) = \langle M_j^\dagger M_j \rangle_{|\psi\rangle}$. Pour que deux états $|\psi_1\rangle$ et $|\psi_2\rangle$ soient discernables en un coup, il faut que les résultats d'une mesure généralisée soient spécifiques à chacun d'eux, c'est-à-dire que si un résultat sort pour un état, alors il ne sort jamais pour l'autre. On dira donc que deux états sont parfaitement *discernables* si et seulement s'il existe une mesure généralisée telle que

$$\forall j, p_j(\psi_1) p_j(\psi_2) = 0. \quad (6.1)$$

Cette définition requiert juste qu'il existe une mesure satisfaisant cette condition. Mais d'autres mesures peuvent ne pas permettre de faire la distinction en un seul coup. Pour que deux états ne soient pas parfaitement discernables, il faut et il suffit qu'aucune mesure ne permette de les discerner. Les questions qui se posent sont donc de savoir s'il est possible de discerner à coup sûr entre deux états. Si oui lesquels, si non pourquoi.

6.1.1 Indiscernabilité de deux états non orthogonaux

Le premier théorème que nous allons prouver ici stipule qu'il n'est pas possible avec une unique mesure de discerner à coup sûr deux états quantiques non orthogonaux. La preuve qui suit utilise un raisonnement par l'absurde.

Considérons deux états quantiques normés $|\psi_1\rangle$ et $|\psi_2\rangle$ fixés et supposons qu'il existe un processus de mesure généralisé qui permette de les discerner à coup sûr. Sans nuire à la généralité, on va supposer que, dans ce processus décrit par la donnée d'opérateurs M_j associés à différents résultats de mesure et tels que $\sum_j M_j^\dagger M_j = \mathbb{1}$, M_1 permet d'identifier l'état $|\psi_1\rangle$ à coup sûr, c'est-à-dire que l'on a

$$\langle \psi_1 | M_1^\dagger M_1 | \psi_1 \rangle = 1 \Leftrightarrow \forall j \neq 1, M_j |\psi_1\rangle = 0. \quad (6.2)$$

La probabilité de trouver 1 comme résultat de la mesure avec $|\psi_2\rangle$ est donc nulle : $p_1(\psi_2) = \langle \psi_2 | M_1^\dagger M_1 | \psi_2 \rangle = 0$. En toute généralité, nous pouvons décomposer l'état $|\psi_2\rangle$ en une partie colinéaire à $|\psi_1\rangle$ et une partie orthogonale

$$|\psi_2\rangle = \alpha |\psi_1\rangle + |\chi_2\rangle \text{ où } \langle \psi_1 | \chi_2 \rangle = 0. \quad (6.3)$$

Calculons maintenant la probabilité d'obtenir un résultat $j \neq 1$ avec l'état $|\psi_2\rangle$,

$$p_{j \neq 1}(\psi_2) = \langle \psi_2 | M_j^\dagger M_j | \psi_2 \rangle = \langle \chi_2 | M_j^\dagger M_j | \chi_2 \rangle. \quad (6.4)$$

Mais nous savons que l'ensemble des opérateurs $M_j^\dagger M_j$ somme à l'identité, ce qui entraîne que

$$\sum_{j \neq 1} p_j(\psi_2) = \langle \chi_2 | \chi_2 \rangle - \langle \chi_2 | M_1^\dagger M_1 | \chi_2 \rangle \leq \langle \chi_2 | \chi_2 \rangle. \quad (6.5)$$

De plus, nous avons $|\langle \psi_2 | \psi_2 \rangle|^2 = |\alpha|^2 + \langle \chi_2 | \chi_2 \rangle = 1$. Ainsi, si $\alpha \neq 1$, la norme de l'état $|\chi_2\rangle$ est strictement inférieure à l'unité. L'équation (6.5) montre alors qu'il n'existe pas de mesure généralisée M_j telle que $\sum_{j \neq 1} p_j(\psi_2) = 1$. On conclut donc que, dès que l'état $|\psi_2\rangle$ est non orthogonal à l'état $|\psi_1\rangle$, les résultats $j \neq 1$ ne permettent pas de le discerner de l'état $|\psi_1\rangle$ à coup sûr. Ceci montre le théorème d'indiscernabilité.

6.1.2 Ce que l'indiscernabilité ne dit pas

Le résultat que l'on vient de montrer est clair : il n'est pas possible en une seule mesure de discerner parfaitement deux états quantiques non orthogonaux. Chacun des termes de ce théorème est important.

Premièrement, si les états sont orthogonaux, alors il est tout à fait possible de les discerner à tous les coups avec un processus de mesure unique. Il suffit en effet d'avoir l'appareil de mesure défini par $M_1 = |\psi_1\rangle\langle\psi_1|$, $M_2 = |\psi_2\rangle\langle\psi_2|$, les autres étant définis grâce à des états qui complètent $|\psi_1\rangle$ et $|\psi_2\rangle$ en une base orthonormée.

Ensuite, si l'on autorise notre appareil de mesure à répondre « je ne sais pas » ou même si l'on s'autorise un taux d'erreur non nul dans les réponses, le problème est différent et le théorème ne s'applique pas. Dans ces situations, il est possible de quantifier notre capacité à les discerner et de penser à des protocoles optimums. Lorsque deux états ne sont pas discernables parfaitement, les limites quantitatives, par exemple une borne sur la probabilité d'arriver à les discerner, fait l'objet de la dernière section de ce chapitre.

Enfin, nous devons noter que dans un monde classique, il est toujours possible de discriminer parfaitement entre deux états. Ce théorème d'impossibilité est donc notre première illustration de la différence entre l'information contenue dans un état classique *versus* un état quantique.

6.2 Théorème de non-clonage

Si la non-indiscernabilité à coup sûr différencie une théorie classique d'une théorie quantique, la démonstration que nous en avons donnée ne permet pas d'en comprendre simplement les raisons physiques. Nous allons voir que ce résultat est en fait relié à la question de clonage d'un état quantique inconnu.

Une machine à cloner est une machine qui transforme un état quantique $|\phi\rangle$ et un état blanc $|B\rangle$ en deux copies de l'état $|\phi\rangle$ d'une manière compatible avec le principe de superposition. La figure 6.1 à gauche représente schématiquement

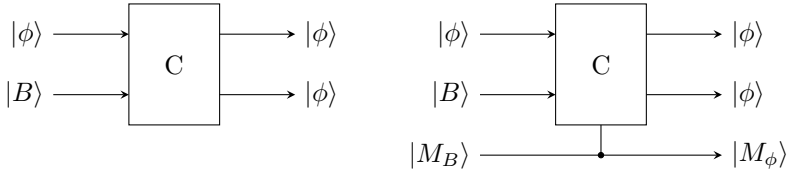


FIGURE 6.1 – Machines à cloner. À gauche : une machine à cloner idéale, sans état interne. À droite : une machine à cloner non idéale, possédant un état interne. Ce dernier ne doit pas être intriqué avec les états de sortie.

une telle machine. Notons qu'il s'agit là d'une machine idéale, classique, dont on ne précise pas l'état physique.

Dans une description plus réaliste, une machine à cloner est elle-même un système quantique dont l'état est altéré par le processus de clonage : on passe de l'état initial $|M_B\rangle$ à l'état $|M_\phi\rangle$ après le clonage. On suppose qu'il n'y a pas d'intrication entre les états de la machine et les états de « sortie » car dans le cas contraire cela voudrait dire qu'on ne pourrait pas oublier l'état de la machine et réaliser le clonage.

6.2.1 Le théorème de non-clonage

Machine idéale (ou classique)

Prenons le cas d'une machine à cloner idéale, c'est-à-dire sans état interne ou classique. Prenons deux états $|0\rangle$ et $|1\rangle$ orthogonaux, et notons $|B\rangle$ l'état blanc. Se restreindre à un qubit comme système à cloner ne fait pas perdre en généralité le contenu du théorème. Le fonctionnement quantique de la machine à cloner sera décrit par un opérateur C tel que

$$C(|0\rangle \otimes |B\rangle) = |0\rangle \otimes |0\rangle \equiv |00\rangle, \quad (6.6a)$$

$$C(|1\rangle \otimes |B\rangle) = |1\rangle \otimes |1\rangle \equiv |11\rangle. \quad (6.6b)$$

Un des prérequis du fonctionnement de la machine est qu'elle respecte le principe de superposition. L'opérateur C doit être un opérateur linéaire et unitaire. Le fonctionnement de la machine peut être écrit de deux façons différentes. Le première utilise la linéarité

$$\begin{aligned} C((\alpha|0\rangle + \beta|1\rangle) \otimes |B\rangle) &= \alpha C(|0\rangle \otimes |B\rangle) + \beta C(|1\rangle \otimes |B\rangle) \\ &= \alpha |00\rangle + \beta |11\rangle. \end{aligned} \quad (6.7)$$

La seconde utilise directement la définition même de la machine

$$\begin{aligned} C((\alpha|0\rangle + \beta|1\rangle) \otimes |B\rangle) &= (\alpha|0\rangle + \beta|1\rangle) \otimes (\alpha|0\rangle + \beta|1\rangle) \\ &= \alpha^2 |00\rangle + \alpha\beta |01\rangle + \beta\alpha |10\rangle + \beta^2 |11\rangle. \end{aligned} \quad (6.8)$$

Les deux états sont clairement différents comme on peut le voir sur le comportement quadratique ou non des amplitudes. D'un point de vue plus physique, le premier est un état intriqué tandis que le second est factorisé. Cette incompatibilité avec la linéarité prouve que l'opération de clonage d'un état quantique inconnu n'est pas possible dans le cadre de la théorie quantique. C'est le *théorème de non-clonage*.

Le même résultat peut être prouvé en disant que la machine correspond à une évolution physique qui doit être unitaire. En clonant deux états $C(|\psi_1\rangle \otimes |B\rangle) = |\psi_1\rangle \otimes |\psi_1\rangle$ et $C(|\psi_2\rangle \otimes |B\rangle) = |\psi_2\rangle \otimes |\psi_2\rangle$, l'unitarité nous dit alors que $(\langle\psi_1| \otimes \langle\psi_1|)(|\psi_2\rangle \otimes |\psi_2\rangle) = \langle\psi_1|\psi_2\rangle^2$ devrait être égal à $(\langle\psi_1| \otimes \langle 0|)(|\psi_2\rangle \otimes |0\rangle) = \langle\psi_1|\psi_2\rangle\langle 0|0\rangle = \langle\psi_1|\psi_2\rangle$. Or cela n'est possible que si les deux états sont orthogonaux ou bien égaux (états de norme un). Dans le cas général, on conclut de nouveau que le clonage n'est pas possible.

Machine non idéale

Considérons le cas des machines non idéales, c'est-à-dire où l'on prend en compte l'état de la machine elle-même. Nous avons alors :

$$C(|0\rangle \otimes |B\rangle \otimes |M_B\rangle) = |0\rangle \otimes |0\rangle \otimes |M_0\rangle, \quad (6.9a)$$

$$C(|1\rangle \otimes |B\rangle \otimes |M_B\rangle) = |1\rangle \otimes |1\rangle \otimes |M_1\rangle. \quad (6.9b)$$

Prenons alors $|X\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, on devrait avoir par linéarité, ainsi que par définition de C ¹

$$C\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |B\rangle \otimes |M_B\rangle\right) = \frac{|0\rangle^{\otimes 2} \otimes |M_0\rangle + |1\rangle^{\otimes 2} \otimes |M_1\rangle}{\sqrt{2}} \quad (6.10a)$$

$$\text{et } C\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |B\rangle \otimes |M_B\rangle\right) = \frac{(|0\rangle + |1\rangle)^{\otimes 2}}{2} \otimes |M_X\rangle. \quad (6.10b)$$

Mais là encore, on voit que le second état combine des termes $|0\rangle \otimes |1\rangle$ et $|1\rangle \otimes |0\rangle$ même après avoir ignoré l'état de la machine alors que dans le premier état, la probabilité d'observer deux valeurs différentes pour σ^z sur chaque qubit est zéro. Ainsi, la machine à cloner non idéale ne fonctionne pas non plus.

6.2.2 Ce que le non-clonage ne dit pas

Notons ici quelques points importants. D'abord, le clonage d'une information classique est faisable : il existe des photocopieurs qui réalisent cette tâche. Mais, s'il n'existe pas de « photocopieur quantique » parfait, il est toujours possible de réaliser un clonage imparfait.

1. La notation $|x\rangle^{\otimes 2}$ signifie $|x\rangle \otimes |x\rangle$.

Ensuite, remarquons que le théorème de non-clonage n'est pas en contradiction avec l'existence d'une tomographie quantique. Certes, il est impossible de caractériser parfaitement un état quantique à partir d'un nombre fini de systèmes préparés dans cet état (chapitre 4). La tomographie est donc un processus asymptotique réalisé à la limite d'un nombre infini d'expériences. On ne peut pas réaliser une copie d'un état quantique dont on ne dispose que d'une seule réalisation.

Enfin, le non-clonage n'est pas en contradiction avec la possibilité de préparer plusieurs réalisations d'un état quantique donné. En fait, un état quantique est la manière dont la nature encode l'information sur un processus physique de préparation. C'est une manière de comprendre la discussion des états intriqués en termes d'états relatifs du chapitre 5. Pour citer Michel Le Bellac, un état quantique inconnu est un oxymore : si vous recevez un système physique dans un état quantique que vous ne connaissez pas, c'est que quelqu'un l'a préparé et donc sait quel est cet état. Et cette personne peut, en principe, en préparer plusieurs.

6.2.3 Non-clonage et indiscernabilité

Commençons par discuter la relation entre le théorème de non-clonage et l'indiscernabilité de deux états quantiques. On va démontrer l'assertion suivante : s'il existe une machine à cloner, alors on peut fabriquer une machine à discerner deux états non orthogonaux à coup sûr.

Imaginons que l'on dispose d'une machine à cloner. On peut à partir d'une telle machine en fabriquer une autre à cloner N fois

$$C_N(|\psi\rangle \otimes |B\rangle^{\otimes N} \otimes |M_B^N\rangle) = |\psi\rangle^{\otimes(N+1)} \otimes |M_\psi^N\rangle. \quad (6.11)$$

Prenons alors $|\psi_1\rangle$ et $|\psi_2\rangle$ non orthogonaux. On clone alors chacun de ces états N fois à l'aide de la machine dont on dispose. Soit $|\psi_\alpha\rangle$ un état inconnu, qui peut être uniquement l'un des deux états $|\psi_1\rangle$ ou $|\psi_2\rangle$. On va le cloner N fois et utiliser ensuite un processus de mesure simultanée sur $N + 1$ réalisations. Selon le choix de mesure que l'on fait, on va avoir les probabilités suivantes :

$$p_1(\psi_\alpha) = |\langle\psi_\alpha|\psi_1\rangle|^{2(N+1)}, \quad (6.12a)$$

$$p_2(\psi_\alpha) = |\langle\psi_\alpha|\psi_2\rangle|^{2(N+1)}. \quad (6.12b)$$

À la limite où $N \rightarrow \infty$, comme $|\langle\psi_1|\psi_2\rangle| < 1$, on a :

$$\begin{aligned} p_1(\psi_1) &= 1 & p_1(\psi_2) &= 0, \\ p_2(\psi_1) &= 0 & p_2(\psi_2) &= 1. \end{aligned} \quad (6.13)$$

En conséquence, on sait discerner les deux états non orthogonaux à coup sûr. L'idée est que le clonage permet de fabriquer des états d'un système plus grand dont le recouvrement diminue. Le clonage permet en quelque sorte d'amplifier la discernabilité d'une paire d'états non orthogonaux.

L'autre résultat qu'il reste à démontrer est la réciproque : si on sait discerner à coup sûr deux états non orthogonaux, alors on sait effectuer le clonage parfait d'un état quantique arbitraire. Pour cela, préparons de manière aléatoire des états dans l'espace des états à un qubit puis, en utilisant la discernabilité d'états non orthogonaux, testons si on a effectué un clonage ou pas. L'espace des états physiques à un qubit étant la sphère de Bloch, le processus est convergent : à la limite d'un nombre d'itérations infini, on finit par arriver sur l'état $|\psi\rangle \otimes |\psi\rangle$. Distinguer parfaitement permet donc de cloner un état quantique, au moins de manière asymptotique.

Nous venons donc de montrer que les théorèmes de non-clonage et d'indiscernabilité sont tous les deux équivalents. Ils mettent en exergue le caractère différent de l'information quantique comparée au classique. Le principe de superposition (linéarité) est au cœur de cette différence. Est-ce que cela caractérise complètement la théorie quantique ? Malheureusement non. Il existe de nombreuses autres théories différentes de la théorie quantique qui possèdent ces mêmes propriétés. De façon plus prudente, il semble que ces résultats permettent *a minima* de dire que la théorie quantique est une théorie non classique, aspect que l'on reverra au travers des inégalités de Bell au chapitre 11.

6.3 Théorème de non-effacement

Le problème du clonage amène naturellement à se poser la question opposée : est-t-il possible d'effacer, c'est-à-dire de « remettre à zéro » l'état inconnu d'un système ? La théorie quantique répond de nouveau par la négative à cette question. Dans le cas de l'effacement, il faut être d'autant plus vigilant sur le sens des mots que l'on utilise. Les subtilités sur la notion d'effacement seront discutées après la démonstration du théorème.

6.3.1 Le théorème de non-effacement

Machine à une entrée

La situation la plus intuitive à considérer est la suivante : peut-on effacer un unique état, c'est-à-dire fabriquer une machine telle que

$$\forall |\psi\rangle, E(|\psi\rangle \otimes |M_I\rangle) = |B\rangle \otimes |M_\psi\rangle \quad (6.14)$$

sans que l'état interne $|M_\psi\rangle$ de la machine quantique véhicule les mêmes informations que $|\psi\rangle$? Cela pose problème au niveau de l'unitarité. Considérons les produits scalaires

$$(\langle B| \otimes \langle M_{\psi_2}|) (|B\rangle \otimes |M_{\psi_1}\rangle) = \langle M_{\psi_2} | M_{\psi_1} \rangle, \quad (6.15)$$

$$(\langle \psi_2| \otimes \langle M_I|) (|\psi_1\rangle \otimes |M_I\rangle) = \langle \psi_2 | \psi_1 \rangle. \quad (6.16)$$

Si pour tout couple de vecteurs, les membres de droite de ces deux équations sont égaux, cela veut dire qu'il existe une transformation unitaire qui envoie

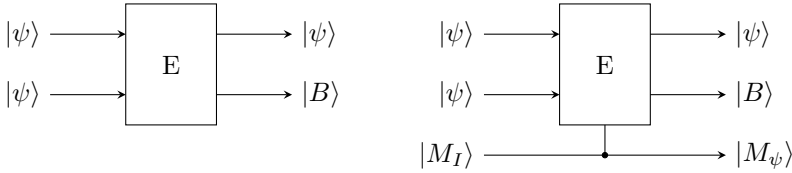


FIGURE 6.2 – Machines à effacer. À gauche : une machine à effacer idéale, sans état interne. À droite : une machine à effacer réelle, possédant un état interne.

$|\psi\rangle$ sur $|M_\psi\rangle$, donc qu'on a transféré l'état $|\psi\rangle$ sur l'état interne de la machine. Ce dernier contient alors les mêmes informations que l'état $|\psi\rangle$, ce que l'on cherchait à éviter. Ceci montre que si on part de $|0\rangle$ et $|1\rangle$ deux états orthogonaux et que l'on cherche à les initialiser à zéro au moyen d'une machine quantique, alors on a juste envoyé l'état $|M_I\rangle$ de la machine sur deux états orthogonaux $|M_0\rangle$ et $|M_1\rangle$. Ce résultat est une première forme du *théorème de non-effacement*.

Machine à deux entrées

Le type de machine à effacer que l'on considère pour ce théorème correspond à l'analogie quantique de l'opération **cnot** qui est une façon de réaliser la remise à zéro. Imaginons que l'on prépare deux réalisations d'un système physique dans le même état inconnu. La machine à effacer initialise un des deux systèmes dans un état fixé et cela quel que soit l'état en entrée (voir figure 6.2). Là aussi, il s'agit d'une machine idéale, classique, et une véritable machine à effacer possède son propre état interne.

Considérons d'abord le cas d'une machine idéale sans état interne. Formellement, son principe de fonctionnement est le suivant

$$\forall |\psi\rangle, E(|\psi\rangle \otimes |\psi\rangle) = |\psi\rangle \otimes |B\rangle. \quad (6.17)$$

L'opérateur E étant inversible, si une telle machine existe, alors son inverse réalise l'opération de clonage de manière idéale et donc ne peut exister. Nous venons de montrer une forme du *théorème de non-effacement*.

Cependant, il se pourrait qu'une machine à effacer non idéale, c'est-à-dire ayant son propre état quantique interne, existe. Notant de nouveau son état quantique interne $|M\rangle$, on se demande s'il existe un opérateur unitaire E vérifiant :

$$\forall |\psi\rangle, E(|\psi\rangle \otimes |\psi\rangle \otimes |M_I\rangle) = |\psi\rangle \otimes |B\rangle \otimes |M_\psi\rangle. \quad (6.18)$$

Remarquons que, dans ce cas, l'inverse n'est pas une machine à cloner non idéale car l'opération de clonage n'est réalisée qu'en mettant en entrée de l'inverse un état auxiliaire $|M_\psi\rangle$ qui dépend de ψ . Pour les besoins de la

démonstration, on choisit comme système un qubit. Posons $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. La machine à effacer, qui est linéaire, doit nécessairement vérifier les relations suivantes :

$$E\left((\alpha|0\rangle + \beta|1\rangle)^{\otimes 2} \otimes |M_I\rangle\right) = (\alpha|0\rangle + \beta|1\rangle) \otimes |B\rangle \otimes |M_\psi\rangle, \quad (6.19a)$$

$$\begin{aligned} E\left((\alpha|0\rangle + \beta|1\rangle)^{\otimes 2} \otimes |M_I\rangle\right) &= \alpha^2|0\rangle \otimes |B\rangle \otimes |M_0\rangle + \beta^2|1\rangle \otimes |B\rangle \otimes |M_1\rangle \\ &\quad + \alpha\beta E\left((|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle) \otimes |M_I\rangle\right). \end{aligned} \quad (6.19b)$$

Pour que ces deux états soient identiques, il faut vérifier les relations suivantes :

$$|M_\psi\rangle = \alpha|M_0\rangle + \beta|M_1\rangle, \quad (6.20a)$$

$$\begin{aligned} E(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle) \otimes |M_I\rangle &= |0\rangle \otimes |B\rangle \otimes |M_0\rangle \\ &\quad + |1\rangle \otimes |B\rangle \otimes |M_1\rangle. \end{aligned} \quad (6.20b)$$

De plus, l'état quantique en sortie de la machine doit être de norme 1, ce qui signifie que

$$\langle B|B\rangle \times \|\alpha|0\rangle + \beta|1\rangle\|^2 \times \|\alpha|M_0\rangle + \beta|M_1\rangle\|^2 = 1. \quad (6.21)$$

En développant cette expression, nous arrivons à $\langle M_0|M_1\rangle = 0$. L'état interne de la machine après la procédure d'effacement, $|M_\psi\rangle$, est donc un état quantique complètement équivalent à l'état $|\psi\rangle$ dont on avait reçu deux copies en entrée. Ainsi, notre machine à effacer ne fait que transférer l'information contenue dans l'état $|\psi\rangle$ vers son état interne².

Ce qui est important à retenir, c'est que l'effacement d'un état quantique ne peut se faire qu'en transférant l'état quantique inconnu dans l'environnement. En revanche, le théorème de non-effacement signifie que si on interdit ce transfert, il n'est pas possible de supprimer une copie d'un état quantique inconnu. Ces aspects seront transparents et auront des conséquences physiques majeures lors de notre discussion sur l'émergence du monde classique au chapitre 14, volume II.

6.3.2 Le problème de l'effacement

Pour bien comprendre les résultats sur l'effacement, commençons par discuter ces questions dans le contexte classique. Nous distinguons deux opérations en apparence très proches : la mise à zéro d'une mémoire et l'effacement d'un bit.

La mise à zéro consiste à envoyer un bit sur une valeur de référence, par exemple 0, et ce quelle que soit la valeur initiale du bit. C'est ce que l'on fait

2. Bien entendu, pour un état d'entrée qui n'est pas de la forme $|\psi\rangle \otimes |\psi\rangle$, la machine à effacer fonctionne d'une manière plus compliquée, qui doit vérifier l'équation (6.20b).

lorsque l'on reformate un disque dur. Un point absolument crucial remarqué par Landauer [246] est qu'une telle opération ne peut être réalisée sans dissipation. En effet, si elle l'était, on pourrait la réaliser par une dynamique réversible vu la micro-réversibilité des équations de la mécanique. Or ce n'est pas le cas vu qu'elle envoie les deux états possibles d'un bit sur un seul. L'opération de remise à zéro est donc logiquement et physiquement irréversible. Elle nécessite une certaine dissipation d'énergie dans l'environnement, ce que nous étudierons de manière approfondie au chapitre 12, volume II.

La seconde opération consiste à envoyer deux bits ayant la même valeur (x, x) sur le bit $(x, 0)$. On a donc effacé un des deux bits. En classique, l'effacement est facile à réaliser au moyen de l'opération **cnot** qui envoie (x, y) sur $(x, x + y)$ et donc (x, x) sur $(x, 0)$. L'opération **cnot** étant inversible, cela montre que l'effacement est réalisable en classique : c'est une opération logiquement et physiquement réversible. Notons que **cnot** qui est son propre inverse fournit également l'opération de clonage.

Nous voyons donc déjà que, même au niveau classique, les deux opérations ne sont pas identiques : effacement et remise à zéro sont fondamentalement différents.

Passons maintenant au cas quantique. L'argument utilisé dans le cas classique montre que la remise à zéro d'une mémoire quantique ne peut se faire sans environnement et sans dissipation. L'impossibilité d'une mise à zéro sans dissipation est donc vraie aussi bien dans un monde classique que dans un monde quantique et nous verrons que la quantité d'énergie à dissiper pour effacer un bit d'information est la même dans les deux mondes.

Cependant, dans le cas d'une machine classique, l'effacement pour deux bits serait l'inverse du clonage. En conséquence, classiquement clonage et effacement sont tous les deux possibles. Mais dans le cas d'une machine possédant des degrés de liberté quantiques, le seul effacement possible consiste à envoyer l'état quantique que l'on souhaite effacer dans les degrés de liberté de la machine. Somme toute, ces considérations montrent que la différence entre le problème de l'effacement dans le cas classique et dans le cas quantique réside en fait dans l'existence de l'opération **cnot** classique qui réalise à la fois le clonage et l'effacement. Dans le monde quantique, l'effacement n'est pas réalisable car le clonage ne l'est pas.

Nous voyons donc que la différence essentielle entre information classique et information quantique réside dans le théorème de non-clonage.

6.4 Impossibilités non locales

Les théorèmes d'impossibilité que nous avons considérés sont locaux : ils énoncent l'impossibilité pour un opérateur de disposer d'une machine physique réalisant certaines opérations sur les états quantiques. Cependant, la perspective de la communication au moyen de ressources quantiques incite à chercher s'il n'existe pas des contraintes sur les opérations auxquelles ont accès plusieurs

opérateurs. Nous allons présenter ici plusieurs résultats dans ce sens dont la plupart sont des conséquences du théorème de non-clonage. Ils expriment le fait qu'il est impossible de transférer l'information quantique associée à un état quantique $|\psi\rangle$ dans une chaîne de bits classiques (inconvertibilité) ou même dans les corrélations entre deux systèmes quantiques (non-dissimulation).

Le seul résultat qui ne découle pas du non-clonage est le théorème de non-communication qui stipule qu'il n'est pas possible de communiquer de l'information au moyen de la seule intrication. Comme nous le verrons dans le chapitre 13, volume II, ce résultat implique le non-clonage mais n'en est pas une conséquence.

6.4.1 Le théorème de non-communication

Dans cette section, nous allons montrer qu'il n'est pas possible d'utiliser la seule intrication quantique pour communiquer de l'information. De manière plus précise, nous allons montrer que les probabilités de mesures locales effectuées par Alice ne sont pas influencées par les actions de Bob. On dit que la théorie quantique est une théorie « no-signaling ».

Pour le démontrer, considérons deux expérimentateurs Alice A et Bob B faisant chacun de leur côté des opérations de mesures locales sur le sous-système qui globalement peut être dans un état quelconque décrit par la matrice densité ρ_{AB} . Une telle matrice peut toujours être décomposée sous la forme $\rho_{AB} = \sum_k p_k \sigma_k^A \otimes \sigma_k^B$ où les σ_k^A et les σ_k^B sont des opérateurs densités pour Alice et Bob et où les p_k sont des probabilités. Supposons que Bob effectue des mesures généralisées décrites par des opérateurs $M_{B\mu}$ tels que $\sum_\mu M_{B\mu} M_{B\mu}^\dagger = \mathbb{1}$. Après une de ces mesures, l'état global a la forme $(\mathbb{1}_A \otimes M_{B\mu}^\dagger) \rho_{AB} (\mathbb{1}_A \otimes M_{B\mu})$. L'opérateur densité d'Alice, qui décrit tous les résultats de mesures locales qu'elle puisse faire sur son sous-système, est alors obtenu en traçant $(\mathbb{1}_A \otimes M_{B\mu}^\dagger) \rho_{AB} (\mathbb{1}_A \otimes M_{B\mu})$ sur l'espace $\mathcal{H}_B$.

Considérons maintenant qu'Alice effectue une mesure généralisée décrite par l'opérateur M_A . La probabilité p_A qu'Alice obtienne le résultat associé à cette mesure est alors donnée par

$$\begin{aligned}
 p_A &= \text{tr}_A \left(\sum_\mu \text{tr}_B \left((M_A^\dagger \otimes M_{B\mu}^\dagger) \rho_{AB} (M_A \otimes M_{B\mu}) \right) \right) \\
 &= \text{tr}_A \left(\sum_{k\mu} p_k \text{tr}_B \left((M_A^\dagger \otimes M_{B\mu}^\dagger) \sigma_k^A \otimes \sigma_k^B (M_A \otimes M_{B\mu}) \right) \right) \\
 &= \text{tr}_A \left(\sum_{k\mu} p_k \text{tr}_B \left(\sigma_k^A \sigma_k^B M_A M_{B\mu} M_A^\dagger M_{B\mu}^\dagger \right) \right). \tag{6.22}
 \end{aligned}$$

Maintenant, une condition suffisante assurant l'indépendance des mesures d'Alice et de Bob est que $[M_A, M_{B\mu}] = 0$ ³. Remarquons que l'on suppose que ce que fait Alice n'est pas conditionné au résultat de mesures de Bob ce qui, même dans un cadre classique, changerait les probabilités d'Alice. En utilisant $[M_A, M_{B\mu}] = 0$, le calcul ci-dessus peut être poursuivi et donne

$$\begin{aligned} p_A &= \text{tr}_A \left(\sum_k p_k \text{tr}_B \left(M_A^\dagger \sigma_k^A \sigma_k^B M_A \left(\sum_\mu M_{B\mu} M_{B\mu}^\dagger \right) \right) \right) \\ &= \text{tr}_A \left(M_A^\dagger \text{tr}_B \sum_k p_k (\sigma_k^A \sigma_k^B) M_A \right) = \text{tr}_A \left(M_A^\dagger \rho_A M_A \right), \end{aligned} \quad (6.23)$$

où $\rho_A = \sum_k p_k \sigma_k^A$ est l'opérateur densité d'Alice avant toute mesure effectuée par Bob.

Ce raisonnement montre que, quelque soit l'opération locale de mesure que puisse faire Bob sur sa partie du système, cela n'affectera en rien la statistique associée aux observables locales d'Alice. L'intrication seule ne permet donc pas de communiquer une information entre Alice et Bob par des mesures locales sur leur sous-système respectif. Ceci est essentiellement dû au caractère aléatoire de la mesure en théorie quantique. C'est le *théorème de non-communication* de la théorie quantique.

6.4.2 Le théorème d'inconvertibilité

Le théorème d'inconvertibilité stipule qu'il est impossible de mettre au point une machine qui encode un état quantique arbitraire d'un qubit dans une chaîne finie de bits classiques.

Si une telle machine existait, elle permettrait de violer le théorème de non-clonage. En effet, s'il était possible d'encoder l'état d'un qubit dans une information classique, celle-ci pouvant être dupliquée, il serait ensuite possible de régénérer deux copies de cet état inconnu en utilisant deux fois le décodeur.

Un autre argument consiste à dire qu'une telle machine violerait la borne de Massard et Popescu discutée à la section 4.3.3. Pour cela, considérons une source de $N > 1$ réalisations identiques d'un même état quantique $|\psi\rangle$. Passons une de ces copies en entrée de l'encodeur en question : on obtient ainsi une chaîne de bits classiques que l'on peut ensuite envoyer en entrée du décodeur. Celui-ci fournit alors l'état $|\psi\rangle$ en sortie que nous prenons comme estimateur de l'état quantique inconnu de départ. La fidélité de l'estimateur par rapport à l'état de départ est toujours égale à l'unité. On aurait donc construit une machine physique qui donnerait un score d'estimation égal à l'unité dès $N = 1$. C'est une violation de la borne de Massard et Popescu et une telle machine ne peut donc exister.

3. De telles opérations sont naturelles dans un cadre relativiste pour des observateurs séparés par un intervalle de genre espace (chapitre 13, volume II).

Finalement, le théorème d'inconvertibilité de l'information quantique en information classique est une conséquence directe du théorème de non-clonage quantique.

6.4.3 Le théorème de non-dissimulation

Nous avons vu qu'il est impossible de cloner l'information quantique, et qu'il est également impossible de l'effacer. Ces éléments plaident en faveur d'une sorte de conservation de l'information quantique qui ne peut être que déplacée d'un système à un autre. Pour confirmer cette intuition, nous allons examiner ce qui se passe lorsque l'on part d'un système physique initialement préparé dans un état quelconque $|\psi\rangle \in \mathcal{H}_S$ et que l'on effectue une série d'actions sur le système et son environnement qui aboutissent, dans tous les cas, à effacer toute l'information sur l'état $|\psi\rangle$. Bien sûr, cela peut se faire en échangeant l'état quantique dans l'environnement, mais ne serait-il pas possible de dissimuler l'information sur l'état quantique dans des corrélations entre le système et son environnement sans qu'elle ne soit accessible en considérant le système ou l'environnement séparément ?

Tout d'abord, il convient de remarquer que c'est possible en classique. Considérons un dispositif qui prend deux chaînes binaires de taille n , l'une étant appelée le message x et l'autre la clé c et qui effectue l'opération binaire **xor** bit à bit notée $\oplus : (x, c) \rightarrow x \oplus c$. Cette opération, qui vit dans $(\mathbb{Z}/2\mathbb{Z})^n$, vérifie les propriétés suivantes : la commutativité $a \oplus b = b \oplus a$, l'associativité $a \oplus (b \oplus c) = (a \oplus b) \oplus c$, l'idempotence $a \oplus a = 0$ et possède naturellement un élément neutre $a \oplus 0 = a$. Cette opération n'est cependant pas inversible : le résultat $y = x \oplus c$ ne permet pas de retrouver ni x ni c sauf à avoir accès à la clé c . L'information sur le message x est cachée dans les corrélations entre c et $x \oplus c$. C'est le principe de chiffrement de Vernam sur lequel nous reviendrons dans le prochain chapitre.

La question quantique analogue consiste à se demander s'il est possible que Alice et Bob disposent d'un dispositif physique qui permette à Alice d'encoder un état quantique inconnu $|\psi\rangle$ dans un état $|\Psi(\psi)\rangle$ partagé par Alice et Bob de telle sorte que ni l'opérateur densité réduit pour Alice, ni celui pour Bob ne dépendent de $|\psi\rangle$. Nous allons démontrer que cela n'est pas possible : plus précisément, si l'opérateur densité d'Alice ne dépend plus du tout de l'état $|\psi\rangle$, alors Bob dispose de l'état $|\psi\rangle$. C'est le *théorème de non-dissimulation* [67] qui stipule qu'il est impossible de dissimuler un état quantique inconnu dans des corrélations entre deux sous-systèmes.

Pour le démontrer, nous allons considérer le cas idéal où le processus physique de dissimulation permet de passer de manière linéaire et unitaire de l'état $|\psi\rangle \in \mathcal{H}_S$ à un état $|\Psi(\psi)\rangle \in \mathcal{H}_S \otimes \mathcal{H}_E$ tel que l'opérateur densité réduit $\rho_S(\psi) = \text{tr}_{\mathcal{H}_E}(|\Psi(\psi)\rangle\langle\Psi(\psi)|)$ soit indépendant de l'état $|\psi\rangle$. La linéarité et l'unitarité de cette transformation sont une conséquence directe du théorème de Gisin, Hughston, Jozsa et Wootters discuté à la section 5.3.2 : on peut

toujours étendre suffisamment l'environnement pour que ce soit le cas. Nous pouvons alors écrire une décomposition de Schmidt sous la forme

$$|\Psi(\psi)\rangle = \sum_k \sqrt{p_k} |s_k\rangle |E_k(\psi)\rangle, \quad (6.24)$$

où les $p_k \neq 0$ définissent une distribution de probabilité, les états $|s_k\rangle$ forment une base orthonormée de $\mathcal{H}_S$ et les états $|E_k(\psi)\rangle$ de $\mathcal{H}_E$ pour chaque $|\psi\rangle$ dans $\mathcal{H}_S$. La linéarité de la transformation $|\psi\rangle \mapsto |\Psi(\psi)\rangle$ entraîne que $|\psi\rangle \mapsto |E_k(\psi)\rangle$ est linéaire. De même, considérons le produit scalaire

$$\langle \Psi(\psi_-) | \Psi(\psi_+) \rangle = \sum_k p_k \langle E_k(\psi_-) | E_k(\psi_+) \rangle, \quad (6.25)$$

où on a utilisé l'orthogonalité de la famille $(|s_k\rangle)_k$. Pour que deux états orthogonaux soient envoyés sur deux états orthogonaux, il faut que, pour $|\psi_+\rangle$ et $|\psi_-\rangle$ orthogonaux, $\langle E_k(\psi_-) | E_k(\psi_+) \rangle = 0$ pour tout k car les p_k sont des réels strictement positifs. En utilisant une base orthonormée $|\psi_j\rangle$ de $\mathcal{H}_S$, et en notant $|E_{kj}\rangle = |E(\psi_j)\rangle_k$, nous avons une famille orthonormée de vecteurs $|E_{kj}\rangle$. Il est alors toujours possible de la voir comme une base formée de vecteurs qui sont des produits tensoriels $|e_k\rangle \otimes |f_j\rangle$ avec $|e_k\rangle$ et $|f_j\rangle$ deux familles orthonormées définissant dans $\mathcal{H}$ un sous-espace isomorphe à un produit tensoriel. Du coup, nous pouvons écrire

$$|\Psi(\psi)\rangle = \sum_{k,j} \sqrt{p_k} \langle f_j | \psi \rangle |s_k\rangle |e_k\rangle |f_j\rangle \quad (6.26a)$$

$$= \sum_k \sqrt{p_k} |s_k\rangle |e_k\rangle \otimes \left(\sum_j \langle f_j | \psi \rangle |f_j\rangle \right). \quad (6.26b)$$

La dernière ligne montre donc que la transformation a encodé l'état $|\psi\rangle$ dans l'espace des états $\mathcal{H}_E$. Ceci prouve le théorème de non-dissimulation.

À la lumière de ce résultat et du théorème de non-effacement discuté à la section 6.3, nous voyons donc qu'il n'est pas possible d'effacer ni de dissimuler une information de nature quantique. Tout au plus pouvons-nous la transférer dans une autre partie du système ou dans l'environnement de ce dernier. Il n'est pas non plus possible de cloner de l'information quantique. Ces résultats expriment sous une forme qualitative la notion de permanence de l'information quantique.

Plus généralement, les théorèmes d'impossibilité que nous avons explorés tout au long de ce chapitre expriment des contraintes sur ce qu'il est possible de faire avec deux types de ressources en apparence très différentes mais qui coexistent dans un univers quantique. Les lois quantitatives qui régissent les différentes opérations et conversions impliquant information classique et quantique forment la théorie de l'information quantique. Mais avant d'entrer dans cette discussion, nous allons explorer les conséquences opérationnelles des théorèmes d'impossibilités.

6.5 Discernabilité d'états non orthogonaux

Le théorème d'indiscernabilité stipule qu'il est impossible de discerner avec certitude deux états quantiques non orthogonaux. Cependant, il est possible de quantifier dans quelle mesure, c'est-à-dire avec quelle probabilité, une telle distinction est possible. Nous allons donc ici discuter les limites quantitatives sur la probabilité de distinction sans erreur d'états quantiques non orthogonaux. Celles-ci jouent un rôle important non seulement dans la sécurisation des communications quantiques mais aussi dans les télécommunications classiques.

6.5.1 Discrimination sans erreur

Nous allons tout d'abord considérer des protocoles qui reposent sur une détection à coup sûr des états préparés par la source. De tels protocoles sont dits « sans erreurs » car ils fournissent une réponse qui est soit correcte, soit non conclusive, « je ne sais pas ». Leurs performances sont donc mesurées par le taux de réponse conclusives.

Dans un premier temps, nous allons chercher un tel protocole basé sur une mesure projective. Puis nous l'élargirons en un protocole utilisant une mesure généralisée et nous verrons que ce dernier a une probabilité plus élevée de discerner entre deux états non orthogonaux. Ce résultat est une des multiples illustrations du fait que, en physique quantique, procéder à des mesures sur un système élargi peut permettre un gain de performance. Au contraire, dans un cadre classique, élargir le système sur lequel on fait des mesures introduit en général du bruit supplémentaire et dégrade la performance.

Discrimination par mesures projectives

Considérons une source qui ne prépare que deux états purs $|\psi_1\rangle$ et $|\psi_2\rangle$ d'un qubit. Si ces états étaient orthogonaux, il suffirait de prendre l'observable associée à cette base et le tour serait joué. Ici, on ne peut pas. Néanmoins, une mesure adaptée à la détection de l'état $|\psi_1^\perp\rangle$ ne donnera jamais un résultat positif lorsque la source prépare l'état $|\psi_1\rangle$. Comme la source prépare $|\psi_1\rangle$ ou bien $|\psi_2\rangle$, si ce détecteur donne un résultat positif, nous savons avec certitude que la source a préparé $|\psi_2\rangle$. De la même manière, un détecteur qui est sensible à l'état $|\psi_2^\perp\rangle$ détecte l'état $|\psi_1\rangle$ à coup sûr.

La probabilité que le détecteur $|\psi_1^\perp\rangle$ donne un résultat positif (valeur 1) est, selon l'état émis par la source

$$p(\psi_1^\perp|\psi_1) = 0, \quad (6.27a)$$

$$p(\psi_1^\perp|\psi_2) = |\langle\psi_1^\perp|\psi_2\rangle|^2. \quad (6.27b)$$

Pour une source qui émet l'état $|\psi_1\rangle$ avec la probabilité p_1 et l'état $|\psi_2\rangle$ avec la probabilité p_2 , la probabilité d'obtenir un résultat conclusif avec cette mesure

est donc donnée par

$$p(\psi_1^\perp) = p_2 |\langle \psi_1^\perp | \psi_2 \rangle|^2. \quad (6.28)$$

En considérant la détection de $|\psi_2^\perp\rangle$, la probabilité d'obtenir un résultat conclusif est

$$p(\psi_2^\perp) = p_1 |\langle \psi_1^\perp | \psi_2 \rangle|^2 \quad (6.29)$$

et donc, si on connaît les probabilités d'émission $p_{1,2}$, on peut choisir quelle détection va être la plus efficace. Dans ce cas, la probabilité maximale d'obtenir un résultat conclusif est donnée par

$$p_{\text{discr}} = |\langle \psi_1^\perp | \psi_2 \rangle|^2 \max(p_1, p_2). \quad (6.30)$$

Il est naturel de se demander s'il n'existerait pas un autre choix de mesure projective qui permette une détermination à coup sûr d'un des deux états $|\psi_{1,2}\rangle$. Il est facile de voir qu'il n'en est rien. En effet, le théorème de Bayes (voir section 2.4) nous dit que la probabilité que, étant donné que l'on a détecté un état $|\mathbf{m}\rangle$, la source ait émis l'état $|\psi_1\rangle$ ou bien l'état $|\psi_2\rangle$ est donnée par

$$p(\psi_\alpha | \mathbf{m}) = \frac{p(\psi_\alpha) p(\mathbf{m} | \psi_\alpha)}{p(\mathbf{m})}. \quad (6.31)$$

Comme nous n'avons que deux alternatives, $p(\psi_1 | \mathbf{m}) + p(\psi_2 | \mathbf{m}) = 1$ et donc la détection à coup sûr signifie qu'une de ces deux probabilités est égale à un ou à zéro. Mais si une de ces probabilités est égale à zéro, alors en vertu de l'équation (6.31), cela signifie que $p(\mathbf{m} | \psi_\alpha) = 0$ pour la valeur de α correspondante et donc que $\mathbf{m} = -\mathbf{n}_\alpha$ où $\mathbf{n}_\alpha$ est le vecteur de Bloch associé à l'état $|\psi_\alpha\rangle$. Ceci montre bien que les deux possibilités envisagées plus haut sont les seules à permettre une discrimination parfaite entre les deux états considérés.

Lorsque les états $|\psi_1\rangle$ et $|\psi_2\rangle$ sont orthogonaux, alors les deux résultats de mesure de la détection selon $|\psi_1^\perp\rangle$ permettent une discrimination à coup sûr des deux états. Mais, plus important, nous venons de montrer qu'en utilisant une mesure projective, la probabilité de discrimination à coup sûr de deux états non orthogonaux émis de manière équiprobable est finalement donnée par

$$p_{\text{discr}}^{(\text{proj})} = \frac{1}{2} (1 - |\langle \psi_1 | \psi_2 \rangle|^2). \quad (6.32)$$

Remarquons que la probabilité décroît brutalement de $1/2$ à 0 lorsque l'on passe d'états exactement orthogonaux à deux états de produit scalaire non nul. Cela provient du fait que l'on veut un protocole sans erreur : une mesure projective basée sur les états orthogonaux $|\psi_1\rangle$ et $|\psi_1^\perp\rangle$ nous dit si on est dans l'état $|\psi_2\rangle$ mais ne nous permet pas de dire à coup sûr si on est dans l'état $|\psi_1\rangle$. Elle introduit donc des erreurs sauf si $|\psi_1\rangle$ et $|\psi_2\rangle$ sont orthogonaux. Nous allons maintenant voir s'il n'est pas possible de faire mieux avec des mesures non projectives.

Discrimination à coup sûr par mesures généralisées

Rappelons qu'une mesure généralisée est décrite par des opérateurs M_j , pas nécessairement auto-adjoints, et qui vérifient $\sum_j M_j^\dagger M_j = \mathbb{1}$. Chaque j correspond à un résultat j possible et la probabilité d'observer ce résultat sachant que le système était dans un état pur $|\psi\rangle$ est donnée par $p(j|\psi) = \langle M_j^\dagger M_j \rangle_{|\psi\rangle}$. Le cas des mesures projectives revient à prendre comme opérateurs M_j des projecteurs orthogonaux. Puisque prendre $|\phi^\perp\rangle\langle\phi^\perp|$ et $|\psi^\perp\rangle\langle\psi^\perp|$ permet dans certains cas de dire quel état nous avons, il paraît judicieux de construire une mesure généralisée en partant de

$$M_1 = \sqrt{\alpha_1} |\psi_1^\perp\rangle\langle\psi_1^\perp|, \quad (6.33a)$$

$$M_2 = \sqrt{\alpha_2} |\psi_2^\perp\rangle\langle\psi_2^\perp|, \quad (6.33b)$$

où $\alpha_{1,2}$ sont deux nombres positifs que l'on ajustera de manière à pouvoir construire une mesure généralisée optimale pour le problème posé. Les opérateurs auto-adjoints positifs associés sont $E_1 = M_1^\dagger M_1 = \alpha_1 |\psi_1^\perp\rangle\langle\psi_1^\perp|$ et $E_2 = \alpha_2 |\psi_2^\perp\rangle\langle\psi_2^\perp|$. On va donc compléter la mesure avec un troisième résultat possible, que l'on notera 0 et auquel correspond l'opérateur hermitien positif

$$E_0 = \mathbb{1} - E_1 - E_2 = \mathbb{1} - \alpha_1 |\psi_1^\perp\rangle\langle\psi_1^\perp| - \alpha_2 |\psi_2^\perp\rangle\langle\psi_2^\perp|. \quad (6.34)$$

Cette mesure généralisée a donc trois résultats. Comme nous l'avons vu précédemment, le résultat 1 permet d'éliminer l'état $|\psi_1\rangle$. De même, si on mesure 2, on peut éliminer l'état $|\psi_2\rangle$. Chacun de ces deux résultats permet donc de savoir avec certitude quel état a été émis par la source. En revanche, le résultat 0 ne permet pas de discriminer puisque $p(0|\psi_1) = 1 - \alpha_2 \frac{1 - \mathbf{n}_1 \cdot \mathbf{n}_2}{2}$ et $p(0|\psi_2) = 1 - \alpha_1 \frac{1 - \mathbf{n}_1 \cdot \mathbf{n}_2}{2}$ qui, de manière générique, sont non nuls. C'est là qu'apparaît la beauté des mesures généralisées : alors qu'avec une mesure projective il serait impossible d'avoir accès aussi bien au résultat de la détection de $|\psi_1^\perp\rangle$ et de $|\psi_2^\perp\rangle$, il est possible de construire une mesure généralisée qui permet cela au prix cependant d'un troisième résultat qui est non discriminant.

Pour deux états purs équiprobables, l'optimum de la probabilité de discrimination est donné par l'expression [137, 217, 298] :

$$p_{\text{discr}}^{(\text{gen})} = 1 - |\langle\psi_1|\psi_2\rangle|. \quad (6.35)$$

Nous voyons que $p_{\text{discr}}^{(\text{gen})} > p_{\text{discr}}^{(\text{proj})}$ pour deux états orthogonaux et distincts : faire des mesures sur un système étendu permet effectivement de faire mieux que les mesures sur le seul système.

6.5.2 Discrimination sans ambiguïté mais avec erreur

L'autre stratégie de discrimination consiste à utiliser un protocole d'estimation conçu pour fournir une réponse parmi les deux alternatives $|\psi_1\rangle$ et $|\psi_2\rangle$

mais sans réponse non conclusive. Il y aura donc des erreurs. La performance est alors directement reliée au taux d'erreur minimal d'un tel protocole.

Considérons une mesure projective dans la direction $\mathbf{m}$ et utilisons un protocole tel que : si on trouve $+1$, alors on estime le vecteur à $\mathbf{n}_1$ et si on trouve -1 , alors on estime le vecteur à $\mathbf{n}_2$, où $\mathbf{n}_{1/2}$ sont les vecteurs de Bloch associés aux états $|\psi_{1/2}\rangle$. Les erreurs surviennent lorsque la source a préparé $|\psi_1\rangle$ et que l'on trouve -1 ou lorsque la source a préparé $|\psi_2\rangle$ et que l'on trouve $+1$. La probabilité d'erreur dépend du choix de la direction de mesure $\mathbf{m}$ et est donc directement donnée par

$$p_{\text{err}}(\mathbf{m}) = p_1 \cdot \frac{1 - \mathbf{m} \cdot \mathbf{n}_1}{2} + p_2 \cdot \frac{1 + \mathbf{m} \cdot \mathbf{n}_2}{2} \quad (6.36a)$$

$$= \frac{1 - (p_1 \mathbf{n}_1 - p_2 \mathbf{n}_2) \cdot \mathbf{m}}{2}. \quad (6.36b)$$

Le choix de $\mathbf{m}$ qui minimise cette probabilité est alors celui d'un vecteur unitaire dans la direction de $p_1 \mathbf{n}_1 - p_2 \mathbf{n}_2$. La probabilité minimale d'erreur est alors donnée par

$$p_{\text{err}}^{(\min)} = \frac{1}{2} \left(1 - \sqrt{1 - 4 p_1 p_2 |\langle \psi_1 | \psi_2 \rangle|^2} \right) \quad (6.37)$$

où l'on a utilisé que $\|p_1 \mathbf{n}_1 - p_2 \mathbf{n}_2\|^2 = p_1^2 + p_2^2 - 2 p_1 p_2 \mathbf{n}_1 \cdot \mathbf{n}_2$ ainsi que $\mathbf{n}_1 \cdot \mathbf{n}_2 = 2 |\langle \psi_1 | \psi_2 \rangle|^2 - 1$. La probabilité maximale de discriminer avec succès les deux états est alors inférieure à

$$p_{\text{discr}}^{(\text{amb})} = \frac{1}{2} \left(1 + \sqrt{1 - 4 p_1 p_2 |\langle \psi_1 | \psi_2 \rangle|^2} \right). \quad (6.38)$$

Cette borne est connue sous le nom de borne de Helstrom [200].

Dans le cas qui nous intéresse ici, pour des états orthogonaux ($\mathbf{n}_2 = -\mathbf{n}_1$), on retrouve bien une probabilité d'erreur nulle : il est possible de les discerner à coup sûr. Dans ce cas, l'algorithme de détection utilise d'ailleurs une mesure dans la direction $\pm \mathbf{n}_1$. Pour un mélange équilibré ($p_1 = p_2 = 1/2$), la direction de mesure qui optimise l'algorithme de détection est alors donnée par le vecteur normé $\mathbf{m} = (\mathbf{n}_1 - \mathbf{n}_2)/\|\mathbf{n}_1 - \mathbf{n}_2\|$. Ce compromis minimise le taux d'erreur et maximise l'information que l'on peut retirer sur chacun des états.

Pour des états presque orthogonaux, $p_{\text{err}}^{(\min)} \simeq p_1 p_2 |\langle \psi_1 | \psi_2 \rangle|^2$. Le taux d'erreur est proportionnel à la probabilité $|\langle \psi_1 | \psi_2 \rangle|^2$ qu'un état soit détecté dans l'autre. Le protocole choisi est toutefois meilleur que ce qu'on pourrait obtenir pour des objets classiques avec une probabilité de confusion de l'un en l'autre identique pour les deux. En effet, la probabilité d'erreur serait alors donnée par ce taux de confusion et donc par $|\langle \psi_1 | \psi_2 \rangle|^2$.

La borne de Helstrom [200] se généralise facilement au cas d'une source qui émet deux états pris au hasard au sein de deux ensembles statistiques ρ_1 et ρ_2 avec des probabilités respectives p_1 et p_2 . Dans ce cas, le taux d'erreur

minimal est donné par

$$p_{\text{err}}^{(\min)} = \frac{1}{2} (1 - \|p_1\rho_1 - p_2\rho_2\|_1) , \quad (6.39)$$

où $\|A\|_1 = \text{tr}(|A|)$ désigne la norme L_1 sur l'espace des opérateurs. Il est facile de voir que cette expression redonne bien (6.38) en prenant $\rho_1 = |\psi_1\rangle\langle\psi_1|$ et $\rho_2 = |\psi_2\rangle\langle\psi_2|$.

Si on a affaire à plus de deux états, la généralisation de (6.39) n'est pas connue même s'il est toujours possible de chercher la valeur de $p_{\text{err}}^{(\min)}$ numériquement [149] de manière efficace. Au niveau analytique, on connaît tout au plus une borne supérieure pour $p_{\text{err}}^{(\min)}$ due à Barnum et Knill [31] qui a trouvé des applications en cryptographie quantique, ainsi qu'un certain nombre de bornes inférieures [305, 274].

6.5.3 Comparaison des deux stratégies

Sur l'exemple d'une source émettant deux états purs de manière équiprobable, il est intéressant de comparer les bornes pour les protocoles sans erreur (6.30) et (6.35) et la borne (6.38) pour les protocoles sans réponse non discriminante mais avec erreur. Nous voyons sur la figure 6.3 que pour une telle source, la borne d'Helstrom est supérieure aux deux autres bornes. C'est donc le protocole avec erreur qui permet d'extraire l'information maximale sur les deux états générés par la source.

À la limite d'états très peu discernables, les deux familles de protocoles se comportent de manière très différente. Le protocole avec erreur donne un taux de discrimination correcte de 50 %, ce qui correspond à un choix de réponse aléatoire. Dans cette limite, les protocoles sans erreurs sont de moins en moins capables de fournir une réponse conclusive et voient leur probabilité de discrimination tendre vers zéro.

6.5.4 Application aux télécommunications

Ces considérations ont des applications directes dans le domaine des télécommunications. Rappelons qu'il existe deux grandes familles de protocoles de communication : des protocoles dits cohérents dans lesquels l'information est encodée dans la phase d'une onde porteuse et les protocoles dits incohérents dans lesquels l'information sera encodée dans l'intensité du signal. Les premiers sont plus sensibles aux effets de dispersion que les seconds mais présentent une meilleure efficacité spectrale, c'est-à-dire une meilleure capacité de transmission par unité de bande spectrale. En revanche les protocoles incohérents sont plus efficaces en termes d'information ramenée à la puissance utilisée. En fait, les protocoles cohérents semblent plus adaptés lorsque l'on désire un taux de transmission élevé dans une bande spectrale donnée [140].

Les protocoles par modulation de phase (*phase shift keying*, PSK) sont l'archétype de protocole cohérent : une porteuse véhicule une information

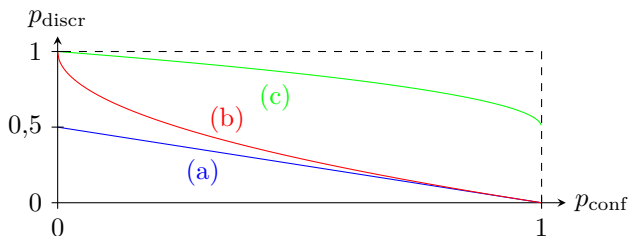


FIGURE 6.3 – Taux maximaux de réponses discriminantes correctes pour les différents protocoles en fonction de la probabilité de confusion $p_{\text{conf}} = |\langle \psi_1 | \psi_2 \rangle|^2$ entre deux états $|\psi_1\rangle$ et $|\psi_2\rangle$: (a) protocole sans erreur par mesures projectives, (b) protocoles sans erreur par mesures généralisées, (c) protocoles sans réponse non conclusive mais avec erreurs.

binaire sous la forme d’une modulation de phase ce qui nécessite, pour décoder, de pouvoir discerner entre deux ondes classiques de même amplitude mais dont la phase diffère de π . Cette technique est largement utilisée dans les télécommunications sans fil du quotidien (Wifi, RFID et Bluetooth) mais aussi par le *Deep Space Network* qui est le réseau de communication de la NASA avec les sondes spatiales interplanétaires⁴. Plusieurs variantes de PSK existent selon comment on encode des signaux binaires sur plusieurs valeurs de la phase mais le plus simple, appelé *binary phase shift keying*, consiste à encoder un bit sur deux phases maximale-ment séparées.

Pour les états cohérents $|\alpha\rangle$ et $|\alpha\rangle$, la probabilité de confusion est $P_{\text{conf}} = |\langle -\alpha | \alpha \rangle|^2 = \exp(-2|\alpha|^2)$. Les deux états n’étant pas discernables à coup sûr, il importe de trouver la méthode permettant de minimiser le taux d’erreur dans le décodage car cela permet d’augmenter la capacité de transmission d’information. Ce problème est également vital pour les communications spatiales à grande distance : l’atténuation par dispersion du faisceau émis par une sonde spatiale fait que l’on peut se retrouver à discerner entre deux états portant un nombre faible de photons ($\lesssim 1$).

Commençons par la méthode effectivement utilisée en pratique (mais qui n’est pas tout le temps optimale) : la *détection homodyne*. Elle est basée sur le fait que les fonctions d’ondes en la variable X de deux états cohérents de paramètres $\pm\alpha$ réels sont deux gaussiennes centrées sur $\pm\alpha$ et de largeur unité. Dans une approche semi-classique, on cherche donc à discerner deux distributions de probabilités gaussiennes. L’utilisation d’une méthode bayésienne donne alors une probabilité d’erreur

$$p_{\text{err}}^{(\text{Gauss})}(\alpha) = \frac{\text{erfc}(\sqrt{2}|\alpha|)}{2}, \quad (6.40)$$

4. Il est opéré par le *Jet Propulsion Laboratory* à Pasadena.

où $\operatorname{erfc}(x) = 2\pi^{-1/2} \int_x^{+\infty} e^{-u^2} du$. C'est ce que l'on peut obtenir par une méthode de détection homodyne et on montre que c'est l'optimal qui puisse être obtenu par une méthode dite gaussienne, c'est-à-dire qui transforme un état gaussien du champ électromagnétique en un autre état gaussien du champ électromagnétique [354].

Il est intéressant de comparer cette probabilité d'erreur au résultat de la borne de Helstrom qui est l'optimum donné par la physique quantique :

$$p_{\text{err}}^{(\text{Hel})}(\alpha) = \frac{1}{2}(1 - \sqrt{1 - e^{-4|\alpha|^2}}). \quad (6.41)$$

Lorsque le nombre moyen de photons $|\alpha|^2$ est grand (régime classique), la borne de Helstrom tend vers zéro comme $e^{-|\alpha|^2}/4 = p_{\text{conf}}/4$ alors qu'elle sature à $1/2$ à la limite des très faibles intensités $|\alpha|^2 \rightarrow 0$. Remarquons que dans le régime classique, $p_{\text{err}}^{(\text{Gauss})}(\alpha)$ est borné par [95]

$$\frac{1}{2} \left(1 - \sqrt{1 - e^{-8|\alpha|^2/\pi}} \right) \leq p_{\text{err}}^{(\text{Gauss})}(\alpha) \leq \frac{1}{2} \left(1 - \sqrt{1 - e^{-2|\alpha|^2}} \right). \quad (6.42)$$

On voit donc que, dans le régime classique, la méthode gaussienne donne un résultat bien moins bon que la borne optimale lorsque $p_{\text{conf}} \ll 1$. Cependant, en pratique, ce n'est pas un réel problème car les probabilités d'erreur sont exponentiellement décroissantes avec le nombre de photons moyen et donc avec l'intensité moyenne du faisceau. Notons que les probabilités d'erreur tendent vers $1/2$ à la limite $|\alpha| \rightarrow 0$, ce qui n'est pas surprenant car dans cette limite les états deviennent parfaitement indiscernables et le taux d'erreur est alors celui d'un choix aléatoire. Notons cependant que pour des petits nombres de photons, la méthode gaussienne donne des résultats proches de l'optimum, ce qui explique que ce soit encore la méthode utilisée dans les télécommunications spatiales à longue distance.

Se pose alors la question de savoir si on ne peut pas quand même faire mieux que la méthode homodyne et, en particulier, si on ne peut pas atteindre la borne de Helstrom. Une première proposition est le *récepteur de Kennedy* [227] qui utilise un photodétecteur idéal. Ce dispositif, que nous avons rapidement introduit à la section 5.4.2 émet un signal 0 lorsqu'il est irradié par le vide et un signal 1 dans tous les autres cas. Un tel détecteur ne transforme pas un état gaussien en un état gaussien. Pour discerner entre les états $|\alpha\rangle$ et $|\alpha\rangle$, commençons par leur appliquer un opérateur déplacement qui les transforme respectivement en $|2\alpha\rangle$ et $|0\rangle$. Dans le premier cas, le détecteur donnera le signal 1 avec la probabilité $|\langle 0|2\alpha\rangle|^2 = e^{-4|\alpha|^2}$ et le signal 0 avec la probabilité complémentaire. Dans le second cas, le détecteur émettra toujours le signal 0. Nous voyons donc que la probabilité d'erreur est

$$p_{\text{err}}^{(\text{Kenn})} = \frac{1}{2} e^{-4|\alpha|^2}. \quad (6.43)$$

Dans le régime classique, ce récepteur fait donc nettement mieux que la détection homodyne et n'est plus mauvais que l'optimum que d'un facteur

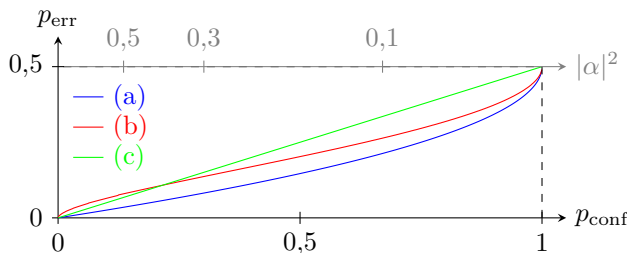


FIGURE 6.4 – Probabilité d’erreur dans la discrimination entre deux états cohérents de phases opposées en fonction de la probabilité de confusion $p_{\text{conf}} = e^{-4|\alpha|^2}$ ou de manière équivalente en fonction du nombre moyen de photons dans l’état considéré, à savoir $|\alpha|^2$. (a) Borne de Helstrom donnant le minimum absolu, (b) borne inférieure atteignable avec des méthodes gaussiennes, (c) taux d’erreur du récepteur de Kennedy. Dans cette représentation, le régime classique (grand nombre de photons) correspond à $p_{\text{conf}} \ll 1$ alors que le régime quantique correspond à p_{conf} proche de l’unité.

deux. En revanche, comme on le voit sur la figure 6.4, dans le régime quantique (petit nombre de photons), la méthode gaussienne reste meilleure.

En pratique, le déplacement dans une expérience d’optique est réalisable au moyen d’un mélange homodyne : on fait interférer le faisceau incident avec un état cohérent d’amplitude donnée sur une lame semi-réfléchissante. Si nous appelons λ l’amplitude de l’état cohérent ainsi envoyé, l’amplitude de l’onde cohérence transmise est alors $\pm\alpha\sqrt{T} + i\sqrt{1-T}\beta$ où T désigne la probabilité de transmission à travers la lame semi-réfléchissante. Pour annuler l’amplitude transmise dans le cas où on envoie $-\alpha$, il faut alors prendre $\beta = -i\sqrt{T/(1-T)}\alpha$. L’amplitude transmise lorsque l’on envoie α est alors $2\alpha\sqrt{T}$ qui tend bien vers 2α à la limite $T \rightarrow 1$. Séduisante sur le papier, cette proposition souffre de plusieurs problèmes : la limite $T \rightarrow 1$ peu atteignable en pratique et surtout une forte sensibilité aux imperfections du photodétecteur qui parfois envoie un signal 1 même lorsqu’il n’a reçu aucun photon (fausses détections) et de plus n’a pas une efficacité de 100 %.

Une variante de ce protocole dans laquelle on effectue un déplacement $\pm\alpha \rightarrow \pm\alpha + \gamma$ avec un choix de γ qui minimise la probabilité d’erreur en prenant en compte le fait que $T \neq 1$ a été proposée puis expérimentalement réalisée [381]. Il est intéressant de voir que pour des petits nombres de photons, cette méthode par déplacement optimisé bat la méthode gaussienne qui est pourtant meilleure que le récepteur de Kennedy dans ce régime. Cela constitue un premier exemple où l’utilisation d’un détecteur spécifiquement quantique permet des performances meilleures qu’un dispositif de détection basé sur des principes classiques.

Cependant, il est possible de concevoir un récepteur qui est capable d’atteindre la borne de Helstrom : c’est le *récepteur de Dolinar* [139]. Il fonctionne

en deux temps : tout d'abord, il éclate le signal d'entrée, qui était concentré sur un mode, sur N modes orthogonaux pour travailler non plus sur les deux états cohérents $|\pm\alpha\rangle$ mais sur deux états produits tensoriels⁵ qui sont :

$$|\psi_{\pm}^{(N)}\rangle = \left| \pm \frac{\alpha}{\sqrt{N}} \right\rangle^{\otimes N}. \quad (6.44)$$

Nous sommes donc ramenés au problème de discrimination de ces N copies qui vivent dans un espace produit tensoriel ayant au minimum une dimension 2^N . La borne de Helstrom s'applique encore mais la méthode que nous avons détaillée dans la section 6.5.2 nécessite d'effectuer une mesure collective sur les N copies comme il est facile de se convaincre en décomposant

$$|\psi_{\pm}^{(N)}\rangle = \cos \theta_N |0_N\rangle \pm \sin \theta_N |1_N\rangle, \quad (6.45)$$

où $|0_N\rangle$ et $|1_N\rangle$ sont des états normés et orthogonaux et où $\cos 2\theta_N = \langle -\alpha/\sqrt{N} | \alpha/\sqrt{N} \rangle^N = \exp(-2|\alpha|^2)$. Mais les mesures collectives sont difficiles à réaliser et, pour N fini, effectuer la même mesure individuelle sur chaque mode donnera forcément un résultat sous-optimal par rapport à la borne de Helstrom.

Le point remarquable est qu'il existe de manière générale un algorithme adaptatif [23, 5] qui permet, avec des mesures individuelles effectuées sur chacune des copies, d'arriver à la même probabilité d'erreur que la borne de Helstrom mais pour chaque valeur de N . Ce protocole, qui est en fait optimal, est réalisable relativement facilement sous réserve de pouvoir effectuer l'éclatement des états cohérents sur N modes, ce qui peut se faire en analysant le signal en temps réel [139]. Néanmoins, cela représente un défi technique qui fait que le protocole n'a pu être mis en œuvre que 30 ans après sa proposition [107]. Le récepteur de Dolinar est un exemple étonnant et instructif où un protocole adaptatif permet d'obtenir les mêmes performances que des mesures collectives. À l'inverse, dans le problème de la reconstitution d'un état quantique à partir d'un nombre fini de réalisation, on peut voir que les protocoles adaptatifs neaturent pas la borne de Massard et Popescu, même s'ils font mieux que les mesures individuelles [161].

La difficulté de mise en œuvre des méthodes optimales fait que, jusqu'à présent, la méthode homodyne reste la méthode de choix pour les télécommunications spatiales à grande distance. Compte tenu des progrès dans l'électronique bas bruit et du fait que l'on peut toujours déployer des antennes supplémentaires pour le réseau de communication en espace profond, il est probable que

5. Pour cela, il suffit de diffuser le mode de départ auquel est associé le couple d'opérateur de destruction et de destruction a et $a^\dagger$ sur de nouveaux modes associés aux opérateurs que nous noterons b_j et $b_j^\dagger$ pour $1 \leq j \leq N$ de sorte que $a = N^{-1/2} \sum_{j=1}^N b_j$ au moyen d'une matrice de diffusion unitaire. De la sorte, si nous mettons en entrée l'état $|\pm\alpha\rangle$ dans le mode de départ et tous les autres modes dans l'état vide, en sortie, nous récupérons un produit tensoriel de N copies de l'état cohérent $|\pm\alpha/\sqrt{N}\rangle$, les autres modes restant vides.

cela restera encore longtemps le cas au niveau du réseau basé sur Terre. En revanche, peut-être que ces technologies quantiques permettant d'atteindre la borne optimale pourraient s'avérer pertinentes pour la communication radio à grande distance entre sondes spatiales, ou éventuellement dans le domaine optique bien que, dans ce cas, on s'oriente plutôt sur des protocoles incohérents pour lesquels le récepteur de Dolinar n'apporte pas vraiment de gain. Il n'en reste pas moins que, du point de vue théorique, un récepteur atteignant la borne optimale et utilisant un protocole d'encodage binaire dans la phase est intéressant, surtout dans le domaine où l'efficacité spectrale par mode est inférieure à un bit par seconde et par hertz [140].

Chapitre 7

Communiquer en utilisant des qubits

Nous allons maintenant voir dans quelle mesure l'utilisation de ressources quantiques permet d'élaborer de nouveaux protocoles de communication et en quoi ceux-ci diffèrent des protocoles que nous connaissons dans le cadre classique. Ainsi, nous verrons que l'impossibilité de copier ou de caractériser un état quantique inconnu font de ceux-ci des ressources de choix pour la cryptographie. Nous verrons aussi que l'intrication peut être non seulement distribuée mais permet également l'élaboration de protocoles qui n'ont aucun équivalent classique. Bien évidemment, cela pose immédiatement la question de la puissance relative de ces différents outils : lorsque l'on se demande ce qu'un état intriqué partagé permet de faire ou si on peut l'utiliser pour communiquer classiquement, on pose la question de la comparaison et de la conversion des ressources utilisées pour communiquer.

Pour en donner une vision globale, nous utiliserons comme cadre de travail la théorie des ressources. Une théorie des ressources décrit un ensemble d'objets qui peuvent être convertis les uns dans les autres et qui permet d'en quantifier le coût. Elle nous permet donc de décrire des phénomènes physiques en prenant en compte l'action de l'observateur (l'agent pour les informaticiens) sur le système.

La théorie de Shannon quantique, qui généralise celle présentée au chapitre 2, se reformule particulièrement bien en ces termes. Cette approche permet de caractériser et comparer les limites et les performances des protocoles basés sur des ressources quantiques, tout en donnant un sens opérationnel aux quantités qu'elle fait intervenir.

7.1 Communiquer quantiquement

7.1.1 Motivation

La théorie de Shannon classique fournit la réponse à deux questions essentielles :

1. Comment compresser ce qui est émis par une source d'information classique ?
2. Quelle est la capacité de transmission d'information à travers un canal de transmission classique ?

Étendre la théorie de Shannon au monde quantique nécessite au minimum de répondre de nouveau à ces deux questions.

Généraliser la première consiste à se demander comment compresser une source quantique. Plus précisément, quelle est la dimension minimale de l'espace des états (ou de manière équivalente le nombre de qubits) dont nous avons besoin pour représenter les états émis par une source quantique ? La réponse est fournie par le théorème de Schumacher (section 7.3.3).

Pour aborder la seconde, nous devons définir la notion de canal quantique (section 7.1.2) qui permet de transférer des états quantiques entre Alice et Bob. Un tel canal peut servir à communiquer une information classique et il est naturel de se demander s'il peut en véhiculer plus qu'un canal classique. Nous verrons que ce n'est pas le cas. En revanche, l'utilisation d'un canal quantique permet une communication sécurisée même en présence d'un espion sur la ligne (section 7.2.1). Enfin, nous verrons que le fait de partager des ressources quantiques intriquées augmente la capacité de communication (section 7.2.2), même si le théorème de non-communication discuté à la section 6.4 stipule que l'intrication seule ne permet pas de communiquer.

L'étape suivante consiste à utiliser les canaux quantiques pour transmettre des états quantiques et non pas de simples messages classiques. Cette question est centrale si on cherche à construire une machine physique capable de manipuler de l'information quantique. En particulier, on doit se demander comment extraire, concentrer ou déplacer de l'intrication entre différents systèmes physiques comme dans un réseau de communications basé sur le partage d'états intriqués.

Ces considérations posent la question des relations entre les différentes ressources classiques et quantiques que l'on peut utiliser pour communiquer. Il ne faut pas perdre de vue qu'en pratique Alice, Bob et leurs camarades de jeu ne sont pas omnipotents : leurs capacités opérationnelles sont limitées. Par exemple, si on s'intéresse aux communications, on peut raisonnablement supposer que les opérations locales, les mesures et les transformations unitaires locales soient facilement réalisables alors que les opérations non locales le sont beaucoup moins. De même, les ressources quantiques non locales, par exemple le partage de paires intriquées ou un canal de communication quantique parfait, sont difficiles à générer.

Dans cette optique, analyser les relations entre ressources classiques et ressources quantiques revient à clarifier les contraintes sur les possibilités opératoires des différents acteurs. Cette question, très importante en pratique, sous-tend également la théorie de Shannon : la capacité d'un canal classique bruité borne ce qu'il est possible de faire en termes de communication. Le cadre conceptuel de la théorie des ressources (section 7.2.4) nous permettra de remettre en perspective les différents protocoles que nous allons présenter dans la première partie de ce chapitre.

7.1.2 Le canal quantique

La communication quantique nécessite de transmettre des états quantiques et donc d'introduire la notion de canal quantique. Un tel canal peut être utilisé soit pour transmettre une information classique, soit pour transmettre une information quantique. Nous allons distinguer ces deux usages grâce aux encodeurs et décodeurs situés aux extrémités du canal quantique. Ces appareils peuvent encoder des messages classiques sous la forme d'états quantiques, ou bien encoder des états quantiques sous forme d'autres états quantiques. Cette architecture nous permettra donc de donner une définition du canal quantique indépendante de son usage.

Alors que la notion de bruit dans un canal classique est assez intuitive, la notion correspondante dans le cas d'un canal quantique demande à être précisée. Pour un physicien, un exemple naturel est la ligne de transmission électrique ou optique imparfaite, c'est-à-dire couplée à son environnement. Comme nous le verrons plus précisément aux chapitres 9 et 10, le principal effet du couplage entre la ligne et son environnement est qu'il crée de l'intrication entre les degrés de liberté du système de transmission et ceux de l'environnement.

En toute généralité, un canal quantique réaliste est donc un système qui transforme tout état $|\psi_A\rangle \in \mathcal{H}_A$ injecté par Alice en un état intriqué entre Bob et les degrés de liberté environnementaux du canal $|\psi_A\rangle \mapsto |\psi_{B,\mathcal{E}}\rangle \in \mathcal{H}_B \otimes \mathcal{H}_{\mathcal{E}}$ tout en préservant les produits scalaires (voir figure 7.1). Cela revient à dire que, si l'environnement du canal est initialement dans l'état $|\mathcal{E}_0\rangle$, le canal quantique envoie, au moyen d'une isométrie, l'état $|\psi_A\rangle \otimes |\mathcal{E}_0\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{\mathcal{E}}$ sur l'état $|\psi_{B\mathcal{E}}\rangle \in \mathcal{H}_B \otimes \mathcal{H}_{\mathcal{E}}$. Lorsque $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}_{\mathcal{S}}$, nous retrouvons la dynamique sous-jacente à une mesure généralisée effectuée dans $\mathcal{H}_{\mathcal{S}}$ (section 5.4.1). En prenant la trace partielle sur l'environnement, l'état pur $|\psi_{B,\mathcal{E}}\rangle$ sera transformé en un opérateur densité

$$\rho_B = \mathcal{N}^{A \rightarrow B}(|\psi_A\rangle\langle\psi_A|) = \text{tr}_{\mathcal{H}_{\mathcal{E}}} |\Psi_{B\mathcal{E}}\rangle\langle\Psi_{B\mathcal{E}}|. \quad (7.1)$$

L'opération linéaire $\mathcal{N}^{A \rightarrow B}$, que nous discuterons plus précisément dans la section 9.5, caractérise totalement le canal quantique. Un canal parfait, qui est le cas particulier d'un canal qui ne crée pas d'intrication avec l'environnement, se contente d'envoyer $|\psi_A\rangle \in \mathcal{H}_A$ sur $U|\psi_A\rangle \in \mathcal{H}_B$ au moyen d'une transformation unitaire U .

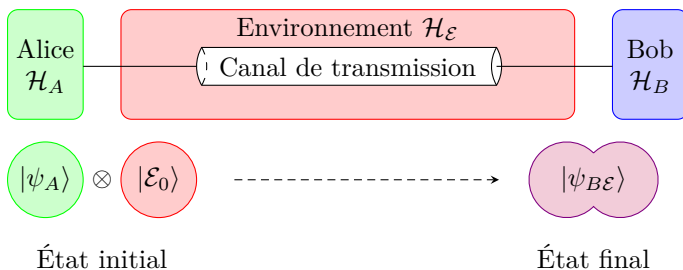


FIGURE 7.1 – Un canal quantique envoie linéairement, en préservant le produit scalaire, un état $|\psi_A\rangle \in \mathcal{H}_A$ sur l'état intriqué $|\psi_{BE}\rangle \in \mathcal{H}_B \otimes \mathcal{H}_E$.

7.2 Les protocoles quantiques élémentaires

Nous allons présenter maintenant les briques de base de la théorie de la communication quantique : ce sont quatre protocoles utilisant les ressources quantiques élémentaires que sont la transmission parfaite d'un qubit et d'une paire de qubit maximalement intriquée.

7.2.1 Le partage de secret au moyen d'états quantiques

Les enjeux de la cryptographie

La cryptographie s'intéresse à la question du partage secret d'information entre différents opérateurs. Apparue très tôt dans le domaine militaire, la protection contre l'espionnage est une question encore plus cruciale à l'heure d'Internet et dans le monde de la finance comme pour nombre d'entreprises. Comme on peut le lire dans un ouvrage de référence [32] : *On peut dire que la cryptographie s'occupe de communication en présence d'adversaires*. Une des façons de se protéger des adversaires est de cacher l'existence du message : c'est ce qu'on appelle la stéganographie. Cela peut être fait en cachant un papier dans la doublure d'un vêtement ou de manière plus moderne, en encodant le message dans un «bruit» ajouté à une image.

Un autre moyen de protéger des informations est la cryptographie, qui fonctionne en chiffrant les données que l'on souhaite protéger de manière à les rendre inintelligibles pour tout autre que le destinataire souhaité. Cela peut se faire de différentes façons. Historiquement, le secret du message reposait sur le secret de l'algorithme utilisé pour le chiffrement. Cependant, une fois cet algorithme deviné ou divulgué, il est très facile de déchiffrer le message. C'est pourquoi la cryptographie moderne applique le principe de Kerckhoffs, qui stipule que, s'il doit bien y avoir une clé secrète, l'algorithme pour chiffrer le message avec cette clé est en revanche connu de tous.

C'est le cas par exemple du chiffre de Vernam, ou masque jetable (*one-time pad*). Deux personnes partagent une clé secrète binaire de la même longueur

que le message à échanger. Le message chiffré est l'addition bit à bit modulo 2 du message et de la clé. Le déchiffrement se fait par l'addition du message chiffré et de la clé. Ce chiffrement est le seul à être incassable, à condition que chaque clé secrète ne soit utilisée qu'une seule fois, qu'elle soit le plus aléatoire possible, et qu'elle reste secrète. Comment échanger cette clé sans risque de copie par un adversaire ? La seule solution sécurisée dans le monde classique consiste à transmettre cette clé « de la main à la main », comme pour le téléphone rouge entre Washington et Moscou : des liasses de 0 et de 1 sont acheminées par la valise diplomatique. C'est ce problème d'échange sécurisé d'une clé totalement aléatoire que résout le protocole quantique BB84 que nous présentons dans la section suivante.

Plus généralement, on distingue deux grands types de cryptographie : la *cryptographie symétrique* et la *cryptographie asymétrique*. Dans la cryptographie symétrique, les deux personnes communiquant de manière secrète partagent un mot binaire secret, la clé, leur permettant de chiffrer ou de déchiffrer un message. C'est le cas du chiffre de Vernam. Mais il existe aussi des clés de chiffrement utilisant des clés aléatoires plus courtes. Le chiffrement en continu comme E0, utilisé par *Bluetooth*, ou par bloc comme AES¹, en sont des exemples actuels. Ces clés sont des graines (*seeds*) qui permettent d'engendrer des suites pseudo-aléatoires plus longues comme dans les machines allemandes Enigma durant la seconde guerre mondiale. Chaque jour, l'opérateur de la machine lisait dans un livre secret un mot (la graine) lui indiquant comment configurer la machine, ce qui définissait le codage du jour. Après la récupération d'une machine Enigma, une percée décisive des cryptologues alliés fut donc la reconstitution de la clé secrète quotidienne en écoutant les messages encodés.

Il est important de remarquer que la sûreté du masque jetable et des méthodes modernes ne repose pas sur le même principe. La sûreté du masque jetable repose sur la complexité informationnelle, ou complexité de Kolmogorov, qui quantifie dans quelle mesure une suite est parfaitement ou imparfaitement aléatoire. En revanche, la sûreté des méthodes modernes repose sur la complexité du calcul : décrypter le message est possible théoriquement, mais les calculs sont très longs. Or un espion qui mettrait plus d'un siècle à décrypter un message n'est pas un espion très utile... Bien sûr, cela signifie aussi que les méthodes de chiffrement doivent progresser avec la puissance des ordinateurs.

Comment prouvons-nous qu'une méthode de chiffrement présente une grande complexité algorithmique de déchiffrement ? Et bien, c'est impossible dans l'absolu. Nous pouvons au mieux comparer les complexités entre elles, et supposer que nous vivons dans le monde « Cryptomania » inventé par Russell Impagliazzo [214].

Dans ce monde, il est possible de communiquer de manière sûre grâce à l'existence de fonctions à sens unique (*one-way function*) : ce sont des fonctions faciles à calculer dont la fonction réciproque n'est pas facile à calculer. Avec une fonction à sens unique, on peut chiffrer facilement un message qui sera difficile

1. Le standard américain largement utilisé pour les transferts sécurisés sur Internet.

à décrypter. Si nous vivions dans « Algorithmica », monde d'Impagliazzo dans lequel $P = NP$ (chapitre 3) et où tout est facile à calculer, les échanges secrets seraient impossibles. Mais peut-être vivons-nous dans « Heuristica », un monde intermédiaire où il existe des fonctions difficiles à calculer dans le pire cas mais faciles en moyenne. Les informaticiens qui travaillent sur ce sujet ont cependant l'intuition qu'il existe des fonctions à sens unique et que notre monde est bien « Cryptomania », mais cela reste à prouver.

Revenons aux différentes méthodes de cryptographie moderne. D'un côté nous avons donc les systèmes symétriques, reposant sur le partage d'une graine. Mais comment faire pour échanger à distance et de manière secrète la clé qui va nous permettre ce chiffrement ? C'est ce que nous permettent les protocoles de cryptographie asymétrique ou cryptographie à clé publique, apparus dans les années 1970. Dans ces protocoles, les clés utilisées pour chiffrer et déchiffrer sont différentes, la première est publique et l'autre privée et personnelle. Le plus connu d'entre-eux est le protocole RSA.

En pratique, imaginons qu'Alice doit recevoir un message privé de Bob. Elle engendre un couple de clés, dont l'une est privée et gardée sur sa machine, et l'autre est publique et peut être transmise à Bob par des canaux non sécurisés. À l'aide de cette clé publique, Bob chiffre son message pour Alice, et le diffuse *via* un canal non sécurisé. Alice est la seule à pouvoir le déchiffrer grâce à sa clé secrète. Ce protocole permet en particulier l'échange d'une clé secrète qui sera ensuite utilisée dans un protocole symétrique. Ces protocoles ont une propriété supplémentaire qui permet l'authentification des auteurs du message : la clé publique permet de déchiffrer un message codé avec la clé privée. Ainsi, Alice peut envoyer à Bob un message qu'elle aura codé avec sa clé secrète et la clé publique de Bob. Bob sera alors le seul à pouvoir déchiffrer le message avec sa clé secrète et la clé publique d'Alice. De cette manière, il pourra être sûr qu'Alice est bien l'auteure du message.

Au final, la sécurité des protocoles asymétriques comme des symétriques repose sur leur complexité algorithmique et sur le fait de vivre dans un des mondes d'Impagliazzo dans lequel il existe ce qu'on appelle un *gap* de complexité ($P \neq NP$).

Un protocole quantique d'échange de clé : BB84

La sécurité d'une communication chiffrée par un protocole symétrique reposant en grande partie sur la confidentialité de la clé utilisée, il est d'une importance capitale que celle-ci ne puisse pas être interceptée ou lue par une tierce partie. Dans les années 1980, l'idée est apparue d'utiliser la physique quantique pour s'échanger cette clé. Puisqu'il est impossible de cloner un état quantique inconnu, c'est-à-dire que l'on ne peut pas lire l'état quantique « sans le modifier », on peut imaginer un protocole permettant de détecter si un état a été lu par un espion et donc si la clé est compromise. C'est l'idée derrière le *protocole BB84* inventé par Charles Bennett et Gilles Brassard en 1984 [41]

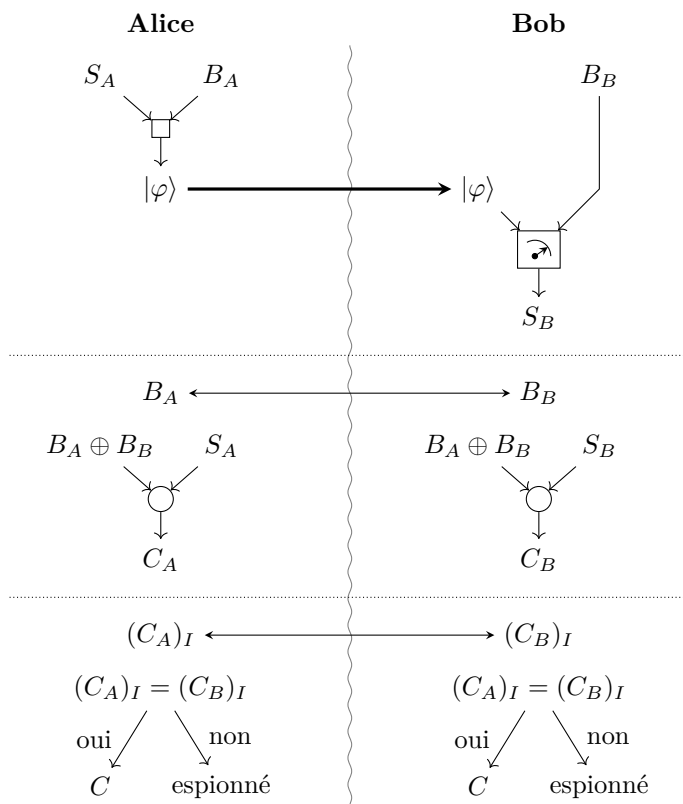


FIGURE 7.2 – Le protocole BB84. Le protocole s’effectue en trois phases. Dans la première phase, Alice prépare des états quantiques à l’aide des suites aléatoires S_A (suite à encoder) et B_A (choix de base), et les envoie à Bob. Bob les mesure dans la base donnée par la suite aléatoire B_B . Dans la deuxième phase, Alice et Bob s’échangent les bases utilisées. Alice trouve un candidat C_A et Bob un candidat C_B pour la clé. Enfin, a lieu une phase de vérification dans laquelle Alice et Bob s’échangent une partie de leur clé candidate. Si ces parties sont similaires, ils en déduisent avec une certaine probabilité que la conversation n’a pas été espionnée.

qui permet de sécuriser l’étape de transmission de la clé dans un protocole de cryptographie symétrique.

Dans ce protocole, Alice et Bob veulent s’échanger un mot binaire C (la clé) tout en s’assurant qu’ils n’ont pas été espionnés et que la clé n’a pas été interceptée par l’espionne Ève. Le protocole se fera en trois phases décrites sur la figure 7.2.

Le qubit nous offre un système quantique idéal pour encoder une information binaire : deux états d’une base orthonormée encoderont deux états

logiques. Ici, nous allons avoir besoin de deux telles bases, $\mathcal{B}_0 = \{|0\rangle, |1\rangle\}$ et $\mathcal{B}_1 = \{|+\rangle, |-\rangle\}$ avec $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Le bit 0 sera alors codé par $|0\rangle$ dans $\mathcal{B}_0$ et par $|+\rangle$ dans $\mathcal{B}_1$ alors que le bit 1 sera codé par $|1\rangle$ dans $\mathcal{B}_0$ et par $|-\rangle$ dans $\mathcal{B}_1$.

La première phase du protocole consiste à transférer un état quantique utilisant ce codage d’Alice vers Bob. Alice tire, au hasard, la suite S_A qui donnera, au final, la clé C . Alice tire également au hasard la suite B_A qui servira au choix de la base d’encodage d’Alice. Par exemple, le tirage de $1 \in B_A$ et de $1 \in S_A$ implique de préparer l’état $|-\rangle$, bit logique 1 dans la base $\mathcal{B}_1$. Alice envoie ensuite ses qubits à Bob.

De son côté, Bob choisit une suite B_B où chaque bit de la suite lui indiquera la base de mesure pour chaque qubit reçu. Après avoir effectué ces mesures, il obtient la suite S_B des résultats.

La seconde phase du protocole consiste à extraire la clé depuis les suites S_A et S_B . Dans le cas où le canal quantique n’est pas écouté, on a $S_{A,i} = S_{B,i}$ lorsque les bases choisies par Alice et par Bob coïncident. Alice et Bob s’échangent donc les bases B_A et B_B par un canal classique, et ne gardent que les bits des suites S_A et S_B pour lesquels $B_{A,i} = B_{B,i}$.

Cependant, Alice et Bob doivent pouvoir vérifier qu’ils n’ont pas été espionnés, ce qui est l’objet de la troisième phase du protocole. Pour cela, Alice va choisir au hasard un ensemble d’indices I de taille p et l’annoncer à Bob. Alice et Bob s’échangent alors les bits correspondant à cet ensemble. Si les clés ne sont pas égales sur l’ensemble des indices envoyés par Alice, alors on déduit que le canal quantique a été espionné par Ève. Dans le cas contraire, on retire à la clé les indices envoyés sur le canal classique et on obtient la clé C .

Les deux premières phases du protocole BB84 sont illustrées par l’exemple donné par le tableau suivant :

Codage								
S_A	0	1	1	0	0	1	0	1
B_A	0	0	1	1	0	0	1	1
φ	$ 0\rangle$	$ 1\rangle$	$ -\rangle$	$ +\rangle$	$ 0\rangle$	$ 1\rangle$	$ +\rangle$	$ -\rangle$
Décodage								
B_B	0	0	1	0	1	0	1	0
S_B	0	1	1	0 ou 1	0 ou 1	1	0	0 ou 1
Clé								
C	0	1	1			1	0	

Quelle est la probabilité qu’Alice et Bob détectent une espionne Ève ? Supposons qu’Ève agisse de la manière suivante. Pour chaque bit A_i qu’elle veut espionner, elle intercepte l’état $|\varphi_i\rangle$ envoyé par Alice. Ensuite, elle tire au hasard une base $\mathcal{B}_0$ ou $\mathcal{B}_1$, et observe $|\varphi_i\rangle$ dans cette base pour en déduire A_i et renvoyer le qubit qu’elle a observé à Bob. L’état renvoyé est le bon si

elle a observé dans la bonne base, donc avec une probabilité $1/2$. Dans ce cas, Alice et Bob ne s'aperçoivent de rien. Si Ève n'a pas la bonne base, alors Bob n'observe le bon résultat qu'avec une probabilité $1/2$. Au final, Alice et Bob ont une probabilité d'une chance sur quatre de détecter chaque observation de l'espionne. Leur probabilité de détecter l'espionnage des p qubits est donc au final de $1 - (3/4)^p$. En augmentant p , nous pouvons rendre cette probabilité aussi proche que souhaité de 1. Après avoir établi la sécurité de la transmission à un niveau suffisant pour les deux parties avec ces p bits de test, il reste en moyenne $(n/2 - p)$ bits connus à la fois d'Alice et Bob qui n'ont pas été transmis en clair, et qui sont donc utilisables pour la clé C .

L'analyse que nous avons faite est cependant très éloignée du cas pratique pour deux raisons. La première est que nous avons supposé que le canal quantique était parfait. Dans la pratique, même sans espion, le canal quantique introduira des erreurs. Il faut donc un protocole de réconciliation, qui permet à Alice et à Bob de corriger les erreurs en faisant fuiter le moins d'information possible. Cette étape de réconciliation, décrite pour la première fois dans [40], donne néanmoins une information indirecte sur les clés. De même, on ne peut pas savoir si les erreurs corrigées viennent du bruit ou d'un espionnage de certains bits par Ève. Il faut donc amplifier le secret en réduisant la taille de la clé.

La seconde raison pour laquelle nous sommes éloignés du cas pratique est qu'il est difficile de faire une source émettant un seul quanta. La méthode la plus simple est d'utiliser des lasers atténués avec une statistique poissonnienne, ce qui implique qu'il existe une probabilité non nulle qu'une impulsion contienne plusieurs photons. Dans ce cas, il est possible d'utiliser une attaque nommée *photon number splitting* que l'on peut traduire par séparation des impulsions multiphotoniques [66]. Cette attaque consiste à jeter les impulsions contenant un seul photon et à couper les impulsions multiphotoniques en deux au moyen d'une lame semi-réfléchissante. Ève garde la moitié de l'impulsion dans une mémoire quantique et envoie le reste à Bob. Si Ève capte l'échange de bases entre Alice et Bob, elle est alors capable de mesurer l'information stockée et ainsi de récupérer la clé sans être détectée. Une façon de se prémunir de ces attaques consiste à utiliser les « états leurres ». L'idée consiste à envoyer côté Alice plusieurs impulsions d'intensités choisies aléatoirement, l'une d'elles véhiculant le signal et les autres étant des leurres. À la fin, Alice annonce publiquement quelle a été l'impulsion codante. En analysant le nombre de clics du photomultiplicateur côté Bob, Alice et Bob peuvent alors détecter une attaque basée sur la séparation pour les impulsions multiphotons [213]. Avec cette méthode et ses variantes, il est possible de transmettre des clés quantiquement à l'échelle intercontinentale au moyen de satellites [258].

Le protocole E91

Le protocole BB84 repose sur l'impossibilité de cloner un état quantique inconnu. Il est également possible d'utiliser l'intrication pour effectuer un

transfert de clé. C'est l'idée du *protocole E91*, inventé par Artur Ekert [148]. Dans ce protocole, Alice et Bob échangent des paires intriquées² et effectuent des mesures analogues à ce qui est fait dans le protocole BB84. Si Ève espionne la communication quantique, les corrélations seront d'origine uniquement classique. Nous avons évoqué dans le chapitre 5 et nous verrons dans le chapitre 11 qu'il existe un test permettant de différencier les corrélations quantiques des corrélations classiques. En utilisant ce test, on peut donc détecter l'espionne Ève. L'intérêt de ce protocole est que l'on peut vérifier la présence d'un espion en observant la structure des corrélations entre Alice et Bob. Ce test est générique et ouvre donc la voie à des protocoles de distribution de clé quantique indépendants du matériel sous-jacent [366].

La distribution de clé quantique en pratique

La distribution de clé quantique utilise les propriétés de la théorie quantique pour s'assurer que le canal quantique n'a pas été espionné. La clé distribuée peut alors servir au chiffre de Vernam, dont la sécurité est assurée par la théorie de l'information. Ainsi, l'intérêt théorique de la distribution de clé quantique est indiscutable : celle-ci permet de faire reposer la cryptographie sur la validité de la théorie quantique plutôt que sur des hypothèses de théorie de complexité.

Si l'intérêt théorique ne fait aucun doute, qu'en est-il en pratique ? Un élément très important, et souvent oublié, est que la sécurité est assurée par une chaîne. La distribution de clé, ainsi que le chiffrement, sont deux éléments importants de cette chaîne, mais ce ne sont pas les seuls. Le protocole de communication, ainsi que celui d'authentification qui permet à Bob de s'assurer qu'il parle bien à Alice font également partie de cette chaîne, de même que les logiciels et le matériel utilisés pour communiquer.

Par exemple, certaines solutions de cryptographie quantique ont été entièrement cassées en saturant les photodiodes à avalanche. Il est en effet possible de prendre le contrôle de la photodiode en émettant une lumière intense dans la fibre et ainsi de récupérer entièrement la clé sans laisser de trace [263]. On a bien là des systèmes parfaitement sûrs en théorie et dont la sécurité est absolument nulle en pratique.

Cela dit, il est possible de mettre en œuvre des protocoles de cryptographie quantique relativement indépendants du système physique sous-jacent, ce qui devrait limiter les attaques visant un matériel spécifique. Reste alors la question des autres éléments de la chaîne de sécurité. L'histoire de la cryptographie moderne montre que la très grande majorité des attaques ne ciblent pas directement le chiffrement mais plutôt des bogues dans les logiciels ou dans les protocoles de communication. La partie cryptographie est ainsi la plus solide de la chaîne, et c'est la seule que la cryptographie quantique prétend remplacer. Cela a conduit le cryptologue Schneier à déclarer que la

2. Alice et Bob peuvent également se reposer sur une source commune.

cryptographie quantique, si l'idée est incroyablement jolie, est presque toujours inutile en pratique [321]. Cet avis est conforté par les services secrets anglais qui déconseillent d'utiliser la cryptographie quantique [182] pour les mêmes raisons.

Notons tout de même une grande différence entre la cryptographie classique et la cryptographie quantique : dans le cas quantique, il est possible de savoir si un espion écoute le canal alors que c'est impossible sur un canal classique. Cette différence vient également avec une importante difficulté : que faire quand un espion écoute le canal quantique ? Si on arrête la communication, cela signifie qu'un adversaire ayant de simples capacités d'écoute peut empêcher la communication.

La distribution de clés quantiques n'est néanmoins pas la seule chose que l'on puisse faire avec des ressources quantiques. L'objet de la section 7.2.2 va être de voir comment on peut utiliser l'intrication comme ressource pour la communication.

7.2.2 Utiliser l'intrication pour communiquer

Distribution d'intrication

La *distribution d'intrication* permet de générer une paire de qubits maximale-ment intriqués dont l'un est à disposition d'Alice et l'autre de Bob. Elle crée donc cette ressource non locale indispensable pour les protocoles de téléportation quantique et de codage superdense que nous discuterons ensuite.

La distribution d'intrication fonctionne en deux temps : Alice commence par créer une paire maximale-ment intriquée, puis elle utilise un canal quantique parfait pour transmettre un des deux qubits à sa disposition à Bob. Le résultat est une paire maximale-ment intriquée partagée entre Alice et Bob. La création de la paire se fait localement très simplement : on applique aux quatre vecteurs de la base canonique l'opération décrite par la matrice de Hadamard

$$H = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \quad (7.2)$$

sur le premier qubit suivi par l'opération **cnot** qui agit sur les deux qubits et dont la matrice est

$$M_{\text{cnot}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (7.3)$$

De telles opérations unitaires sont appelées portes à un et deux qubits et constituent la généralisation des portes logiques vues au chapitre 3. Comme dans le cas classique, elles peuvent être combinées en des circuits quantiques pour réaliser des opérations unitaires portant sur plusieurs qubits. De manière plus générale, on peut combiner bits classiques et bits quantiques dans un circuit quantique.

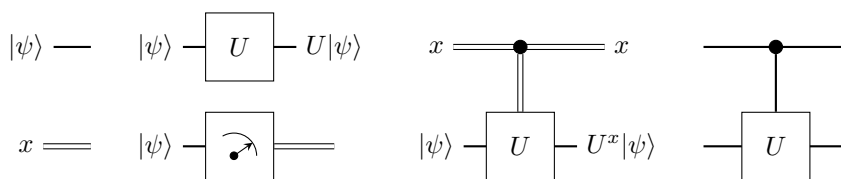


FIGURE 7.3 – Présentation des éléments d’une représentation en circuit quantique. 1^{re} colonne : un qubit est représenté par une ligne épaisse, tandis qu’un bit classique est représenté par une double ligne. 2^e colonne : on peut agir sur un qubit à l’aide d’un opérateur, représenté par un carré, ou en le mesurant, ce qui est représenté par une aiguille. 3^e colonne : on peut également utiliser un bit classique comme contrôle de l’action d’un opérateur, ce que l’on représente avec un point et une double ligne verticale. L’opérateur n’est appliqué que si le bit classique considéré (ici x) vaut 1 d’où la notation en exposant U^x . 4^e colonne : enfin, on peut également contrôler un opérateur par un qubit, et ne l’appliquer que si l’état du qubit de contrôle est $|1\rangle$.

Illustrons ce concept de représentation en *circuit quantique* pour mieux visualiser les différents protocoles que nous allons présenter. Un tel circuit est constitué des éléments visibles sur la figure 7.3 et se lit de gauche à droite. Au cours du temps un qubit (représenté par une ligne épaisse) peut être modifié par l’action d’un opérateur (dans un carré). L’état d’un qubit peut également contrôler l’action d’un opérateur (ce qui est représenté par un point et une ligne de contrôle), ou être mesuré (ce que l’on représente par une aiguille). Si des bits classiques interviennent dans le protocole, on les représente par une double ligne, tandis qu’un opérateur sur plusieurs qubits aura plusieurs lignes en entrée et en sortie.

Par exemple, le protocole de distribution d’intrication complet est représenté sur la figure 7.4. Il transforme chacun des quatre états $|xy\rangle$ en un état maximalelement intriqué que nous noterons $|B_{xy}\rangle$ et qui pour $x = y = 0$ correspond à l’équation (5.22). L’opérateur associé au circuit étant unitaire, on obtient ainsi une base orthonormée de l’espace des états de deux qubits, appelée *base de Bell*. Renverser le circuit permet de fabriquer un appareil faisant une mesure projective dans la base de Bell.

Le codage superdense

Puisqu’un qubit est un système paramétré par deux angles et qu’un bit classique probabiliste ne nécessite qu’un nombre réel, est-il possible de transmettre deux bits classiques d’information en n’utilisant qu’un seul qubit ?

Alice peut évidemment envoyer à Bob un bit d’information classique en utilisant un qubit. Pour cela, elle code 0 en $|0\rangle$ et 1 en $|1\rangle$, puis envoie le qubit correspondant à Bob. Cependant, un tel protocole naïf ne permet pas à Alice d’envoyer à Bob plus d’un seul bit d’information classique par qubit. En effet,

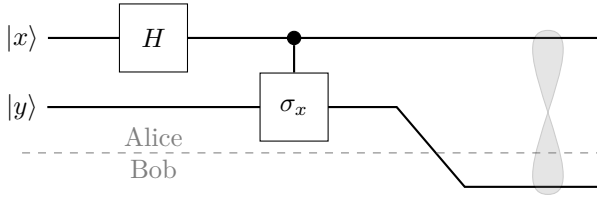


FIGURE 7.4 – Circuit quantique du protocole de distribution d'intrication d'un état de la base de Bell entre Alice et Bob. Alice commence par créer un tel état intriqué en combinant une porte de Hadamard avec une porte **cnot**. Elle utilise ensuite un canal quantique parfait pour transmettre un des deux qubits physiques à Bob. Si on injecte $|xy\rangle$ en entrée côté Alice, les deux parties partagent l'état $|B_{xy}\rangle$ en sortie.

nous savons (chapitre 6) qu'il est impossible de discerner à coup sûr entre deux états non orthogonaux. On doit donc faire une mesure qui distingue entre deux états orthogonaux pour extraire le premier bit. Une mesure projective permet de le faire mais cela altère alors l'état de manière irréversible et on ne peut donc extraire un bit de plus. Pour transmettre deux bits classiques au moyen d'un seul qubit, il faut donc une ressource supplémentaire qui sera l'intrication.

Pour préparer cet échange, Alice et Bob se mettent d'accord sur la procédure suivante : tout d'abord, le protocole de distribution d'intrication leur permet de partager les qubits d'une paire maximalelement intriquée dans l'état de Bell $|B_{00}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ (Alice disposant du second qubit de celle-ci). Décrivons maintenant l'algorithme de *codage superdense* :

1. Alice veut envoyer les bits classiques xy à Bob. Lorsque $x = 1$, elle applique l'opérateur $M_{\text{flip}}^y = \sigma_z$ qui introduit un déphasage de π sur le qubit en sa possession. Sinon elle ne fait rien. De même, lorsque $y = 1$, elle applique l'opérateur **not** et sinon elle ne fait rien. L'état intriqué partagé entre Alice et Bob est alors

$$(I_2 \otimes M_{\text{not}}^y M_{\text{flip}}^x) |B_{00}\rangle = |B_{xy}\rangle . \quad (7.4)$$

2. Alice envoie son qubit à Bob.

3. Bob mesure les qubits dans la base de Bell et récupère alors yx .

Le circuit décrivant le protocole de codage superdense représenté sur la figure 7.5 montre que l'on peut transmettre plus d'un bit classique par qubit dès lors que l'on dispose d'une paire intriquée partagée. Cela ne contredit pas l'idée qu'un qubit ne puisse véhiculer qu'un bit classique : en effet, le partage d'une paire maximalelement intriquée peut se faire au moyen d'opérations locales et de la transmission d'un seul qubit. En composant le codage superdense avec la distribution d'intrication, on voit donc que la transmission de deux bits classiques repose sur la transmission de deux qubits d'Alice vers Bob pour peu que la paire intriquée ait été produite localement chez Alice.

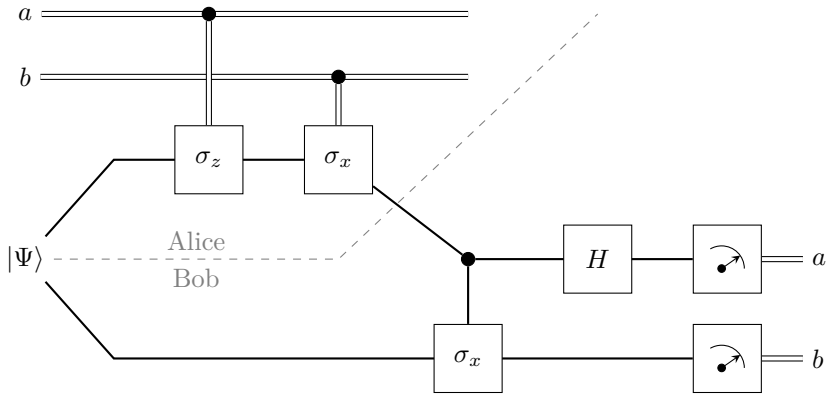


FIGURE 7.5 – Circuit quantique du protocole de codage superdense. Un état $|\psi\rangle$ intriqué est partagé entre Alice et Bob. Alice agit sur son qubit à partir des bits classiques qu'elle souhaite transmettre à Bob puis procède à l'envoi. Lorsque Bob le reçoit, il applique un opérateur **cnot** contrôlé par ce qubit à celui qu'il avait depuis le début, puis un opérateur de Hadamard, avant de mesurer le résultat. Il trouve alors les 2 bits classiques a et b , alors qu'un seul qubit a été transmis. Le seul moyen de communication utilisé entre Alice et Bob est le canal qui transmet un bit quantique.

La téléportation quantique

La dernière idée consiste à utiliser la notion d'état relatif. En effet, si Alice et Bob partagent un état intriqué, la donnée des résultats d'une mesure effectuée par Alice assure que Bob dispose d'un état quantique parfaitement défini (chapitre 5). Si Alice dispose de deux qubits dont un est intriqué avec Bob, on peut espérer que les résultats de mesures collectives, c'est-à-dire affectant les deux qubits de manière corrélée, permettent à Bob d'apprendre quelque chose sur les états des qubits dont dispose Alice.

Cette idée permet un transfert d'un état quantique $|\varphi\rangle$ d'Alice vers Bob dont le protocole associé est appelé protocole de *téléportation quantique*. Bien sûr, cela nécessite la transmission d'une information classique, en l'occurrence deux bits classiques, d'Alice à Bob. *Via* l'intrication, on va donc pouvoir « transférer » l'état d'un qubit sans avoir à déplacer le qubit lui-même. Ce protocole fonctionne comme suit :

1. En préparation de la transmission, Alice et Bob partagent la paire maximalement intriquée $|B_{00}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, et Alice dispose d'un état générique $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ qu'elle souhaite transmettre à Bob.
2. Alice mesure ses qubits dans la base de Bell $|B_{ab}\rangle$ et obtient deux bits d'information a et b . Elle transmet alors (a, b) à Bob par un canal de communication classique.
3. Bob applique les opérateurs $M_{\text{not}}^b M_{\text{flip}}^a = \sigma_x^b \sigma_z^a$ à son qubit.

Montrons que ce protocole produit bien la « téléportation » de l'état quantique $|\varphi\rangle$ d'Alice à Bob. On démarre avec un état à trois qubits de la forme

$$\begin{aligned} |\psi_1\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ &= \frac{\alpha}{\sqrt{2}}(|000\rangle + |011\rangle) + \frac{\beta}{\sqrt{2}}(|100\rangle + |111\rangle). \end{aligned} \quad (7.5)$$

La seconde étape revient à appliquer le circuit qui transforme la base de Bell en la base canonique, puis à mesurer l'état des deux qubits d'Alice dans la base canonique. Avant mesure, l'état global des trois qubits est donc :

$$\begin{aligned} |\psi_2\rangle &= |00\rangle \left(\frac{\alpha}{2}|0\rangle + \frac{\beta}{2}|1\rangle \right) + |01\rangle \left(\frac{\alpha}{2}|1\rangle + \frac{\beta}{2}|0\rangle \right) \\ &\quad + |10\rangle \left(\frac{\alpha}{2}|0\rangle - \frac{\beta}{2}|1\rangle \right) + |11\rangle \left(\frac{\alpha}{2}|1\rangle - \frac{\beta}{2}|0\rangle \right), \end{aligned} \quad (7.6)$$

où l'on a pris soin de diviser l'état entre la partie correspondant aux deux qubits côté Alice et au qubit côté Bob. On voit que les états côté Bob sont très proches de l'état inconnu que l'on cherchait à transmettre. On peut le réécrire

$$\begin{aligned} |\psi_2\rangle &= \frac{1}{2}|00\rangle|\varphi\rangle + \frac{1}{2}|01\rangle(M_{\text{not}}|\varphi\rangle) \\ &\quad + \frac{1}{2}|10\rangle(M_{\text{flip}}|\varphi\rangle) + \frac{1}{2}|11\rangle(M_{\text{flip}} \circ M_{\text{not}}|\varphi\rangle). \end{aligned} \quad (7.7)$$

La dernière étape consiste à transmettre les deux bits classiques (a, b) dont dispose Alice à Bob et à ce que ce dernier applique précisément les inverses des unitaires qui leur sont associés pour récupérer l'état $|\varphi\rangle$. Bob applique donc $M_{\text{not}}^b \circ M_{\text{flip}}^a$ à son qubit physique. Ce protocole permet ainsi de « transporter » un état quantique entre Alice et Bob en n'utilisant que la transmission de deux bits classiques d'information et des opérations unitaires effectuées localement. Insistons ici sur le fait que l'utilisation d'un canal classique est incontournable : c'est la transmission des deux bits classiques qui permet à Bob de savoir quelle opération unitaire appliquer pour préparer son qubit dans l'état $|\varphi\rangle$. C'est pour cette raison que ce protocole ne permet pas la communication supraluminique et n'est donc pas incompatible avec la relativité restreinte. Nous reviendrons dans le chapitre 11 et dans le chapitre 13, volume II, sur la compatibilité entre la relativité restreinte et la théorie quantique. Nous verrons que, de manière tout à fait remarquable, elle est en fait optimale : la théorie quantique est juste calibrée pour ne jamais violer la causalité einsteinienne³ tout en permettant l'émergence de corrélations non locales plus fortes que n'importe quelles corrélations classiques. Remarquons enfin que le théorème de non-clonage d'un état quantique implique que, quel que soit le protocole mis en œuvre, l'état originel est détruit avant que la transmission ne soit complète.

3. C'est-à-dire l'existence d'une vitesse de propagation limite à l'information classique.

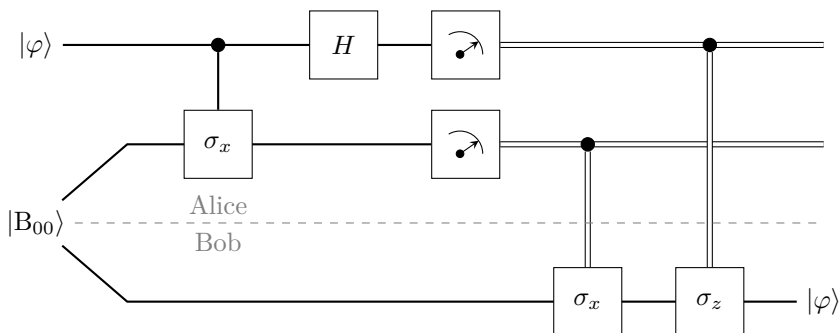


FIGURE 7.6 – Circuit quantique du protocole de téléportation d'un état quantique inconnu d'Alice à Bob. Il faut pour cela qu'ils partagent initialement la paire maximalement intriquée $|B_{00}\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$, et qu'Alice transmette le résultat de mesure effectué à Bob de manière classique, pour qu'il puisse appliquer les bons opérateurs à son qubit. Un canal de communication classique est donc utilisé pour transmettre deux bits classiques.

7.2.3 Échange d'intrication

Le protocole de téléportation quantique permet de faire plus que de mettre le qubit de Bob dans l'état $|\varphi\rangle$: il transfère toutes les propriétés que cet état initial inconnu possédait au qubit de Bob. En effet, si on remplace les nombres complexes α et β de la section précédente par des états quantiques $|\alpha\rangle$ et $|\beta\rangle$ correspondant à un 4^e qubit intriqué avec celui que l'on cherche à transmettre, on peut appliquer le même protocole et voir que l'état final est désormais le même état intriqué qu'initialement, mais que l'une des parties est désormais du côté de Bob. Cette propriété d'*échange d'intrication* permet également de montrer que l'on peut téléporter de la même manière un état mélange quelconque, puisque cet état mélange peut toujours être vu comme provenant d'un état intriqué ayant subi une mesure projective (section 5.3.1).

Puisqu'il est possible de transférer de l'intrication entre des parties éloignées, on peut pousser l'idée plus loin et imaginer un réseau de communication intriqué, à la manière d'un réseau téléphonique, en passant par un opérateur central avec lequel tous les participants partageraient des qubits intriqués [360, 63]. Cette idée simple est représentée sur la figure 7.7. Chaque utilisateur du système partage une paire intriquée dans un état $|\Psi\rangle$ avec l'opérateur central. Si deux utilisateurs veulent être mis en relation pour pouvoir s'échanger un état quantique sans avoir été en contact précédemment, l'opérateur central peut permettre cela. Plus précisément, si on part des deux états quantiques $|\Psi_{33'}\rangle$ et $|\Psi_{44'}\rangle$, où la notation $'$ dénote le qubit situé au niveau de l'opérateur central, il suffit d'effectuer une opération de téléportation quantique utilisant la paire 44' et le qubit 3' comme état à téléporter, pour créer l'état $|\Psi_{34}\rangle$ (figure 7.7).

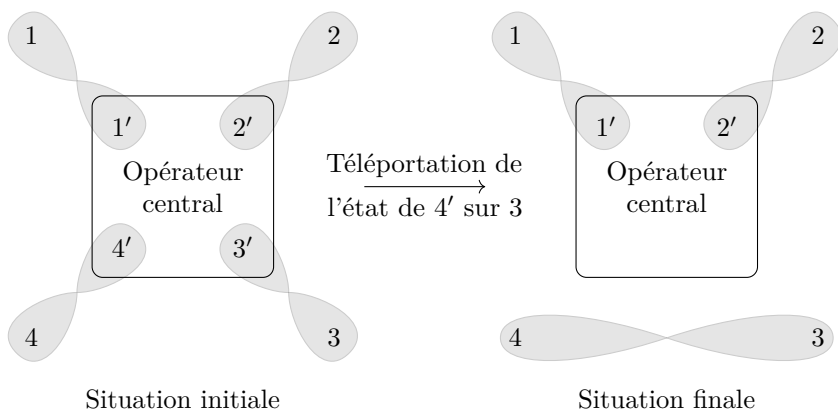


FIGURE 7.7 – Schéma d'un réseau de communication quantique utilisant l'échange d'intrication. Tous les utilisateurs du réseau partagent une paire intriquée dans l'état $|\Psi\rangle$ avec un opérateur central. Si deux utilisateurs (ici, ceux possédant les qubits 3 et 4) souhaitent partager une paire intriquée, par exemple pour envoyer un état quantique de l'un à l'autre, ou pour générer une clé de manière sécurisée entre eux, il suffit à l'opérateur central de téléporter l'état du qubit 3' sur le qubit 4 en utilisant la paire 44'.

De la même manière, on peut créer des états intriqués entre plus de deux utilisateurs, par exemple un état tripartite de la forme $1/\sqrt{2}(|000\rangle + |111\rangle)$.

Insistons sur le fait que tout ceci est local : bien que les utilisateurs ne se soient jamais rencontrés, ils ont dû interagir avec le central pour créer les paires maximales intriquées qu'ils partagent avec lui. Comme pour le protocole de distribution d'intrication, cela se fait au moyen d'opérations quantiques locales même si, à la fin, il faut transmettre le porteur physique d'un état quantique. Il n'y a aucune non-localité cachée dans la construction d'un réseau de communication quantique.

Développer un tel système présente toutefois un avantage, car une fois que le central et les utilisateurs disposent de paires intriquées, il n'est plus nécessaire que ces derniers se rencontrent pour pouvoir partager des paires d'états intriqués. De plus, on peut s'en servir pour permettre un partage de clés quantiques entre deux utilisateurs ne s'étant jamais rencontrés [53], l'opérateur central jouant le rôle d'autorité de certification pour identifier au préalable les deux utilisateurs. En effet, dès qu'Alice et Bob possèdent une paire intriquée dans l'état $|\Psi\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle) = 1/\sqrt{2}(|++\rangle + |--\rangle)$, il suffit que chacun d'eux fasse une mesure dans la base $\mathcal{B}_0 = \{|0\rangle, |1\rangle\}$ ou $\mathcal{B}_1 = \{|+\rangle, |-\rangle\}$, choisie aléatoirement, pour que l'on retrouve un résultat très proche de celui vu dans BB84. On effectue plusieurs fois ce protocole, et Alice et Bob communiquent publiquement par un canal classique leur liste

de choix de bases. Lorsqu'ils ont choisi la même base, ils ont nécessairement obtenu le même résultat s'ils n'ont pas été espionnés et ce résultat est inconnu de tout autre qu'eux. Ils peuvent donc détecter un éventuel espion, et s'ils n'ont pas été espionnés se servir de leur liste de résultats comme d'une clé pour chiffrer leur communication. On peut donc établir une communication sécurisée quantiquement à grande échelle, entre personnes qui n'avaient pas prévu de devoir communiquer entre elles.

Néanmoins, on se rend bien compte que cette architecture nécessite le partage d'un très grand nombre de paires intriquées entre les utilisateurs et l'opérateur central si l'on veut pouvoir communiquer de manière rapide. De plus, il faut imaginer que les états intriqués entre utilisateur et opérateur aient un temps de vie suffisamment long pour pouvoir être utiles lorsqu'on en a besoin. Or, les temps de vie de telles paires intriquées sont encore très courts expérimentalement (chapitres 9 et 10). Le développement d'un tel réseau de communication quantique à grande échelle relève encore de la science-fiction, mais des réalisations expérimentales d'échange d'intrication ont été effectuées en laboratoire à la fin du siècle dernier [294] avant d'être déployées sur des échelles de l'ordre du kilomètre en 2004 [313] pour atteindre 10 puis 100 km en 2017⁴ [353, 342]. La génération d'intrication entre deux mémoires quantiques à l'état solide a également été démontrée récemment [225]. Enfin l'échange d'intrication est utilisé dans la dernière génération d'expériences de test des inégalités de Bell (chapitre 11).

Pour conclure cette présentation des protocoles de base mentionnons qu'ils ont tous été mis en œuvre expérimentalement avec des photons, qu'il s'agisse d'expériences menées avec des photons uniques ou d'expériences utilisant des états cohérents. Les développements les plus spectaculaires sont sans nul doute la démonstration d'un protocole quantique de distribution de clé sur une distance intercontinentale au moyen d'un satellite et l'émergence de produits commerciaux destinés à déployer des solutions de cryptographie quantique. Une revue très personnelle d'Anton Zeilinger [386] constitue un point de départ très intéressant pour aller plus loin sur ces idées.

7.2.4 Ressources et protocoles

Nous allons maintenant revenir sur l'ensemble des protocoles élémentaires en les exprimant dans le *langage des ressources*. Ce que nous appelons une ressource pour la communication est une collection de moyens physiques utilisés par plusieurs opérateurs pour leur permettre de communiquer, classiquement ou quantiquement. Nous allons utiliser ici la notation introduite par I. Devetak *et al.* [131, 132, 133] et reprise dans le livre de M. Wilde [379, Chapitre 26] : une

4. Le lecteur attentif aura noté une corrélation entre le changement de position du premier auteur de la première publication au fil du temps et les progrès accomplis dans le domaine. Cette corrélation qui n'a rien de quantique et qui ne fait pas causalité illustre qu'investir opportunément et franchement sur de jeunes talents temporairement expatriés est une stratégie pertinente.

TABLE 7.1 – Ressources classiques et quantiques élémentaires utilisées en théorie de la communication. La notation $n[R]$, où $n \in \mathbb{N}^*$, signifie que la ressource $[R]$ est utilisée n fois.

Ressource	Symbole
Un bit classique partagé	$[cc]$
Une paire maximale ment intriquée partagée	$[qq]$
Canal classique parfait	$[c \rightarrow c]$
Canal quantique parfait	$[q \rightarrow q]$

ressource sera notée $[R]$ et si elle permet d'accomplir tout ce que la ressource $[R']$ est censée faire, nous noterons cela $[R] \geq [R']$. Diverses terminologies sont employées dans la littérature : on dit que $[R]$ permet de simuler $[R']$ ou encore que l'on peut convertir $[R]$ en $[R']$ ou que l'on peut consommer $[R]$ pour obtenir $[R']$. Lorsque $[R] \geq [R']$ et $[R'] \geq [R]$ les ressources $[R]$ et $[R']$ sont équivalentes, ce que nous noterons $[R'] \equiv [R]$.

Comme nous nous intéressons à la communication, commençons par discuter des ressources partagées entre Alice et Bob. Ainsi $[cc]$ désignera un même bit d'information partagé entre Alice et Bob, alors que $[qq]$ désignera l'analogue quantique, c'est-à-dire une paire maximale ment intriquée. De même, les canaux de transmission entre Alice et Bob seront des ressources en théorie de la communication. On notera $[c \rightarrow c]$ un canal de communication classique parfait entre Alice et Bob et $[q \rightarrow q]$ un canal de communication quantique parfait. Ces ressources de base sont ainsi listées dans le tableau 7.1.

Comme il est possible d'utiliser un bit quantique pour communiquer une information classique (en prenant des états orthogonaux), nous avons donc notre première inégalité en ressources :

$$[q \rightarrow q] \geq [c \rightarrow c]. \quad (7.8)$$

En revanche, si une ressource R ne permet pas de réaliser ce que fait une ressource R' , nous le noterons $\neg([R] \geq [R'])$. Certains théorèmes d'impossibilités (chapitre 6) peuvent ainsi s'exprimer en termes de ressources. C'est le cas des théorèmes de non-communication $\neg(n[qq] \geq [c \rightarrow c])$, $n \in \mathbb{N}^*$ (section 6.4.1) et de non-téléportation $\neg(n[c \rightarrow c] \geq [q \rightarrow q])$, $n \in \mathbb{N}^*$ (section 6.4.2). Ré-exprimons maintenant les protocoles que nous avons déjà vus en termes de ressource :

- Le protocole BB84, qui est un protocole cryptographique, nécessite d'introduire la notion de ressource privée et de ressource publique. Nous les discernons donc par un indice de sorte que $[c \rightarrow c]_{\text{priv}}$ désigne un canal de transmission privé. Le chiffre de Vernam s'exprime alors en termes de ressources par l'inégalité

$$[cc]_{\text{priv}} + [c \rightarrow c] \geq [c \rightarrow c]_{\text{priv}}. \quad (7.9)$$

Un canal de communication privé permettant la communication et la distribution d'une clé privée, nous avons $[c \rightarrow c]_{\text{priv}} \geq [cc]_{\text{priv}}$ et $[c \rightarrow c]_{\text{priv}} \geq [c \rightarrow c]$. Avec ce formalisme, le protocole BB84 qui permet la distribution d'une clé privée classique s'exprime par

$$[q \rightarrow q] + [cc] \geq [c \rightarrow c]_{\text{priv}} . \quad (7.10)$$

- La téléportation quantique exprime qu'il est possible de transmettre parfaitement un état quantique d'Alice vers Bob au moyen d'un canal de transmission classique utilisé deux fois et d'une paire maximalement intriquée partagée entre Alice et Bob :

$$[qq] + 2[c \rightarrow c] \geq [q \rightarrow q] . \quad (7.11)$$

- La distribution d'intrication utilise un canal quantique idéal pour pouvoir partager une paire maximalement intriquée entre Alice et Bob, ce qui, en termes de ressources s'écrit :

$$[q \rightarrow q] \geq [qq] . \quad (7.12)$$

Remarquons cependant qu'il n'est pas possible de communiquer quantiquement uniquement à partir d'une paire intriquée :

$$\neg ([qq] \geq [q \rightarrow q]) . \quad (7.13)$$

Procédons par l'absurde : comme $[q \rightarrow q] \geq [c \rightarrow c]$, l'inégalité $[qq] \geq [q \rightarrow q]$ entraînerait qu'une paire intriquée permette de créer un canal de communication classique et cela viole le théorème de non-communication.

- Le codage superdense permet enfin de transmettre deux bits classiques d'information à l'aide d'une paire maximalement intriquée partagée et d'un canal de transmission utilisé une seule fois. L'inégalité correspondante en termes de ressources s'écrit donc :

$$[qq] + [q \rightarrow q] \geq 2[c \rightarrow c] . \quad (7.14)$$

Les règles du calcul des ressources

Dans la discussion précédente, nous avons introduit des ressources « à l'unité » correspondant à l'utilisation unique du système physique associé. Cependant, dans la théorie de l'information classique présentée au chapitre 2, les théorèmes importants comme le théorème de Shannon sur le codage ou encore celui sur la capacité de transmission d'un canal sont obtenus à la limite asymptotique, c'est-à-dire pour un très grand nombre d'utilisations de la ressource de communication classique. Il en est de même dans le cas quantique.

Très souvent, on considérera donc des ressources dites asymptotiques. Cela signifie que l'on s'autorise à utiliser n fois la ressource $[R]$ en parallèle et donc à considérer les ressources $[R^{\otimes n}]$ pour $n \in \mathbb{N}^*$. Les inégalités de ressource $[R_1] \geq [R_2]$ ont alors un sens asymptotique : elles signifient qu'il existe un protocole permettant, à la limite $n \rightarrow +\infty$ de simuler $[R_2^{\otimes n}]$ grâce à $[R_1^{\otimes n}]$ avec une précision arbitraire. En toute rigueur, cette notion demanderait à être précisée mais dans l'optique du présent ouvrage, nous en resterons à un niveau heuristique et renvoyons le lecteur à l'article original [132]. Un des intérêts de considérer des ressources asymptotiques est de donner un sens à $x[R]$ où $x \in \mathbb{R}^+$: à la limite asymptotique $n \rightarrow +\infty$, cela correspond à la donnée de $[R^{\otimes \lfloor nx \rfloor}]$.

À partir de là, les règles du calcul des ressources sont celles que l'on attend pour des inégalités. Les premières règles correspondent respectivement à la transitivité de la relation de simulation, ainsi qu'à la possibilité d'ajouter des inégalités de ressources ou de les amplifier :

$$[R_1] \geq [R_2] \text{ et } [R_2] \geq [R_3] \implies [R_1] \geq [R_3], \quad (7.15a)$$

$$[R_1] \geq [R_2] \text{ et } [R'_1] \geq [R'_2] \implies [R_1] + [R'_1] \geq [R_2] + [R'_2], \quad (7.15b)$$

$$[R_1] \geq [R_2] \implies \forall x \in \mathbb{R}^+, x[R_1] \geq x[R_2]. \quad (7.15c)$$

On peut aussi simplifier les inégalités de ressources : soit $[R_1]$, $[R_2]$ et $[R]$ trois ressources, alors :

$$[R] + [R_1] \geq [R] + [R_2] \implies [R_1] \geq [R_2], \quad (7.16)$$

ce qui permet d'étendre le multiple d'une ressource à des nombres négatifs et donc de donner un sens à une ressource $[R] - [R']$ au moyen de l'équivalence logique :

$$[R_1] - [R_2] \geq [R'_1] - [R'_2] \iff [R_1] + [R'_2] \geq [R'_1] + [R_2]. \quad (7.17)$$

Les règles (7.15) et (7.16) vont nous permettre de discuter la question de l'optimalité des protocoles élémentaires, puis d'énoncer les résultats de la théorie de Shannon quantique sous forme d'inégalités de ressources et surtout, de comprendre intuitivement comment ils se déduisent les uns des autres.

Optimalité des protocoles

Il est naturel de se demander s'il serait possible d'avoir des résultats plus forts que les protocoles déjà étudiés. Par exemple, pourrions-nous transmettre plus de deux bits classiques au moyen d'une paire intriquée et d'un canal quantique parfait ? Ceci s'exprimerait par l'inégalité de ressources

$$[q \rightarrow q] + [qq] \geq 2r[c \rightarrow c] \quad (7.18)$$

avec $r > 1$. Si cela était vrai, alors grâce à la téléportation quantique, nous aurions l'inégalité $[qq] + 2[c \rightarrow c] + [qq] \geq 2r[c \rightarrow c]$, ce qui se réécrit formellement

$$2[qq] \geq (2r - 2)[c \rightarrow c]. \quad (7.19)$$

Cette inégalité s'interprète en disant qu'il est possible d'émuler un canal classique parfait $[c \rightarrow c]$ en utilisant $1/(r - 1)$ paires intriquées partagées entre Alice et Bob. Ceci est en contradiction avec le théorème de non-communication.

De la même manière, supposons que nous puissions transférer un bit quantique en utilisant moins de deux bits classiques, c'est-à-dire que

$$2r[c \rightarrow c] + [qq] \geq [q \rightarrow q] \quad (7.20)$$

avec $r < 1$. Dans ce cas, l'inégalité de codage superdense nous dit que $2r[c \rightarrow c] + 2[qq] \geq 2[c \rightarrow c]$ et donc $2[qq] \geq 2(1 - r)[c \rightarrow c]$, ce qui est en contradiction avec le théorème de non-communication.

Il est également impossible de téléporter plus d'un qubit en utilisant une paire intriquée et deux bits classiques. En effet, en combinant $2[c \rightarrow c] + [qq] \geq r[q \rightarrow q]$ pour $r > 1$ avec l'inégalité de codage superdense (7.14), nous obtenons $(r + 1)[qq] \geq 2(r - 1)[c \rightarrow c]$, ce qui est également en contradiction avec le théorème de non-communication.

Examinons enfin la distribution d'intrication : supposons que l'on puisse distribuer plus d'une paire maximale intriquée avec une seule utilisation d'un canal quantique parfait : $[q \rightarrow q] \geq r[qq]$ avec $r > 1$. En combinant cette inégalité avec la téléportation quantique, nous aurions $2[c \rightarrow c] \geq (r - 1)[q \rightarrow q]$, ce qui signifie qu'il existe $n \in \mathbb{N}^*$ tel que $n[c \rightarrow c] \geq [q \rightarrow q]$. Cela contredirait donc le théorème de non-téléportation. Un canal quantique parfait utilisé une fois permet donc de créer au plus une paire maximale intriquée.

Cette discussion montre que les inégalités de ressources (7.11) et (7.14) sont optimales : il faut deux bits classiques pour téléporter un bit quantique et deux bits classiques ne peuvent pas transférer plus d'un bit quantique. Un canal quantique assisté d'une paire intriquée ne peut permettre de transmettre plus de deux bits classiques.

Exhaustivité des protocoles élémentaires

Plus généralement, on peut se demander à quelle condition il est possible de réaliser physiquement un protocole de communication entre Alice et Bob caractérisé par l'inégalité de ressources

$$0 \geq C[c \rightarrow c] + Q[q \rightarrow q] + E[qq], \quad (7.21)$$

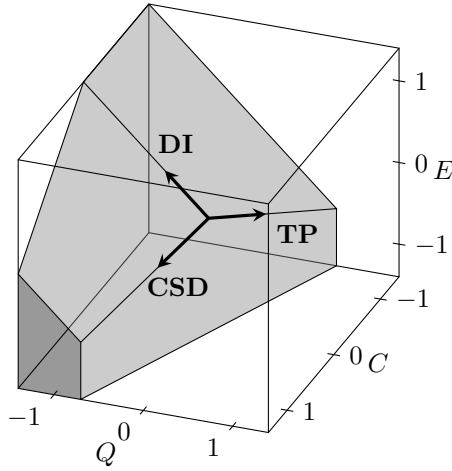


FIGURE 7.8 – Représentation dans l'espace (C, Q, E) de la zone des protocoles physiquement réalisables obtenus en combinant la téléportation **TP**, la distribution d'intrication **DI** et le codage superdense **CSD**. Une valeur négative d'une coordonnée correspond à une ressource consommée tandis qu'une valeur positive correspond à ce qui est réalisé.

où, lorsque C , Q ou E est négatif, il est passé à gauche du signe $\geq$. La réponse est que cela est possible si et seulement si [211] :

$$C + Q + E \leq 0, \quad (7.22a)$$

$$Q + E \leq 0, \quad (7.22b)$$

$$C + 2Q \leq 0. \quad (7.22c)$$

La zone correspondante est représentée sur la figure 7.8. Cela équivaut à combiner les trois protocoles fondamentaux que sont la distribution d'intrication, la téléportation quantique et le codage superdense. En d'autres termes, ces protocoles sont suffisants pour générer tous les protocoles physiquement autorisés entre Alice et Bob.

La preuve de ce résultat se fait en deux temps : d'abord on montre que combiner les trois protocoles élémentaires donne les inégalités (7.22). Ensuite, on doit démontrer la réciproque, à savoir que tout point vérifiant les inégalités (7.22) est bien réalisable physiquement en combinant les trois protocoles élémentaires. Les trois éléments de base sont les deux contraintes physiques que sont la non-communication et la non-téléportation mais il faut également utiliser une troisième contrainte qui est qu'un qubit ne peut véhiculer plus d'un bit d'information classique. Ce résultat est dû à Holevo et sera discuté dans la section 7.3.2 [379, Chapitre 8].

On voit ici tout l'intérêt de la formulation en termes de ressources : elle donne une vision synthétique de la puissance des différentes ressources élémentaires dont nous disposons. Ainsi le canal quantique $[q \rightarrow q]$ est à la fois plus puissant que le canal classique $[c \rightarrow c]$ et qu'une paire intriquée $[qq]$. Ces deux dernières ressources ne sont en revanche pas fongibles l'une en l'autre à cause des théorèmes d'impossibilité. La théorie des ressources nous permet ainsi de voir clairement les limitations à certaines conversions de ressources entre elles. En approfondissant cette démarche, nous pourrions ainsi mieux comprendre les différentes relations entre divers protocoles plus complexes et avoir un cadre unificateur pour la théorie de l'information quantique.

7.3 Contenu en information d'un qubit

Après avoir vu quelles sont les conversions possibles entre ressources quantiques et classiques élémentaires, il est nécessaire de quantifier la valeur informationnelle des ressources quantiques afin de pouvoir pleinement développer la théorie de Shannon quantique. Une manière d'aborder ce problème consiste à se demander quelles sont les différentes quantités d'information que l'on peut associer à un qubit.

Un qubit peut être vu sous deux angles (section 7.1.1), soit en tant qu'objet quantique véhiculant une information classique, soit en tant que porteur d'une information purement quantique.

Le premier angle de vue se fonde sur l'inégalité de ressources $[q \rightarrow q] \geq [c \rightarrow c]$ qui stipule qu'un canal quantique parfait permet de communiquer une information classique. On peut donc imaginer un protocole dans lequel Alice gère une source d'information classique qu'elle encode au moyen d'états d'un qubit qui sont ensuite récupérés sans être altérés par Bob. Alice étant en charge de la préparation, la question qui se pose est celle de la quantité d'information nécessaire pour préparer ce qui est émis par la source, soit évaluer une *information de préparation* I_{prep} . De son côté, Bob se demande quelle est la quantité d'information véhiculée par les qubits qu'il reçoit, c'est-à-dire quelle est l'information classique maximale qu'il peut obtenir au moyen de mesures effectuées sur ce qu'il reçoit, soit évaluer une *information accessible* I_{acc} . Enfin, en considérant le duo Alice & Bob, on peut se demander quelle est l'information classique qui manque à Bob pour reconstruire sans ambiguïté, c'est-à-dire purifier, ce qu'Alice lui a envoyé. On cherche donc à calculer une *information manquante* I_{manq} . Nous voyons donc apparaître trois notions d'information dont nous allons montrer qu'elles doivent être différenciées en physique quantique alors qu'elles se confondent en classique.

Le second angle de vue se propose de considérer un qubit comme vecteur d'une information purement quantique. Cela pose la question de la compression d'une source quantique : quelle est l'utilisation minimale d'un canal quantique parfait qu'il faut mobiliser pour transférer l'information quantique émise par

cette source d'Alice à Bob. La réponse, qui généralise le théorème de codage de Shannon, est donnée par le théorème de Schumacher (section 7.3.3).

7.3.1 Bits classiques *versus* qubits

Dans le cas classique, les notions d'information de préparation, manquante, et accessible coïncident avec l'entropie de Shannon. Supposons qu'Alice dispose d'une source classique émettant le message i avec la probabilité p_i . La source classique associée à Alice nécessite $I_{\text{prep}} = S[p_i]$ bits d'information. Bob de son côté accède juste à un ensemble de messages classiques dont il peut faire la statistique. Sans surprise, il aura besoin de $I_{\text{acc}} = S[p_i]$ bits pour encoder ce qu'il a reçu. Enfin, l'information qu'Alice a besoin de communiquer à Bob pour lui permettre de décoder ce qu'elle a envoyé vaut $I_{\text{manq}} = S[p_i]$. Ainsi, les trois notions d'information sont confondues, la raison principale étant que les messages classiques peuvent être discernés sans erreur dès lors qu'ils sont différents.

Dans le cas quantique, la situation est radicalement différente. Alice encode le message i au moyen d'un état quantique appartenant à l'ensemble statistique décrit par l'opérateur densité ρ_i . Cela permet de décrire le cas où le processus d'encodage par Alice n'est pas parfait et ne produit pas toujours le même état pur. Pour garder une notation compacte, on notera A l'ensemble décrit par les opérateurs densité ρ_i avec leurs probabilités d'émission p_i .

L'information de préparation est toujours l'entropie de Shannon associée à la distribution de probabilité $p_i : I_{\text{prep}} = S[p_i]$ reflétant les choix d'Alice. Cependant, du fait de l'indiscernabilité des états quantiques non orthogonaux, les deux autres notions sont bien différentes.

L'information manquante est l'information que doit transmettre Alice à Bob pour supprimer toute incertitude statistique quant aux états qu'il reçoit, c'est-à-dire pour décrire ce qu'il reçoit en termes d'états purs. Dans le cas où les messages ρ_i sont des états purs, nous savons (section 5.3.3) que l'entropie de von Neumann du mélange $\rho_A = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ quantifie la quantité minimale d'information classique qu'Alice doit transmettre à Bob pour que celui-ci explicite le mélange qu'il reçoit sous forme d'un ensemble statistique d'états purs $I_{\text{manq}} = S_{\text{vn}}[\rho_A]$. Lorsque les états $|\psi_i\rangle$ ne sont pas mutuellement orthogonaux, l'information manquante est donc strictement inférieure à l'information de préparation.

Dans le cas où Alice utilise des messages encodés par des opérateurs densité ρ_i qui ne correspondent pas à des états purs, l'information manquante est l'information nécessaire pour que Bob puisse décrire ce qu'Alice émet comme un ensemble statistique d'états purs. Là encore, l'information nécessaire pour purifier totalement ce que Bob reçoit est l'entropie de von Neumann de l'opérateur densité représentant tout ce qui est émis côté Alice, à savoir

$\rho_A = \sum_i p_i \rho_i$. Or $S_{\text{vn}}[\rho_A]$ vérifie les inégalités suivantes :

$$\sum_i p_i S_{\text{vn}}[\rho_i] \leq S_{\text{vn}} \left[\sum_i p_i \rho_i \right] \leq S[p_i] + \sum_i p_i S_{\text{vn}}[\rho_i]. \quad (7.23)$$

La borne inférieure reflète la concavité de l'entropie de von Neumann que nous démontrerons à la section 7.4. Pour obtenir la borne supérieure, nous utilisons la décomposition de Schmidt de chaque $\rho_i = \sum_j p_j^{(i)} |\psi_j^{(i)}\rangle \langle \psi_j^{(i)}|$ où $(|\psi_j^{(i)}\rangle)$ forme une base orthonormée de $\mathcal{H}$ et $\sum_j p_j^{(i)} = 1$. Nous avons alors $\rho_A = \sum_{i,j} p_i p_j^{(i)} |\psi_j^{(i)}\rangle \langle \psi_j^{(i)}|$ et donc $S_{\text{vn}}[\rho_A] \leq S[(p_i p_j^{(i)})_{i,j}]$. L'équation (2.19) permet de calculer cette quantité et, en identifiant $S_{\text{vn}}[\rho_i] = -\sum_j p_j^{(i)} \log(p_j^{(i)})$, nous obtenons la borne supérieure dans (7.23).

Considérons maintenant le point de vue de Bob. À quelle information peut-il accéder par des mesures ? Bob dispose d'un laboratoire dans lequel il peut effectuer n'importe quelle expérience et donc potentiellement obtenir les résultats de n'importe quelle mesure généralisée qu'il peut effectuer sur les systèmes quantiques reçus d'Alice. Tel Sherlock Holmes, Bob cherche à remonter à partir de ses résultats de mesure aux messages émis par Alice. L'information accessible est donc définie en termes du gain d'information dans un processus d'inférence bayésienne. Si R désigne la source constituée par les résultats de mesure obtenus par Bob lorsqu'il considère un protocole de mesure donné, alors l'information accessible n'est autre que

$$I_{\text{acc}}[A] = \max_R (I[R, A]) . \quad (7.24)$$

où $I[R, A]$ désigne l'information mutuelle entre les résultats expérimentaux obtenus par Bob et les choix d'Alice A qui sont caractérisés par la distribution de probabilité p_i . La maximisation au niveau de Bob correspond au choix d'un protocole optimal. Comme on doit maximiser une expression faisant apparaître une fonction log sur toutes les mesures généralisées possibles, il n'existe pas, en général, de formule pour cette quantité. Les seuls résultats généraux qui existent sont des bornes. L'inégalité $I[A, R] \leq S[A]$ entraîne que $I_{\text{acc}} \leq I_{\text{prep}}$ mais c'est une borne bien mauvaise. Fort heureusement, une borne bien meilleure a été obtenue et constitue un des premiers résultats importants de la théorie de l'information quantique.

7.3.2 La borne de Holevo

Le théorème et ses conséquences

Le *théorème de Holevo* [204] stipule que l'information accessible par Bob lorsque Alice émet les opérateurs densités ρ_i avec les probabilités p_i (on note $A = [(\rho_i, p_i)]$) est bornée supérieurement par la quantité de Holevo

$$\chi[A] = S_{\text{vn}} \left[\sum_i p_i \rho_i \right] - \sum_i p_i S_{\text{vn}}[\rho_i]. \quad (7.25)$$

Ce résultat entraîne que, sans autre ressource mobilisée, un qubit ne peut véhiculer plus d'un bit d'information classique. En effet, $\rho_A = \sum_i p_i \rho_i$ étant un opérateur densité pour un qubit, $S_{\text{vn}}[\rho_A] \leq 1$. En conséquence, l'information accessible pour un qubit est au plus égale à un bit. En termes de ressources :

$$\forall n \geq 2, \neg ([q \rightarrow q] \geq n [c \rightarrow c]) , \quad (7.26)$$

ce qui permet de démontrer les inégalités (7.22). Cette borne montre aussi que, si Alice émet des états purs, alors l'information accessible est au plus égale à l'entropie de von Neumann de ρ_A . Ce résultat est compatible avec l'interprétation de cette quantité donnée dans la section 5.3.3 : en représentant ρ_A comme provenant d'une purification partagée avec Carole, nous avons montré que $S_{\text{vn}}[\rho_A]$ représente la quantité d'information minimale que Carole doit transmettre à Alice pour décrire ρ_A sous la forme d'un mélange d'états purs qui, avec ce choix, sont parfaitement discernables. Or, intuitivement, c'est cette forme qui permet la meilleure communication de messages au moyen d'états purs sous réserve que le canal qui relie Alice à Bob soit parfait et que Bob mesure dans la bonne base.

Finalement, en combinant la borne de Holevo avec les inégalités (7.23), nous voyons que l'information accessible, la quantité de Holevo, l'information manquante et l'information de préparation s'ordonnent selon

$$I_{\text{acc}} \leq \chi[A] \leq I_{\text{manq}} = S_{\text{vn}}[\rho_A] \leq I_{\text{prep}} = S[(p_i)] . \quad (7.27)$$

Lorsque Alice émet des états purs parfaitement discernables, les trois quantités coïncident mais en général les inégalités seront strictes, ce qui reflète l'impossibilité de discerner deux états différents à coup sûr.

✂ Preuve

Pour prouver la borne de Holevo, supposons que Bob effectue une mesure généralisée décrite par des opérateurs M_y qui vérifient $\sum_y M_y^\dagger M_y = \mathbb{1}$. Nous allons introduire deux systèmes auxiliaires qui gardent trace du processus de choix par Alice d'un message x et de la mesure utilisée par Bob. L'état du système complet avant mesure par Bob est

$$\rho_{\text{avant}} = \sum_i p_x |x\rangle\langle x| \otimes \rho_i \otimes |y_0\rangle\langle y_0| \quad (7.28)$$

où les états $|x\rangle$ mutuellement orthogonaux auxiliaires encodent les messages classiques qu'Alice souhaite transmettre au moyen des ρ_x . De même, $|y_0\rangle$ désigne l'état avant mesure du système auxiliaire chargé de garder trace des résultats des mesures effectuées par Bob. La base de mesure utilisée par Bob est formée d'états $|y\rangle$ indexés par y . L'état après transmission à travers le canal quantique puis mesure par Bob est alors donné par

$$\rho_{\text{après}} = \sum_{x,y} p_x |x\rangle\langle x| \otimes M_y \rho'_x M_y^\dagger \otimes |y\rangle\langle y| , \quad (7.29)$$

où ρ'_x désigne ce qui ressort du canal de communication quantique reliant Alice et Bob lorsque ρ_x y a été injecté. Introduisons l'opérateur densité associé au résultat de mesure y et conditionné au choix x par Alice

$$\rho'(y|x) = \frac{M_y \rho'_x M_y^\dagger}{p(y|x)}, \quad (7.30)$$

où la probabilité conditionnelle d'obtenir le résultat y sachant qu'Alice a choisi d'émettre ρ_x est $p(y|x) = \text{tr}(M_y \rho'_x M_y^\dagger)$. Avec cette notation

$$\rho_{\text{après}} = \sum_{x,y} p_x p(y|x) |x\rangle\langle x| \otimes \rho'(y|x) \otimes |y\rangle\langle y| \quad (7.31)$$

décrit le mélange statistique des opérateurs densités $\rho'(y|x)$ pondérés par la probabilité $p_x p(y|x) = p(x, y)$. En traçant sur les degrés de liberté quantiques qui ont traversé le canal quantique, nous obtenons donc un opérateur densité qui véhicule les corrélations classiques établies entre Alice (X degrés de liberté associés aux états auxiliaires $|x\rangle$) et Bob (Y degrés de liberté associés aux états $|y\rangle$ codant le résultat de la mesure) :

$$\rho_{X,Y}^{(\text{après})} = \sum_{x,y} p(x, y) |x\rangle\langle x| \otimes |y\rangle\langle y|. \quad (7.32)$$

En conséquence, l'information mutuelle $I[A, B_M]$ qui quantifie les corrélations classiques entre les choix d'Alice notés A et les résultats de mesure de Bob notés⁵ B_M , s'exprime sous la forme :

$$I[A, B_M] = S_{\text{vn}}[\rho_Y^{(\text{après})}] + S_{\text{vn}}[\rho_X^{(\text{après})}] - S_{\text{vn}}[\rho_{XY}^{(\text{après})}], \quad (7.33)$$

où $\rho_Y = \text{tr}_X(\rho_{XY})$ et $\rho_X = \text{tr}_Y(\rho_{XY})$ désignent respectivement les opérateurs densité réduits pour les degrés de liberté quantiques X et Y que l'on appellera des registres quantiques associés à Alice et Bob. Cette expression relie une information mutuelle classique $I[A, B_M]$ avec une quantité qui, formellement, est définie de manière analogue à l'information mutuelle introduite à la section 2.4.3 mais en utilisant des entropies de von Neumann :

$$I_{\text{vn}}[X, Y] = S_{\text{vn}}[\rho_X] + S_{\text{vn}}[\rho_Y] - S_{\text{vn}}[\rho_{XY}]. \quad (7.34)$$

Cette quantité, appelée *information mutuelle quantique*, généralise l'information mutuelle classique. Intuitivement, c'est une mesure des corrélations quantiques et classiques entre X et Y .

L'entropie de von Neumann de l'opérateur densité ρ_{avant} est facile à calculer car le registre quantique Y de Bob est découplé. En conséquence

$$S_{\text{vn}}[\rho_{\text{avant}}] = S_{\text{vn}} \left[\sum_x p_x |x\rangle\langle x| \otimes \rho_x \right], \quad (7.35)$$

5. L'indice rappelle que Bob utilise la mesure généralisée définie par les opérateurs (M_y) .

ce qui permet de calculer l'entropie associée de manière directe (voir section 7.4) :

$$S_{\text{vn}}[\rho_{\text{avant}}] = S[(p_x)] + \sum_x p_x S_{\text{vn}}[\rho_x]. \quad (7.36)$$

Cette expression fait apparaître l'entropie de préparation qui est l'entropie des choix d'Alice $S[(p_x)]$ ainsi que la somme pondérée des entropies $S_{\text{vn}}[\rho_x]$. Les seuls systèmes corrélés étant les registres quantiques X et Y que nous désignerons par Q , on peut calculer l'information mutuelle quantique entre eux $I[(Q, \text{Aux})_{\text{avant}}]$ avant transmission à travers le canal et montrer qu'elle est égale à $\chi[A]$. Notons qu'à ce stade, les registres Y sont initialement décorrelés d'Alice et des registres X .

Après transmission à travers le canal quantique et mesure par Bob, les corrélations présentes entre les objets quantiques Q reçus par Bob et le registre Y sont donc égales à $I[A, B_M]$. Ce n'est là qu'une partie des corrélations entre Q et les registres X et Y . Intuitivement, on s'attend à ce qu'oublier un partenaire diminue les corrélations et donc, dans le cas présent :

$$I[A, B_M] \leq I[(Q, \text{Aux})_{\text{après}}]. \quad (7.37)$$

Cette propriété, évidente en classique, découle d'un résultat fondamental de la théorie de l'information quantique que, pour le moment, nous admettrons⁶. Pour conclure notre preuve de la borne de Holevo, nous devons finalement comparer deux informations mutuelles entre les qubits auxiliaires Aux et les objets quantiques transmis à travers le canal de communication quantique entre avant et après la transmission. C'est ici que se cache l'essentiel de la difficulté technique.

On peut avoir une idée du résultat dont nous aurons besoin en revenant au cas où nous avons un canal de communication classique bruité entre Alice et Bob. Nous voulons comparer la corrélation classique entre d'une part les qubits Q utilisés pour la transmission et ceux des registres Aux et d'autre part les bits reçus par Bob et le même ensemble de bits auxiliaire. Intuitivement, le bruit introduit par le canal de transmission va brouiller ces corrélations et donc, on s'attend à ce que

$$I[(Q, \text{Aux})_{\text{après}}] \leq I[(Q, \text{Aux})_{\text{avant}}]. \quad (7.38)$$

Pour le démontrer, on exprime l'information mutuelle comme une divergence de Kullback-Leibler (voir équation (2.59)) et la contractivité de cette quantité dans une dynamique stochastique (voir équation (2.41)) prouve l'inégalité ci-dessus.

De manière remarquable, le raisonnement est identique dans le cas quantique : il nous faut introduire la divergence de Kullback-Leibler quantique,

6. On peut montrer que ce résultat découle aussi de la contractivité de la divergence de Kullback-Leibler quantique à laquelle nous allons faire appel plus bas.

aussi appelée entropie relative quantique, et qui est définie formellement par ⁷

$$D[\rho\|\rho'] = \text{tr}(\rho(\log \rho - \log \rho')) . \quad (7.39)$$

Lorsque les deux opérateurs densité sont diagonaux dans la même base, cette quantité coïncide avec la divergence de Kullback-Leibler classique évaluée sur les spectres de ces deux opérateurs. Dans ce cas, cette quantité mesure l'éloignement de ρ' par rapport à ρ vus comme des sources classiques dans leur base propre commune. Comme nous le verrons dans la section 7.4, $D[\rho\|\rho'] = 0$ si et seulement si $\rho = \rho'$.

La propriété dont nous avons besoin pour conclure la preuve de la borne de Holevo est l'analogie quantique de l'équation (2.41). Pour cela, il suffit de considérer que ρ provient d'une purification avec un environnement $\mathcal{E}$, ce qui est toujours vrai, et que ρ' est obtenu en prenant la trace partielle sur celui-ci après évolution unitaire du système complet. C'est exactement ce qui définit la transmission à travers un canal quantique imparfait (section 7.1.2). De manière générale, si on note $\rho'_{1,2}$ les évolués dans le temps de deux opérateurs densité $\rho_{1,2}$, alors l'entropie relative quantique est contractante :

$$D[\rho'_1\|\rho'_2] \leq D[\rho_1\|\rho_2] . \quad (7.40)$$

Comme nous le verrons dans une prochaine section, cette propriété permet de justifier l'inégalité (7.37) que nous avons déjà utilisée. Dans le cas présent, la contractivité de la divergence de Kullback-Leibler quantique nous donne la dernière inégalité qui nous permet de compléter la preuve :

$$I[A, B_M] \leq I[(Q, \text{Aux})_{\text{après}}] \leq I[(Q, \text{Aux})_{\text{avant}}] = \chi[A] , \quad (7.41)$$

où $\chi[A]$ désigne la quantité de Holevo définie par l'équation (7.25).

7.3.3 Le théorème de codage de Schumacher

Le théorème et ses conséquences

Nous allons maintenant aborder le problème du codage quantique. Cela consiste à déterminer quelles sont les ressources quantiques qu'il faut utiliser pour transférer de la manière la plus fidèle possible l'information quantique émise par une source quantique donnée.

Le *théorème de Schumacher* [324] fournit la réponse à cette question et constitue donc l'analogie quantique de celui de Shannon. Il s'énonce de la manière suivante : pour transmettre les états quantiques émis par une source

7. Une précaution s'impose dans la définition : si l'image de l'opérateur densité ρ intersecte le noyau de ρ' , on dira que $D[\rho\|\rho'] = \infty$. Cette prescription est compatible avec l'interprétation de $D[\rho\|\rho']$ comme une mesure de l'éloignement de ρ' par rapport à ρ . En effet, si ce dernier émet des états qui ne sont jamais émis par le premier, alors il y a un éloignement maximal (infini) entre les deux. Dans les cas où elle est finie, nous verrons que $D[\rho\|\rho'] \geq 0$.

quantique décrite par un opérateur densité ρ_S , il faut mobiliser $S_{\text{vn}}[\rho_S]$ qubits par message.

De manière plus précise, il existe un dispositif d'encodage/décodage quantique qui permet de transmettre avec une excellente fidélité $N \gg 1$ états émis par la source S en utilisant $NS_{\text{vn}}[\rho_S]$ fois un canal quantique idéal. Remarquons que pour un mélange statistique d'états non orthogonaux décrit par $\rho_S = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, le fait que l'information manquante $S_{\text{vn}}[\rho_S]$ soit plus petite que l'information de préparation $S[p_j]$ montre que le protocole de compression quantique est plus efficace que le protocole naïf qui consisterait à compresser les données classiques décrivant le catalogue des états du mélange.

Ce théorème est démontré en détail dans de nombreux ouvrages de référence. L'idée de preuve donnée ci-dessous est librement adaptée de la démonstration présentée dans l'ouvrage de Mark Wilde [379, Chapitre 18]. Elle s'appuie sur une notion de typicalité quantique : lorsque l'on prend une collection de $N \gg 1$ systèmes, l'opérateur densité $\rho_S^{\otimes N}$ des N systèmes est localisé sur un sous-espace réduit de dimension environ $2^{NS_{\text{vn}}[\rho_S]}$. Ce sous-espace des états typiques possède donc une dimension exponentiellement plus petite que la dimension de l'espace des états sur lequel vit $\rho_S^{\otimes N}$ dès lors que $S_{\text{vn}}[\rho_S] < \log_2 \dim(\mathcal{H}_S)$. Notons de plus que lorsque $N \gg 1$, $\rho_S^{\otimes N}$ tend vers un opérateur densité équiréparti sur l'espace des états typiques. Ces propriétés sont au cœur de la preuve rigoureuse du théorème de Schumacher.

✂ Idée de la preuve

Pour avoir une idée de la preuve générale, il est utile d'analyser une source qui émet des qubits. Considérons une source quantique qui émet des qubits préparés à 50 % dans l'état $|0\rangle$ et à 50 % dans l'état $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Dans la base canonique, la matrice densité correspondante s'écrit :

$$\rho_S = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix} \quad (7.42)$$

et peut donc être diagonalisée dans une base orthonormée $|u_{\pm}\rangle$ avec comme valeurs propres $p_{\pm} = 1/2 \pm 1/2\sqrt{2}$, ce qui permet d'évaluer directement l'entropie de von Neumann $S_{\text{vn}}[\rho_S] = -p_+ \log_2 p_+ - p_- \log_2 p_- \simeq 0,609$. L'opérateur densité réduit peut alors être purifié (section 5.3.1) en

$$|\Psi\rangle = \sqrt{p_+} |u_+\rangle |r_+\rangle + \sqrt{p_-} |u_-\rangle |r_-\rangle, \quad (7.43)$$

où $|r_{\pm}\rangle$ sont des états ancillaires normés et mutuellement orthogonaux. Regardons maintenant un grand nombre $N \gg 1$ de copies de ρ_S qui sont donc purifiées au moyen de l'état $|\Psi\rangle^N$. Nous pouvons alors développer ce dernier sous la forme

$$|\Psi\rangle^{\otimes N} = \sum_{k=0}^N \sqrt{C_N^k p_+^k p_-^{N-k}} |\psi_{k, N-k}^{(N)}\rangle, \quad (7.44)$$

où $|\psi_{k,N-k}^{(N)}\rangle$ désigne l'état normé proportionnel au symétrisé de l'état produit tensoriel de k copies de $|u_+\rangle |r_+\rangle$ et de $N - k$ copies de $|u_-\rangle |r_-\rangle$. Pour $k \neq k'$, $\langle \psi_{k,N-k}^{(N)} | \psi_{k',N-k'}^{(N)} \rangle = 0$. L'état $|\psi\rangle^{\otimes N}$ est donc décomposé sur les états $|\psi_{k,N-k}^{(N)}\rangle$ avec des amplitudes qui sont les racines carrées des probabilités données pour la spécialisation aux chaînes binaires de l'équation (2.9). Comme nous l'avons vu à la section 2.2.3, cette probabilité est très piquée autour de $k \simeq Np_+$.

Nous allons utiliser cela pour adapter au cas quantique l'idée qui conduit au théorème de Shannon sur le codage. Pour $\eta > 0$ et $\varepsilon > 0$, considérons les séquences typiques telles que les inégalités (2.14) soient vérifiées. Elles définissent ainsi une partition de l'espace étendu $\mathcal{H}_{\text{et}}$ engendré par les $|\psi_{k,N-k}^{(N)}\rangle$ ($k = 0, \dots, N$) en un espace typique et un espace atypique : $\mathcal{H}_{\text{et}} = \mathcal{H}_{\text{typ}} \oplus \mathcal{H}_{\text{atyp}}$. Le poids total de la projection orthogonale de $|\psi\rangle^{\otimes N}$ sur $\mathcal{H}_{\text{atyp}}$ est alors inférieur à ε . L'idée consiste à abandonner cette partie et à se focaliser uniquement sur la projection sur l'espace $\mathcal{H}_{\text{typ}}$. Tout le problème consiste maintenant à déterminer quelles sont les ressources de communication quantiques qu'il faut mobiliser pour faire passer les états quantiques typiques d'Alice à Bob tout en préservant leur intrication avec les qubits ancillaires de sorte que Bob se retrouve en possession d'un opérateur densité réduit arbitrairement proche de ρ_S .

Pour cela, regardons quelle est la structure de la projection orthogonale de $|\psi\rangle^{\otimes N}$ sur l'espace $\mathcal{H}_{\text{typ}}$. Cet état est constitué d'un nombre de termes deux à deux orthogonaux qui n'est autre que le nombre de séquences typiques $N(\varepsilon, \eta)$ comportant k copies $|u_+\rangle |r_+\rangle$ et $N - k$ copies de $|u_-\rangle |r_-\rangle$ avec des coefficients qui sont proches de $1/\sqrt{N(\varepsilon, \eta)}$ pour assurer la normalisation. On voit donc que la dimension de l'espace engendré par les états typiques à transférer est $N(\varepsilon, \eta) \simeq 2^{NH_2(p_+)}$ où $H_2(p_+) = -p_+ \log_2 p_+ - p_- \log_2 p_-$ qui n'est autre que l'entropie de von Neumann $S_{\text{vn}}[\rho_S]$. Nous venons donc de voir qu'il nous faut transférer $S_{\text{vn}}[\rho_S]$ qubits de manière parfaite.

Cette discussion, qui s'appuie sur les idées sous-jacentes au théorème de codage de Shannon, suggère que l'entropie de von Neumann possède une interprétation opérationnelle purement quantique : c'est le nombre minimal de fois qu'il faut utiliser un canal quantique parfait pour transmettre l'information quantique émise par une source.

D'un point de vue constructif, le protocole de compression peut être mis en œuvre par le circuit représenté sur la figure 7.9. Le point essentiel pour y arriver est l'étape de mesure projective de typicalité : il faut construire un dispositif de mesure qui permette d'extraire la projection orthogonale sur l'espace des états typiques de manière cohérente. En effet, on souhaite préserver toutes les cohérences dans ce sous-espace. Bien que cela soit toujours possible en théorie, la réalisation pratique peut s'avérer être un défi considérable. Remarquons aussi que dans ce circuit, la mesure correspondant au sous-espace typique est une mesure collective effectuée sur $\mathcal{H}_A^{\otimes N}$ et l'étape $\mathcal{E}_2$ est une transformation unitaire qui agit globalement sur l'espace des états typiques.

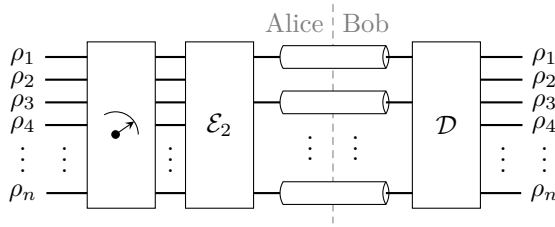


FIGURE 7.9 – Schéma d'un protocole de compression quantique : Alice dispose d'une source quantique décrite par l'opérateur densité ρ_S . Pour une émission de $N \gg 1$ messages, on commence par effectuer une mesure projective de typicalité. Seule la composante typique est encodée, *via* une isométrie appropriée, dans un espace de dimension $2^{NS_{\text{vn}}[\rho_S]}$ et transmise à travers $NS_{\text{vn}}[\rho_S]$ canaux quantiques idéaux.

Ces deux étapes génèrent donc inévitablement de l'intrication entre les N copies. Nous reviendrons sur ce point important dans la section 7.6.5.

7.4 Propriétés de l'entropie de von Neumann

Nous comprenons bien dorénavant l'importance toute particulière de la notion d'entropie de von Neumann pour l'information quantique. Parallèlement à la section 2.3, nous allons ici déduire les relations vérifiées par l'entropie de von Neumann lors de l'extraction d'information, lors de la concaténation de messages issus de sources différentes où les corrélations jouent un rôle primordial et enfin lors d'un mélange aléatoire de sources.

Extraction d'information L'extraction d'information quantique consiste à supposer que l'opérateur densité se décompose sous la forme d'un mélange statistique d'opérateurs densité réduits qui ont des supports sur des sous-espaces orthogonaux. En termes physiques, il est possible de les discerner à coup sûr. Soit $\rho = \sum_i p_i \rho_i$ une telle décomposition, nous avons alors :

$$S_{\text{vn}}[\rho] = - \sum_i p_i \text{tr}(\rho_i \log \rho) = S[p_i] + \sum_i p_i S_{\text{vn}}[\rho_i]. \quad (7.45)$$

Démonstration. Ceci se démontre en utilisant $\text{tr}(\rho_i \log \rho) = \text{tr}(\rho_i \log(p_i \rho_i))$ venant de l'orthogonalité des supports des ρ_i . On obtient alors $\text{tr}(\rho_i \log \rho) = \log p_i + \text{tr}(\rho_i \log \rho_i)$ ce qui démontre l'équation (7.45), l'analogie quantique de l'équation (2.19). ■

Sous-additivité Soit ρ l'opérateur densité d'une source composée décrivant un ensemble statistique d'états pour un système partagé entre Alice et Bob. En notant ρ_A et ρ_B les opérateurs densités réduits relatifs à Alice

et Bob, nous avons alors la propriété de sous-additivité de l'entropie de von Neuman :

$$S_{\text{vn}}[\rho] \leq S_{\text{vn}}[\rho_A] + S_{\text{vn}}[\rho_B]. \quad (7.46)$$

Bien que la sous-additivité de l'entropie de von Neumann soit une propriété qu'elle partage avec l'entropie de Shannon, le cas quantique nous réserve une surprise intéressante. En effet, nous savons que tout mélange statistique peut être purifié (section 5.3.1). Ainsi, si Alice et Bob partagent un mélange statistique décrit par un opérateur densité ρ_{AB} qui n'est pas un projecteur, celui-ci provient d'un état pur partagé avec un environnement E . Dans ce cas, en partitionnant ce système tripartite en deux de toutes les manières possibles, la sous-additivité implique les deux inégalités $S_{\text{vn}}[\rho_A] + S_{\text{vn}}[\rho_E] \geq S_{\text{vn}}[\rho_{AE}] = S_{\text{vn}}[\rho_B]$ et $S_{\text{vn}}[\rho_B] + S_{\text{vn}}[\rho_E] \geq S_{\text{vn}}[\rho_{BE}] = S_{\text{vn}}[\rho_A]$. Comme $S_{\text{vn}}[\rho_E] = S_{\text{vn}}[\rho_{AB}]$, nous obtenons l'inégalité dite d'Araki-Lieb :

$$S_{\text{vn}}[\rho_{AB}] \geq |S_{\text{vn}}[\rho_A] - S_{\text{vn}}[\rho_B]|. \quad (7.47)$$

Elle diffère de la borne inférieure (2.52) valable dans le cas classique. En fait, si Alice (A) et Bob (B) partagent un état pur, la borne d'Araki-Lieb est nulle, en accord avec le fait que $S_{\text{vn}}[\rho_{AB}] = 0$ dans ce cas. Dans cette situation, l'entropie de von Neumann de l'ensemble est plus petite que celle de chacune des parties, ce qui n'est jamais le cas pour l'entropie de Shannon.

La borne supérieure de l'information mutuelle quantique n'est donc pas la même que celle pour l'information mutuelle classique. Elle vaut $2S_{\text{vn}}[A]$ lorsque Alice et Bob partagent un état quantique pur, alors qu'elle ne dépassera jamais $\min(S[A], S[B])$ si Alice et Bob partagent des bits classiques. La borne inférieure d'Araki-Lieb exprime le fait que dans la théorie quantique, le tout peut être moins désordonné que chacune des parties alors que cela n'est jamais possible classiquement.

Démonstration. Pour prouver (7.46), nous devons nous appuyer sur un analogue quantique de l'inégalité (2.27). Plus précisément, si ρ et ρ' sont deux opérateurs densité, alors

$$\text{tr}(\rho \log \rho) \geq \text{tr}(\rho \log \rho'). \quad (7.48)$$

La preuve de (7.48) est plus complexe que dans le cas classique car, *a priori*, les deux opérateurs densité ne sont pas diagonalisables simultanément. Introduisons donc $(|\lambda\rangle)$ et $(|\mu\rangle)$ les bases propres de ρ et ρ' respectivement. En notant p_λ et p'_μ les valeurs propres associées, nous avons :

$$\text{tr}(\rho \log \rho') = \sum_{\mu} \langle \mu | \rho | \mu \rangle \log p_{\mu} = \sum_{\lambda, \mu} \langle \mu | \lambda \rangle \langle \lambda | \mu \rangle p_{\lambda} \log p'_{\mu}, \quad (7.49a)$$

$$\text{tr}(\rho \log \rho) = \sum_{\lambda} p_{\lambda} \log p_{\lambda}. \quad (7.49b)$$

En combinant ces résultats, et en utilisant $1 = \sum_{\mu} |\langle \mu | \lambda \rangle|^2$, nous obtenons

$$\mathrm{tr}(\rho \log \rho' - \rho \log \rho) = \sum_{\lambda, \mu} p_{\lambda} |\langle \mu | \lambda \rangle|^2 \log \left(\frac{p'_{\mu}}{p_{\lambda}} \right). \quad (7.50)$$

Comme dans le cas classique, on utilise alors la concavité du logarithme pour borner $\log \left(\frac{p'_{\mu}}{p_{\lambda}} \right) \leq \frac{p'_{\mu}}{p_{\lambda}} - 1$, ce qui donne :

$$\mathrm{tr}(\rho \log \rho' - \rho \log \rho) \leq \sum_{\lambda, \mu} |\langle \mu | \lambda \rangle|^2 (p'_{\mu} - p_{\lambda}). \quad (7.51)$$

Comme $\sum_{\lambda, \mu} |\langle \mu | \lambda \rangle|^2 p_{\lambda} = \sum_{\lambda, \mu} |\langle \mu | \lambda \rangle|^2 p'_{\mu} = 1$, cette inégalité montre que $\mathrm{tr}(\rho \log \rho' - \rho \log \rho) \leq 0$. Finalement, la borne (7.46) s'obtient en appliquant l'inégalité (7.48) avec $\rho' = \rho_A \otimes \rho_B$. ■

Concavité Le multiplexage permet de construire une source ρ , mélange statistique de deux autres $\rho = p_1 \rho_1 + p_2 \rho_2$. Nous avons alors l'inégalité :

$$S_{\mathrm{vn}}[p_1 \rho_1 + p_2 \rho_2] \geq p_1 S_{\mathrm{vn}}[\rho_1] + p_2 S_{\mathrm{vn}}[\rho_2]. \quad (7.52)$$

Démonstration. Cette inégalité n'est autre que la concavité de l'entropie de von Neumann. Elle découle directement de l'inégalité (7.48) au couple (ρ, ρ_1) et (ρ, ρ_1) avec $\rho = p_1 \rho_1 + p_2 \rho_2$. ■

Divergence de Kullback-Leibler Comme dans le cas classique, pour démontrer la sous-additivité (7.52), nous avons dû faire appel à la quantité définie par l'équation (7.39) qui généralise la divergence de Kullback-Leibler au cas quantique. Nous avons montré que cette quantité est positive (équation (7.51)) et il est évident qu'elle est additive : $D[\rho_1 \otimes \rho_2 \| \rho'_1 \otimes \rho'_2] = D[\rho_1 \| \rho'_1] + D[\rho_2 \| \rho'_2]$.

Nous allons maintenant chercher à quelle condition $D[\rho \| \rho'] = 0$. En utilisant la concavité du logarithme, nous voyons que

$$\log \left(\sum_{\mu} |\langle \mu | \lambda \rangle|^2 \rho'_{\mu} \right) \geq \sum_{\mu} |\langle \mu | \lambda \rangle|^2 \log(\rho'_{\mu}).$$

L'équation (7.45) nous donne un minorant de $D[\rho \| \rho']$ par une divergence de Kullback-Leibler classique :

$$D[(\rho_{\mu})_{\mu} \| (q_{\mu})_{\mu}] \leq D[\rho \| \rho'] \quad (7.53)$$

avec $q_{\mu} = \sum_{\nu} |\langle \mu | \lambda \nu \rangle|^2 \rho'_{\nu}$. En conséquence, si $D[\rho \| \rho'] = 0$, cela signifie que les distributions de probabilité $(p_{\mu})_{\mu}$ et $(q_{\mu})_{\mu}$ sont identiques. Mais la matrice des probabilités de $|\langle \mu | \nu \rangle|^2$ étant bistochastique, cela entraîne qu'elle est réduite à l'identité. Les bases propres de ρ et ρ' sont identiques ainsi que les éléments de matrice diagonaux : $\rho_{\mu} = \rho'_{\mu}$ et donc $\rho = \rho'$. Ceci généralise au niveau quantique la propriété vérifiée par la divergence de Kullback-Leibler classique.

7.5 Conditionnement en théorie quantique

Au niveau classique, le conditionnement est l'établissement de corrélations entre Alice et Bob de sorte que les probabilités conditionnelles pour l'un relativement à l'autre ne sont pas les probabilités marginales. Ces corrélations peuvent être utilisées pour communiquer en utilisant deux lignes en parallèle, les performances étant données par la borne Slepian-Wolf dans la théorie de Shannon (chapitre 2). Celle-ci nous dit que Bob a besoin de transmettre $S[A|B]$ bits à Alice pour qu'elle ait accès à toute l'information partagée entre eux.

La même question se pose dans le cadre de la théorie quantique de la communication : imaginons qu'Alice et Bob partagent des états quantiques issus d'une source imparfaite et décrits par un opérateur densité ρ_{AB} , que doit faire Bob pour qu'Alice ait accès à toute l'information de son côté ? Dans ce cas, il est crucial que Alice et Bob partagent des corrélations quantiques qui trouvent leur origine dans le partage d'états intriqués. C'est donc l'intrication qui joue le rôle du conditionnement quantique et nous verrons comment définir une bonne notion d'entropie quantique conditionnelle pour la théorie de Shannon quantique.

7.5.1 Conditionnement classique en théorie quantique

La première notion de conditionnement à laquelle nous pouvons penser est le conditionnement par rapport à des données classiques comme des résultats de mesure. Imaginons qu'Alice et Bob partagent un opérateur densité ρ_{AB} . Supposons que Bob mesure une quantité X associée à une base orthonormée de vecteurs $|x\rangle \in \mathcal{H}_B$. Pour un résultat x donné, Alice reçoit un mélange statistiques d'états $\rho(A|x)$ conditionné au résultat x . Si on note $p_B(x)$ la probabilité que Bob obtienne le résultat x , l'opérateur densité post-mesure pour Alice et Bob est donné par

$$\rho_{A,B_X} = \sum_x p_B(x) \rho(A|x) \otimes |x\rangle\langle x|, \quad (7.54)$$

où la notation B_X nous rappelle la dépendance en la quantité X mesurée par Bob. Il paraît alors naturel de définir l'entropie conditionnelle d'Alice par la moyenne des entropies de von Neumann des opérateurs $\rho(A|x)$ comme

$$S[A|B_X] = \sum_x p_B(x) S_{\text{vn}}[\rho(A|x)]. \quad (7.55)$$

La trace partielle sur $\mathcal{H}_A$ de ρ_{A,B_X} est l'opérateur densité réduit post-mesure pour Bob $\rho_{B_X} = \sum_x p_B(x) |x\rangle\langle x|$. L'équation (7.45) nous montre que, dans ce

cas précis,

$$S_{\text{vn}}[\rho_{A,B_X}] = S[p_B(x)] + \sum_x p_B(x) S_{\text{vn}}[\rho(A|x)], \quad (7.56a)$$

$$S_{\text{vn}}[\rho_{B_X}] = S[p_B(x)]. \quad (7.56b)$$

L'entropie conditionnelle définie par l'équation (7.55) est alors donnée par

$$S[A|B_X] = S_{\text{vn}}[\rho_{A,B_X}] - S_{\text{vn}}[\rho_{B_X}]. \quad (7.57)$$

Cette entropie conditionnée à des résultats de mesure classique est donc bien de la forme $S[\rho_{AB}] - S[\rho_B]$ de manière tout à fait analogue à l'expression classique (2.57) mais ce sont les opérateurs densités post-mesure par Bob qui interviennent. Ceci est bien naturel car c'est la mesure qui génère le conditionnement classique.

Le problème consiste maintenant à cerner la notion de conditionnement quantique en général et à disposer de l'entropie conditionnelle associée. Pour cela, nous allons procéder en faisant un pas inductif en proposant directement une définition naturelle de l'entropie conditionnelle quantique. Puis, nous verrons que cette définition permet d'effectuer des bilans d'entropie cohérents y compris dans un protocole de communication quantique. Enfin, c'est en approfondissant l'interprétation physique de l'entropie conditionnelle quantique en termes de communication quantique que nous comprendrons vraiment ce qu'est le conditionnement quantique.

7.5.2 Entropie conditionnelle quantique

Définition

La discussion de la section précédente suggère de définir *l'entropie conditionnelle quantique* par

$$S_{\text{vn}}[A|B] = S_{\text{vn}}[A, B] - S_{\text{vn}}[B]. \quad (7.58)$$

qui est la transposition au cas quantique de l'expression (2.57) de la théorie de Shannon. La sous-additivité de l'entropie de von Neumann (7.46) entraîne que $S_{\text{vn}}[A|B] \leq S_{\text{vn}}[A]$ comme dans le cas classique. Cependant, une différence importante apparaît lorsque l'on considère un état pur $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. En effet, $S_{\text{vn}}[A, B] = 0$ alors que $S_{\text{vn}}[\rho_A] = S_{\text{vn}}[\rho_B] > 0$ si ρ_A et ρ_B ne sont pas purs. L'entropie conditionnelle quantique peut donc être négative.

Avant d'éclairer l'origine de cette étrangeté, une discussion comparée du chiffre de Vernam et de la téléportation quantique va nous apprendre que cela est nécessaire pour obtenir des bilans d'information quantique cohérents.

Dans le chiffre de Vernam (section 7.2.1), Alice et Bob cherchent à communiquer un bit classique en utilisant une paire classique corrélée, ce que nous avons exprimé en termes de ressources par l'inégalité (7.9). Comparons les

bilans des flots d'entropie faits par Alice et Bob. Alice veut transmettre un bit aléatoire c_A en utilisant celui qui provient de la paire classique (c_1, c_2) partagée avec Bob. Elle dispose donc de $S[c_A c_1] = S[c_A] + S[c_1] = 2$ bits d'information. Elle calcule alors le **xor** de manière irréversible, c'est-à-dire en oubliant un bit, et transmet le résultat c_r à Bob, soit $S[c_r] = S[c_A c_1] - 1 = 1$ bit d'information.

Maintenant, considérons que Bob veuille calculer l'information $S[c_r c_2]$ qu'il reçoit. En regardant d'où elle provient, il va dire qu'elle est égale à $S[c_A c_1 c_2] - 1$, où le -1 provient du bit classique oublié dans l'évaluation du **xor**. Bob utilise alors le fait que c_1 et c_2 sont classiquement corrélés. Il écrit que $S[c_A c_1 c_2] = S[c_A c_1] + S[c_2 | c_1]$ avec $S[c_2 | c_1] = 0$ (corrélation classique parfaite). Bob en déduit donc qu'il dispose de $2 + 0 - 1 = 1$ bit. Les résultats d'Alice et Bob sont cohérents.

Effectuons les bilans analogues pour la téléportation quantique. Alice dispose d'un qubit q_A qui code un message classique aléatoire et reçoit un qubit provenant d'une paire maximalelement intriquée $[e_1 e_2]$ qu'elle partage avec Bob. Elle dispose donc d'une information quantique $S[q_A e_1] = S[q_A] + S[e_1] = 2$. D'après le théorème de Holevo, elle peut en extraire au maximum deux bits classiques qu'elle transmet à Bob. Ce dernier dispose donc des deux bits transmis et de e_2 . Il utilise ses deux bits classiques pour faire une action de contrôle sur un qubit physique et il lui reste donc un bit quantique.

Maintenant, considérons le point de vue de Bob. Il dispose de ressources qui proviennent de $q_A e_1 e_2$. Dans l'esprit du calcul précédent, Bob va écrire que cela correspond à $S[q_A e_1 e_2] = S[q_A e_1] + S[e_2 | q_A e_1]$ de manière à introduire dans le bilan d'information le fait que la paire $[e_1 e_2]$ est maximalelement intriquée. Sachant que le qubit q_A est décorrélié de la paire $[e_1 e_2]$, $S[e_2 | q_A e_1] = S[e_2 | e_1]$. Pour que ce bilan fait par Bob donne $S[q_A e_1 e_2] = 1$ sachant que $S[q_A e_1] = 2$, nous avons nécessairement $S[e_2 | e_1] = -1$. Un bilan cohérent nécessite donc bien des valeurs négatives pour l'entropie conditionnelle.

Bornes inférieures classiques et quantiques

Des bornes sur l'entropie conditionnelle découlent directement des inégalités sur l'entropie d'une source composée. La figure 7.10 représente les zones permises pour l'entropie conditionnelle classique et quantique. Dans le premier cas, S désigne l'entropie de Shannon. Dans le second, S désigne l'entropie de von Neumann. Cette figure représente donc en quelque sorte le « diagramme des phases informationnel » pour un système quantique bipartite. Sur la partie gauche de la figure, il est exprimé dans les variables $S[B]/S[A]$ et $S[A|B]/S[A]$ qui sont adaptées à la discussion des valeurs possibles pour entropie conditionnelle quantique. Rappelons que la donnée de ces deux nombres détermine $S[A, B]/S[A]$ et donc $I[A, B]/S[A]$ et aussi $S[B|A]/S[A]$. On peut donc aussi représenter le « diagramme de phases informationnel » en les variables $\lambda = S[B]/S[A]$ et $I[A, B]/S[A]$ (voir la partie droite de la figure 7.10) ce qui permet de mieux visualiser le diagramme en termes d'une mesure des corrélations entre A et B .

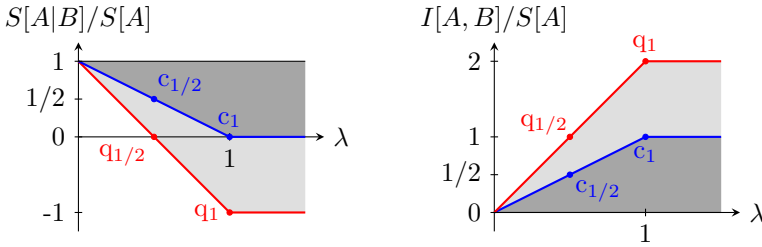


FIGURE 7.10 – Zones autorisées pour l’entropie conditionnelle de Shannon (zone gris foncé, borne inférieure ou supérieure en bleu) et pour l’entropie de von Neumann (zone gris clair, borne inférieure ou supérieure en rouge) dans les variables $\lambda = S[B]/S[A]$ et $S[A|B]/S[A]$ (diagramme de gauche) ainsi que dans les variables $\lambda = S[B]/S[A]$ et $I[A, B]/S[A]$ (diagramme de droite) lorsque $S[A] \geq 0$. Les points (c_λ) et (q_λ) correspondent aux bornes inférieures classiques et quantiques pour $\lambda = S[B]/S[A]$ fixé.

Déjà, $S[A|B]/S[A] \leq 1$ et $I[A, B]/S[A] \geq 0$ dans les deux régimes. Par contre, les bornes inférieures dans les régimes classiques et quantiques diffèrent. Dans le cas classique $S[A, B] \geq \max(S[A], S[B])$ et donc $S[A|B] \geq 0$ lorsque $S[B] \geq S[A]$, et $S[A|B] \geq S[A] - S[B]$ lorsque $0 \leq S[B] \leq S[A]$. L’information conditionnelle est alors bornée supérieurement par $\min(S[A], S[B])$ et inférieurement par zéro. En revanche, dans le cas quantique, lorsque $S_{\text{vn}}[A] \leq S_{\text{vn}}[B]$, la borne inférieure d’Araki-Lieb montre que $S_{\text{vn}}[A|B] \geq S_{\text{vn}}[A] - 2S_{\text{vn}}[B]$ pour $S_{\text{vn}}[B] \leq S_{\text{vn}}[A]$ et que $S_{\text{vn}}[A|B] > -S_{\text{vn}}[A]$ lorsque $S_{\text{vn}}[B] \geq S_{\text{vn}}[A]$. L’information conditionnelle de von Neumann est alors bornée supérieurement par $2 \min(S_{\text{vn}}[A], S_{\text{vn}}[B])$, soit 2λ pour $\lambda < 1$ et 1 au-delà.

Le cas du chiffre de Vernam correspond à Alice et Bob qui partagent chacun un bit classique : $S[B]/S[A] = 1$ et $S[A|B] = 0$ (point (c_1)). Pour la téléportation quantique, on utilise une paire maximalelement intriquée (point (q_1)). Dans ce cas, $S_{\text{vn}}[B]/S_{\text{vn}}[A] = 1$ et $S_{\text{vn}}[A|B] = -1$.

Cependant, la figure 7.10 montre que certaines zones interdites ne correspondent pas à des valeurs négatives de $S[A|B]$. Pour mieux comprendre leur signification, considérons une situation asymétrique pour laquelle $S[B] = 1$ et $S[A] = 2$: dans le cas classique $S[A|B] \geq 1 = S[A]/2$ (point $(c_{1/2})$). En clair, si Alice dispose de deux lancers de pile ou face, en connaître un parfaitement lui en laisse encore un. En revanche, dans le cas non classique, la borne inférieure sur $S_{\text{vn}}[A|B]$ est zéro (point $(q_{1/2})$). Le modèle que nous allons discuter permet d’interpoler entre ces deux situations et va nous éclairer sur le rôle de l’intrication quantique, sur l’importance de la purification et enfin sur l’interprétation des entropies conditionnelles négatives en termes de communication quantique.

Au point $(c_{1/2})$, Bob dispose d’un qubit physique et Alice de deux. Nous voulons que $S_{\text{vn}}[B] = 1$ et donc que l’état de son qubit soit décrit par l’opérateur

densité réduit $\rho_B = (|0\rangle\langle 0| + |1\rangle\langle 1|)_B/2$. De même, nous voulons que l'opérateur densité réduit des deux qubits en possession d'Alice soit

$$\rho_A = \frac{1}{4} \sum_{(x,y) \in \{0,1\}^2} |xy\rangle_A \langle xy|_A . \quad (7.59)$$

La situation classique maximalement corrélée ($c_{1/2}$) pour laquelle $S_{\text{vn}}[A|B] = 1$ est obtenue en corrélant la valeur du bit y dont dispose Alice dans le mélange à la valeur du bit dont dispose Bob. Un exemple d'opérateur densité possible qui satisfait cette contrainte est

$$\rho_{AB}^{(c_{1/2})} = \frac{1}{4} \sum_{(x,y) \in \{0,1\}^2} |xy\rangle_A \langle xy|_A \otimes |y\rangle_B \langle y|_B . \quad (7.60)$$

Dans cet exemple simple, $S_{\text{vn}}[A, B] = 2$ et donc $S_{\text{vn}}[A|B] = 1 = S[A]/2$.

Pour trouver comment battre cette borne classique, nous allons purifier ce mélange en introduisant un environnement E . On considère alors quatre états $|E_{x,y}\rangle \in \mathcal{H}_E$ et on définit

$$|\Psi_{ABE}^{(c_{1/2})}\rangle = \frac{1}{2} \sum_{(x,y)} |xy\rangle_A |y\rangle_B |E_{x,y}\rangle . \quad (7.61)$$

Lorsque les quatre états $|E_{x,y}\rangle$ sont orthogonaux, la trace partielle sur $\mathcal{H}_E$ redonne bien l'opérateur densité $\rho_{AB}^{(c_{1/2})}$. Dans cette situation, Alice, Bob et l'environnement partagent un état intriqué où les deux bits classiques (x, y) d'Alice sont associés à des états $|E_{x,y}\rangle$ discernables à coup sûr.

Cependant, $S_{\text{vn}}[\rho_A] = 2$ impose une condition plus faible que la précédente sur les états $|E_{x,y}\rangle$, à savoir l'orthogonalité seulement pour $x \neq x'$ car les états $|0\rangle_B$ et $|1\rangle_B$ constituent déjà deux « empreintes » maximalement discernables du bit classique y . Remarquons enfin que dans tous les cas, l'état $|\Psi\rangle_{ABE}$ conduit à $S[B] = 1$ car la base $|xy\rangle_A$ pour Alice est orthonormée. La situation du paragraphe précédent présente une redondance car pour discerner entre les deux valeurs du bit classique y , nous avons une double empreinte : une dans l'orthogonalité des états $|E_{x,y}\rangle$ et l'autre dans l'orthogonalité des états $|y\rangle_B$. Nous pouvons donc ouvrir un espace de liberté supplémentaire en éliminant la redondance dans le codage du bit y *via* des états orthogonaux dans $\mathcal{H}_B \otimes \mathcal{H}_E$.

Ainsi, si nous prenons des états $|E_{x,y}\rangle$ qui ne sont pas orthogonaux pour $y \neq y'$, l'opérateur densité réduit ρ_{AB} ne sera plus d'entropie maximale alors que nous aurons toujours $S_{\text{vn}}[A] = 2$ et $S_{\text{vn}}[B] = 1$. Dans cet exemple, c'est l'entropie conditionnelle $S_{\text{vn}}[B|A]$ qui va devenir négative. Le cas extrême est celui où $|E\rangle_{x,y} = |\varphi_x\rangle_E$ avec $\langle \varphi_0 | \varphi_1 \rangle = 0$. Dans ce cas, le codage des deux bits classiques (x, y) est réparti de la manière la plus économique possible entre Bob et l'environnement. Dans ce cas limite,

$$|\psi_{ABE}^{(q_{1/2})}\rangle = \frac{1}{2} \sum_{(x,y)} |xy\rangle_A |y\rangle_B |\varphi_x\rangle_E \quad (7.62)$$

et l'opérateur densité réduit pour Alice et Bob est donné par

$$\rho_{AB}^{(q_{1/2})} = \frac{1}{2} \sum_x |\Psi_x\rangle_{AB} \langle \Psi_x|_{AB} , \quad (7.63a)$$

$$|\Psi_x\rangle_{AB} = \frac{1}{\sqrt{2}} \sum_y |xy\rangle_A |y\rangle_B . \quad (7.63b)$$

Alice et Bob partagent donc un mélange de deux états intriqués orthogonaux. En conséquence, $S_{\text{vn}}[(A, B)] = 1$. Cet exemple atteint donc la borne inférieure $S_{\text{vn}}[A|B] = 1 - 1 = 0$ pour l'entropie conditionnelle quantique dans le cas où $S_{\text{vn}}[B] = S_{\text{vn}}[A]/2$. De plus dans ce cas, $S_{\text{vn}}[B|A] = -1$ qui reflète le fait qu'Alice et Bob partagent en fait la même paire de Bell : $|\Psi_x\rangle_{AB} = |x\rangle_{A_1} \otimes |B_0\rangle_{A_2B}$ (voir équation (5.22)). Au point $(q_{1/2})$, Alice et Bob ne partagent pas de corrélations classiques mais des corrélations quantiques sous la forme de paires maximalement intriquées. Là réside la différence entre un conditionnement purement classique et un conditionnement quantique : alors que dans le cas classique, Alice et Bob partagent des paires classiquement corrélées, ils partagent des paires maximalement intriquées dans le cas quantique.

Pour compléter cette discussion, nous invitons le lecteur à analyser les cas intermédiaires dans lesquels les états $|E_{x,0}\rangle$ et $|E_{x,1}\rangle$ avec x fixé ne sont pas parfaitement discernables et à voir que $S_{\text{vn}}[A|B]/S_{\text{vn}}[A]$ varie bien de zéro (cas totalement indiscernable) à $1/2$ (cas discernable à coup sûr) et que dans ce régime $S_{\text{vn}}[B|A]$ est négatif! De manière générale, le lecteur vérifiera que $S_{\text{vn}}[B|A] < 0$ dans la zone comprise entre les bornes classiques et quantiques où $S_{\text{vn}}[A|B] > 0$. De manière générale, le régime quantique correspond aux cas où au moins une des deux entropies conditionnelles $S_{\text{vn}}[A|B]$ ou $S_{\text{vn}}[B|A]$ est strictement négative.

7.6 La théorie de Shannon quantique

À ce stade, prenons un instant pour regarder où nous en sommes : nous disposons des briques de base que sont les protocoles quantiques élémentaires et nous avons quantifié les limites physiques à la fongibilité des ressources quantiques idéales entre Alice et Bob que sont les paires maximalement intriquées ou les utilisations de canaux parfaits. Nous avons également explicité les différentes valeurs en information classique d'un qubit et quelles sont les limites à la compression de l'information quantique. Enfin, nous avons compris que le conditionnement quantique n'était autre que l'établissement de corrélations quantiques entre Alice et Bob.

Tout ce travail préliminaire a permis de formuler dans des cas simples les conséquences en termes de communication quantique des théorèmes d'impossibilité vus au chapitre 6 et des théorèmes fondamentaux liés à l'intrication du chapitre 5. Nous sommes donc maintenant en bonne position pour discuter de

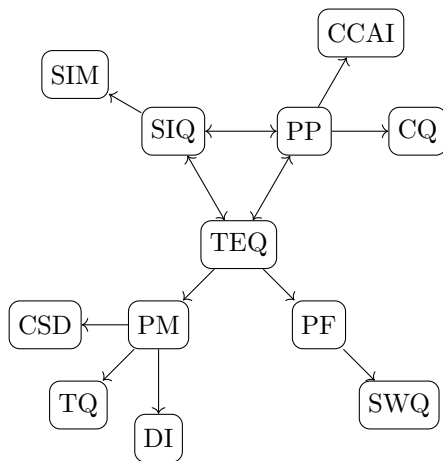


FIGURE 7.11 – Arbre généalogique des protocoles de communication quantique et de leurs descendants [4] : les trois protocoles fondamentaux que sont le transfert d'état quantique (TEQ, voir l'inégalité (7.71)), le protocole père (PP, voir l'inégalité (7.89)) [142] et Shannon inverse quantique (SIQ, voir l'inégalité (7.84)) sont reliés par diverses dualités [130]. Le protocole père permet de déduire les résultats relatifs à la capacité du canal quantique (CQ, inégalité (7.100)) [190] ainsi qu'à la capacité classique assistée par intrication (CCAI). L'inégalité de Shannon inverse quantique permet de voir comment simuler un canal quantique à partir d'intrication et de communication classique (SIM). Enfin le protocole de transfert d'état quantique nous a permis de déduire d'une part le protocole de fusion (PF, inégalité (7.72)) [207] qui fournit la réponse au problème de Slepian-Wolf quantique (SWQ), ainsi que le protocole mère (PM) d'où découlent les généralisations du codage superdense (CSD, inégalité (7.81)), de la téléportation quantique (TQ, inégalité (7.80)) ainsi que l'inégalité caractérisant la distillation d'intrication (DI, inégalité (7.78)) [42].

la théorie de Shannon quantique qui vise à formaliser en toute généralité les limites physiques aux échanges d'information classiques et quantiques entre Alice et Bob.

Nous verrons qu'elles peuvent être déduites à partir de deux résultats fondamentaux : d'une part le protocole de transfert d'état quantique qui définit comment transférer optimalement l'information quantique d'Alice vers Bob lorsque ces deux partenaires partagent des corrélations quantiques imparfaites et, d'autre part, le protocole dual qui décrit quelles sont les limites au partage d'information quantique de Bob vers Alice. La généalogie complète de ces différents résultats est résumée par la figure 7.11.

Les descendants du protocole de transfert d'état quantique concernent le partage d'une ressource dite statique comme la donnée de corrélations quantiques tripartites. En revanche, les protocoles duaux et leurs descendants

impliquent tous la donnée d'une ressource dynamique comme un canal de communication quantique. Généralisant la théorie de Shannon présentée au chapitre 2, ces résultats sont cruciaux pour décrire les limites aux communications quantiques.

7.6.1 Information conditionnelle et communication

L'entropie conditionnelle classique trouve son interprétation naturelle en théorie de la communication *via* la borne de Slepian-Wolf (section 2.5.4). Celle-ci quantifie le nombre de bits qu'Alice doit communiquer à Bob pour que celui-ci récupère toute l'information partagée entre Alice et Bob *via* des corrélations classiques. Or, nous avons vu dans la section précédente que le conditionnement quantique consiste non pas à partager des corrélations classiques mais des corrélations quantiques. Ces deux remarques nous conduisent donc naturellement à chercher une interprétation opérationnelle de l'entropie conditionnelle quantique en termes de communication quantique y compris dans le cas où Alice et Bob partagent des corrélations quantiques. Pour aborder cette question importante, nous allons procéder en revenant sur les exemples déjà considérés précédemment pour dégager intuitivement cette interprétation.

L'idée générale consiste à décrire les corrélations entre Alice et Bob, indépendamment de leur caractère classique ou quantique, par un état pur partagé entre Alice, Bob et leur environnement. Dans cette optique, la question posée précédemment consiste à trouver comment et à quel prix transférer l'information partagée entre Alice et Bob de manière à ce qu'elle se retrouve uniquement chez Bob sans altérer les corrélations de ce dernier avec l'environnement. C'est un problème de transfert d'état quantique entre Alice et Bob qui rappelle la téléportation quantique à ceci près que l'on part d'un état intriqué tripartite au lieu de bipartite. La question est alors de savoir quelles sont les ressources de communication classique ou quantique qu'il faut mobiliser pour cela.

Le cas classique

Considérons tout d'abord le cas général dans lequel Alice et Bob partagent des corrélations classiques

$$\rho_{AB} = \sum_{(x,y)} p(x,y) |x\rangle_A \langle x|_A \otimes |y\rangle_B \langle y|_B \quad (7.64)$$

au travers d'états orthogonaux $|x\rangle_A$ pour Alice (resp. $|y\rangle_B$ pour Bob) associés aux variables aléatoires x (resp. y) distribuées selon la loi $p(x,y)$. Une purification de l'opérateur densité ρ_{AB} est obtenue en introduisant un environnement E et une famille orthonormée $|xy\rangle_E$ de $\mathcal{H}_E$ de sorte que l'état tripartite

$$|\Psi_{ABE}\rangle = \sum_{x,y} \sqrt{p(x,y)} |x\rangle_A |y\rangle_B |xy\rangle_E \quad (7.65)$$

vérifie $\rho_{AB} = \text{tr}_{\mathcal{H}_E} |\Psi_{ABE}\rangle \langle \Psi_{ABE}|$. La famille orthonormée $|xy\rangle_E$ supprime les effets d'interférence quantique entre couples (x, y) différents. Réécrivons maintenant $|\psi_{ABE}\rangle$ en décomposant l'environnement en deux fragments E_1 et E_2 , chacun encodant une des deux variables aléatoires (x, y) :

$$|\psi_{ABE}\rangle = \sum_y \sqrt{p_B(y)} \sum_x \sqrt{p(x|y)} |x\rangle_A |y\rangle_B |x\rangle_{E_1} |y\rangle_{E_2}. \quad (7.66)$$

Pour chaque valeur de y , Alice et l'environnement partagent un état intriqué $|\Psi(AE_1|y)\rangle = \sum_x \sqrt{p(x|y)} |x\rangle_A \otimes |x\rangle_{E_1}$. Toute la question consiste donc à comprendre comment traduire en termes quantiques le processus de communication classique de x distribué selon la loi $p(x|y)$ entre Alice et Bob. Cela revient à se demander combien de fois nous allons devoir utiliser un canal quantique parfait entre Alice et Bob pour transférer $|\Psi(AE_1|y)\rangle$ d'Alice et E_1 à Bob et E_1 . Le théorème de Schumacher énoncé dans la section 7.3.3 répond précisément à cette question : il faut $S_{\text{vn}}[\text{tr}_{\mathcal{H}_{E_1}}(|\Psi(AE_1|y)\rangle \langle \Psi(AE_1|y)|)]$ utilisations d'un canal quantique à y fixé pour effectuer le transfert. L'expression de l'état $|\Psi(AE_1|y)\rangle$ montre que cette entropie est égale à l'entropie de Shannon de la loi conditionnelle $p(\cdot|y)$. Une fois toutes les valeurs de y considérées avec leur poids statistique $p_B(y)$, il faudra donc $S[p|p_B] = S_{\text{vn}}[A|B]$ utilisations d'un canal quantique parfait pour effectuer le transfert. Nous obtenons ainsi une interprétation de l'entropie conditionnelle classique en termes de communication quantique lorsque Alice et Bob partagent des corrélations classiques : $S_{\text{vn}}[A|B] = S[p|p_B] \geq 0$ est le nombre de qubits à transférer d'Alice à Bob pour passer à ce dernier l'information quantique qui initialement était partagée entre Alice et Bob.

Le cas quantique

Passons maintenant au cas quantique. Pour commencer, nous allons considérer les valeurs interdites classiquement de $S_{\text{vn}}[A|B]$ correspondant aux points $(q_{1/2})$ et (q_1) de la figure 7.10.

Considérons pour commencer le cas $(q_{1/2})$ ($S_{\text{vn}}[A|B] = 0$ et $S_{\text{vn}}[B|A] = -S_{\text{vn}}[B]$) dans lequel Alice et Bob partagent un mélange de deux états intriqués (équation (7.62)). D'un point de vue global, les deux qubits dont dispose Alice font chacun partie d'une paire maximalement intriquée : celle impliquant le bit classique x est maximalement intriquée entre Alice et l'environnement alors que celle impliquant y l'est entre Alice et Bob. Une fois transférée à Bob l'information quantique en possession d'Alice, celui-ci se retrouvera avec trois qubits physiques⁸ dont deux porteront l'information sur la variable y et une celle sur x . Pour aboutir à cela, Bob commence par dupliquer l'information sur la variable classique y . En clair, il transforme $|y\rangle_B$ en $|y\rangle_B^{\otimes 3}$ ce qui est une opération de clonage de deux états orthogonaux connus que Bob réalise au

8. Deux correspondant à ceux anciennement en possession d'Alice auquel s'adjoint celui dont il dispose déjà.

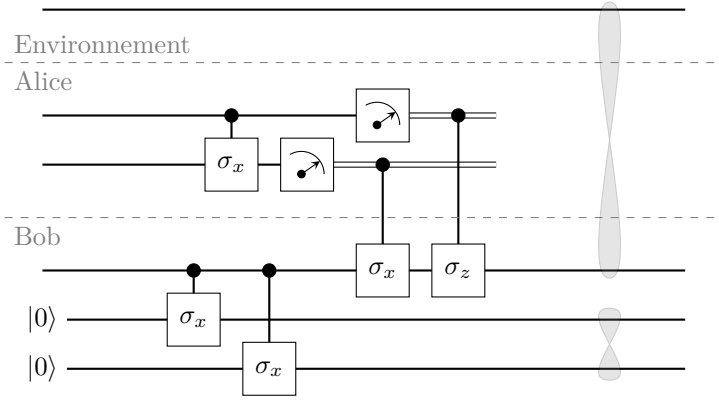


FIGURE 7.12 – Représentation des opérations effectuées pour transférer sur Bob et l'environnement les corrélations quantiques tripartites partagées *via* $|\psi_{ABE}^{(q)1/2}\rangle$ donné par l'équation (7.62). Deux portes **cnot** permettent à Bob de cloner son qubit dans l'état $|y\rangle$ ($y = 0, 1$) sur deux autres qubits auxiliaires initialisés à $|0\rangle$. Alice effectue alors une mesure dans la base de Bell sur ses deux qubits puis transfère le résultat à Bob qui applique alors deux opérations conditionnées à ces résultats, réalisant ainsi un échange d'intrication avec Alice sur la paire de Bell qu'elle partage avec l'environnement.

moyen de portes **cnot**. L'étape suivante consiste en un échange d'intrication entre l'environnement et le premier qubit physique dont dispose maintenant Bob. Pour cela, Alice va mesurer les deux qubits physiques dont elle dispose et transmettre les résultats de ses mesures classiques (2 bits) à Bob qui appliquera alors l'opération unitaire adaptée sur son premier qubit physique. Remarquons que dans ce processus, le bit classique y reste une variable aléatoire sur laquelle on somme. Finalement, Alice a mesuré ses qubits physiques qui ont donc été laissés de côté et l'état complet partagé par Bob et l'environnement est

$$|\psi_{B'E}^{(q_1/2)}\rangle = \frac{1}{2} \sum_{(x,y)} |xyy\rangle_{B'} |x\rangle_E . \quad (7.67)$$

Nous avons donc transféré toute l'information quantique partagée par Alice et Bob à ce dernier. Cela s'est fait au moyen d'un protocole représenté sur la figure 7.12 qui n'implique que des opérations locales au niveau d'Alice et de Bob et l'échange de deux bits classiques entre eux. Il n'a pas été nécessaire d'utiliser un canal de communication quantique entre Alice et Bob. Rappelons que pour cet exemple $S_{\text{vn}}[A|B] = 0$ et que $I_{\text{vn}}[A, B] = 2$. Ceci suggère que $S_{\text{vn}}[A|B] = 0$ représente encore le nombre d'utilisations d'un canal quantique et que $I_{\text{vn}}[A, B] = 2$ désigne le nombre de bits classique qu'Alice doit transférer à Bob.

Au point (q_1) , Alice et Bob partagent une paire intriquée maximale-
ment intriquée

$$|\psi_{AB}^{(q_1)}\rangle = \frac{|0_A 0_B\rangle + |1_A 1_B\rangle}{\sqrt{2}}, \quad (7.68)$$

de sorte que $S_{\text{vn}}[A] = S_{\text{vn}}[B] = 1$, $S_{\text{vn}}[A, B] = 0$ et donc $S_{\text{vn}}[A|B] = -1$ et $I_{\text{vn}}[A, B] = 2$. Alice et Bob savent qu'ils partagent une paire maximale-
ment intriquée. Bob va donc juste en produire une autre, dans le même état. Il
disposera alors de l'information qui est contenue dans ρ_{AB} et, en plus, d'une
paire maximale-ment intriquée avec Alice qu'il sera possible d'utiliser pour
une communication quantique ultérieure au moyen du protocole de télépor-
tation quantique. Alice n'a pas eu besoin de transférer de qubit à Bob et
 $-S_{\text{vn}}[A|B] > 0$ semble apparaître comme le nombre de paires maximale-
ment intriquées subsistant entre Alice et Bob après le transfert. Remarquons que si
Alice s'autorise le transfert de deux bits classiques à Bob, alors elle dispose
d'une capacité de transfert d'un qubit vers Bob, qui par ailleurs dispose des
corrélations quantiques contenues dans ρ_{AB} .

Un argument de comptage vient préciser cette intuition. En effet, comme
l'état tripartite est pur, $I_{\text{vn}}[A, E] = S_{\text{vn}}[A] + S_{\text{vn}}[A|B]$. Si on admet que
les corrélations quantiques entre Alice et l'environnement sont portées par
 $I_{\text{vn}}[A, E]/2$ paires maximale-ment intriquées partagées entre eux, et que l'on
peut en isoler les qubits correspondant au moyen d'opérations locales du
côté d'Alice, les transférer à Bob nécessite la transmission de $I_{\text{vn}}[A, E]$ bits
classiques d'Alice vers Bob. Ceux-ci proviennent des mesures des $I_{\text{vn}}[A, E]/2$
paires maximale-ment intriquées partagées avec E et de $I_{\text{vn}}[A, E]/2$ paires maxi-
malement partagées avec Bob. Il reste alors $S_{\text{vn}}[A] - I_{\text{vn}}[A, E] = -S_{\text{vn}}[A|B]$
qubits disponibles côté Alice dans le cas où $S_{\text{vn}}[A|B] < 0$. Si on s'autorise
en tout $I_{\text{vn}}[A, B]$ utilisations d'un canal classique d'Alice vers Bob, comme
 $I_{\text{vn}}[A, B] = S_{\text{vn}}[A] - S_{\text{vn}}[A|B]$, il reste encore

$$I_{\text{vn}}[A, B] - I_{\text{vn}}[A, E] = -2S_{\text{vn}}[A|B] \quad (7.69)$$

utilisations possibles du canal classique. Avec les $-S_{\text{vn}}[A|B]$ paires intriquées
partagées avec Bob, cela permet de simuler $-S_{\text{vn}}[A|B] > 0$ transmissions de
qubits d'Alice vers Bob [86]. En revanche, comme nous l'avons vu, si Alice
et Bob partagent des corrélations classiques, $S_{\text{vn}}[A|B] \geq 0$ est le nombre de
fois qu'Alice doit utiliser un canal quantique idéal pour que Bob récupère
toute l'information qu'il partageait avec Alice après envoi de $I_{\text{vn}}[A, B]$ bits
classiques.

Vers un protocole de transfert d'état quantique

Un des points clés dans cette discussion intuitive est l'idée que l'information
mutuelle quantique compte le nombre de paires maximale-ment intriquées qui
portent les corrélations quantiques entre deux systèmes et qu'il est possible,

au moyen d'opérations locales effectuées sur un des partenaires, d'isoler les qubits correspondants chez celui-ci. À la lumière de cette remarque, l'égalité

$$I_{\text{vn}}[A, E] + I_{\text{vn}}[A, B] = 2S_{\text{vn}}[A] \quad (7.70)$$

qui découle directement de la pureté de l'état tripartite suggère qu'il est possible de partitionner, après opération locale côté Alice, tous ses qubits entre $I_{\text{vn}}[A, E]/2$ qubits portant les corrélations avec E et $I_{\text{vn}}[A, B]/2$ portant les corrélations avec Bob. Ceci semble indiquer également que, si les qubits d'Alice sont maximalelement intriqués avec Bob (respectivement avec l'environnement), ils ne peuvent être maximalelement intriqués avec l'environnement (respectivement avec Bob). Cette intuition physique s'avère correcte et s'appelle la *monogamie de l'intrication*.

On peut donner un argument physique qualitatif simple pour justifier ce principe en montrant que son contraire, que l'on appelle la polygamie de l'intrication, est équivalente au clonage. Tout d'abord, donnons un sens à la polygamie : il s'agit de la possibilité, par un processus physique, d'étendre le partage de corrélations quantiques entre deux systèmes à plusieurs couples de systèmes impliquant à chaque fois un partenaire commun. Imaginons donc qu'Alice et Bob partagent des corrélations quantiques sous la forme la plus générale possible, c'est-à-dire *via* un mélange d'états intriqués décrit par un opérateur densité ρ agissant sur $\mathcal{H}_A \otimes \mathcal{H}_B$ où $\mathcal{H}_A$ correspond au système dont dispose Alice. On dira qu'il y a polygamie s'il est possible, par un processus physique, de créer un état partagé $\rho_{AB_1B_2}$ entre Alice, Bob (B_1) et un troisième système (Bernard, alias B_2), tel que $\mathcal{H}_B = \mathcal{H}_{B_1} \simeq \mathcal{H}_{B_2}$ et $\text{tr}_{B_2} \rho_{AB_1B_2} = \text{tr}_{B_1} \rho_{AB_1B_2} = \rho$. Un tel état « ménage à trois » est obtenu en injectant $\rho_{AB} = \rho$ dans une machine à cloner du côté de Bob et en ne faisant rien côté Alice. Réciproquement, la polygamie permet de réaliser un cloneur. En effet, imaginons disposer d'un état $\rho_{AB_1B_2}$ polygame. Imaginons que l'on effectue une téléportation quantique à trois : Alice disposant aussi d'un état quantique $|\varphi\rangle$ générique, elle effectue une mesure dans la base de Bell sur la paire constituée de ce qubit et de celui du ménage à trois dont elle dispose puis transfère alors le résultat classique à Bob et Bernard qui appliquent chacun à leur qubit l'unitaire correspondant dans le protocole de téléportation quantique. La polygamie entraîne que Bob et Bernard se retrouvent chacun avec une copie de $|\varphi\rangle$, ce qui réalise le clonage. Monogamie de l'intrication et non-clonage sont donc équivalents.

Tout cela suggère qu'en transférant les $I_{\text{vn}}[A, E]/2$ qubits d'Alice corrélés avec l'environnement vers Bob, Alice ne conserve que les qubits portant les corrélations qu'elle partage avec Bob. Ce dernier dispose alors en plus des corrélations qu'elle partageait avec l'environnement. Comme nous allons le voir, c'est précisément l'idée sous-jacente au protocole de transfert d'état quantique qui permet d'engendrer toute la généalogie des protocoles. L'intérêt de cette approche par rapport aux discussions précédentes est qu'elle nous donne un

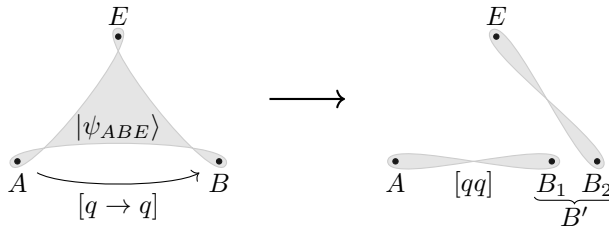


FIGURE 7.13 – Schéma de ce que réalise un protocole de transfert d'état quantique. À gauche : au départ, Alice, Bob et leur environnement partagent un état pur intriqué. Alice et Bob ne peuvent qu'agir localement et utiliser un canal de communication parfait entre eux. À droite : après réalisation du protocole, l'état ρ_{AB} partagé par Alice et Bob est totalement en possession de Bob et ses corrélations quantiques avec l'environnement sont préservées. Nous verrons qu'il peut subsister un certain nombre de paires intriquées partagées entre Alice et Bob.

protocole purement quantique : à aucun moment nous n'avons mentionné une transmission d'information classique entre Alice et Bob.

7.6.2 Le transfert d'état quantique

L'idée du protocole

Contrairement à la borne de Holevo ou aux protocoles de téléportation quantique ou de codage superdense, le théorème de Schumacher est un résultat d'information quantique sans aucune référence à une information classique. Il donne un sens opérationnel à l'entropie de von Neumann et présente un protocole qui décrit comment Alice peut transférer de manière optimale l'information quantique dont elle dispose à Bob. Cependant, pour ce résultat, les seules corrélations quantiques qui existent avant transmission sont exclusivement entre Alice et l'environnement.

Pour discuter du transfert d'information quantique entre Alice et Bob de manière très générale, nous devons supposer qu'il existe un conditionnement quantique ou classique entre eux. Cette situation peut en toute généralité être décrite par une purification d'un état pur $|\Psi_{ABE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ partagé entre trois parties : Alice, Bob et un environnement (E). La question est alors de savoir s'il est possible de trouver un protocole qui, de manière asymptotique, permette de transférer les corrélations quantiques partagées entre les trois parties (A, B, E) sur uniquement deux d'entre elles (Bob et l'environnement) au moyen d'opérations locales effectuées par Alice et Bob et de l'utilisation d'un canal de communication quantique entre Alice et Bob. Et si oui, quel est le nombre minimal d'utilisations du canal quantique nécessaires pour réaliser ce transfert ?

Le *protocole de transfert d'état quantique* représenté schématiquement sur la figure 7.13 apporte une réponse à ces questions. Comme pour le théorème de Schumacher, il s'agit d'un résultat asymptotique, c'est-à-dire obtenu en considérant $N \gg 1$ copies de $|\Psi_{ABE}\rangle$. À ce résultat correspond une inégalité de ressources correspondante. Nous verrons que l'on peut en déduire un certain nombre de résultats sur la distillation de paires intriquées, la téléportation, le codage superdense bruité ainsi que la généralisation quantique de la borne de Slepian-Wolf qui est la question discutée dans la section 7.6.1.

Représentation en circuit

Le protocole de transfert d'état quantique peut être réalisé au moyen du circuit représenté sur la figure 7.14. Cette représentation est à la base de la preuve de l'existence de ce protocole.

Le point de départ est un état pur partagé $|\Psi_{ABE}\rangle \in \mathcal{H}_{ABE}$ entre Alice, Bob et leur environnement E . Dans la théorie de l'information quantique, on voit cet état comme préparé à l'avance, puis partagé entre ces trois parties. On le décrit par une ressource $\langle U^{S \rightarrow AB} : \psi_S \rangle$ qui signifie que nous disposons d'un opérateur extérieur capable d'envoyer isométriquement l'état $|\psi_S\rangle \in \mathcal{H}_S$ sur $|\Psi_{ABE}\rangle \in \mathcal{H}_{ABE}$. Ceci se fait au moyen d'un canal quantique entre d'un côté S et de l'autre côté Alice et Bob qui, sur l'état $|\psi_S\rangle$, se comporte comme l'isométrie $U^{S \rightarrow AB}$. Cette ressource, dite relative [132], décrit le partage de corrélations imparfaites entre Alice et Bob.

On considère alors $|\Psi_{ABE}\rangle^{\otimes N}$ pour $N \gg 1$. Alice effectue une compression de Schumacher sur son entrée et applique alors un opérateur unitaire bien choisi U_A . En sortie, elle transmet $N I_{\text{vn}}[A, E]/2$ qubits A_2 à Bob où la notation $I_{\text{vn}}[A, E]$ désigne l'information mutuelle quantique définie par l'équation (7.34) entre les parties A et E , évaluée dans l'état $\rho_{AE}^{\otimes N}$ avec $\rho_{AE} = \text{tr}_{\mathcal{H}_B} |\Psi_{ABE}\rangle\langle\Psi_{ABE}|$. Bob combine alors les qubits reçus avec ceux en sa possession pour les faire passer à travers un opérateur unitaire U_B . Le point remarquable, qui constitue le cœur de la démonstration [4], est que l'on peut choisir U_A et U_B de sorte que, asymptotiquement, le résultat de l'ensemble de ces opérations soit un état produit tensoriel $|\Psi_1^{(\text{out})}\rangle \otimes |\Psi_2^{(\text{out})}\rangle$ où le premier état vit sur les qubits A_1 qui restent en la possession d'Alice et une partie des qubits de Bob que l'on notera B_1 . En clair, il est possible de choisir l'opération U_A de manière à détruire toutes les corrélations entre Alice et l'environnement. Le second état qui vit dans $(\mathcal{H}_{B_2} \otimes \mathcal{H}_E)^{\otimes N}$ contient alors toutes les corrélations quantiques entre Alice, Bob et l'environnement. Asymptotiquement, cet état est proche de l'état $|\Psi_{ABE}\rangle^{\otimes N}$, ce qui signifie que l'on peut l'obtenir au moyen d'une ressource relative de S dans B_2 . Le premier état $|\Psi_1^{(\text{out})}\rangle$ qui vit dans $(\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1})^{\otimes N}$ est un état intriqué entre Alice et Bob encodé au travers de $N I_{\text{vn}}[A, B]/2$ paires maximales intriquées. Comme on l'a vu, cela correspond bien à la décomposition des qubits d'Alice donnée par l'équation (7.70).

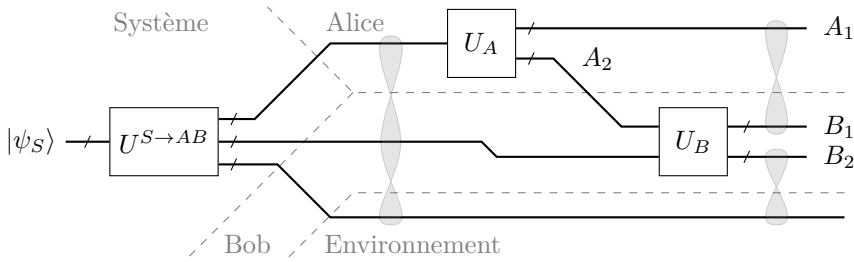


FIGURE 7.14 – Circuit du protocole de transfert d'état quantique. Après l'unitaire U_A , $I_{\text{vn}}[A, E]/2$ qubits sont transférés à Bob laissant $I_{\text{vn}}[A, B]/2$ qubits disponibles pour former des paires maximales intriquées avec Bob. À la fin, Bob dispose de toutes les corrélations quantiques initialement créées entre Alice & Bob d'une part et l'environnement E d'autre part. Les traits en biais sur les fils indiquent que l'on a plusieurs qubits sur chaque fil.

Remarquons que l'opération U_B permet de bien séparer en sortie d'une part les qubits de Bob portant les corrélations quantiques avec les qubits A_1 encore en possession d'Alice et d'autre part ceux portant les corrélations quantiques avec l'environnement. Le point clé de la démonstration est donc véritablement cette possibilité de « découpler » les corrélations quantiques partagées avec deux partenaires sur deux ensembles de qubits distincts.

L'inégalité en ressource du protocole de transfert

Au niveau des ressources, ce protocole nécessite donc au minimum $I_{\text{vn}}[A, E]$ utilisations d'un canal quantique parfait entre Alice et Bob et il reste à la fin $I_{\text{vn}}[A, B]$ paires intriquées disponibles. On aurait pu s'en douter : la discussion du cas (q_1) à la section 7.6.1 nous a montré que lorsque $S_{\text{vn}}[A|B] < 0$ un tel transfert d'état est réalisable en laissant des paires intriquées résiduelles en nombre $-S_{\text{vn}}[A|B]$. Le protocole de transfert s'exprime alors sous la forme de l'inégalité de ressources suivante :

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + \frac{1}{2} I_{\text{vn}}[A, E] [q \rightarrow q] \geq \frac{1}{2} I_{\text{vn}}[A, B] [qq] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle. \quad (7.71)$$

Dans cette inégalité $\langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle$ signifie qu'à la fin du protocole, les corrélations quantiques véhiculées par $|\Psi_{ABE}\rangle$ sont partagées entre l'environnement E et les qubits B_2 de Bob. Celui-ci dispose également de $I_{\text{vn}}[A, B]/2$ qubits B_1 qui sont maximales intriqués avec Alice.

Remarquons que l'on retrouve bien le théorème de codage de Schumacher lorsque $|\Psi_{ABE}\rangle = |\Psi_{AE}\rangle |\psi_B\rangle$, c'est-à-dire lorsque Alice et Bob ne partagent aucune corrélation. Dans ce cas, nous avons $I_{\text{vn}}[A, B] = 0$ et $I_{\text{vn}}[A, E] = 2S_{\text{vn}}[A]$ où $S_{\text{vn}}[A]$ est l'entropie de von Neumann de $\rho_A = \text{tr}_{\mathcal{H}_E}(|\Psi_{AE}\rangle\langle\Psi_{AE}|)$. L'inégalité de ressources (7.71) nous dit que l'on peut transférer l'état $|\Psi_{AE}\rangle$

sur des qubits de Bob au moyen de $S_{\text{vn}}[A]$ utilisations d'un canal quantique parfait entre Alice et Bob. Il ne reste alors aucune paire intriquée disponible entre Alice et Bob.

7.6.3 Conséquences du protocole de transfert

Passons maintenant en revue les résultats qui peuvent être dérivés du protocole de transfert. L'idée consiste à combiner celui-ci avec les protocoles élémentaires que sont la téléportation quantique, le codage superdense ou encore la distribution d'intrication. Comme nous allons le voir, les résultats correspondants sont les généralisations des inégalités de ressources discutées dans la section 7.2.4 aux situations réalistes, c'est-à-dire lorsque le lien entre Alice et Bob est un canal quantique et que les corrélations entre eux ne sont pas portées par des paires maximale-ment intriquées mais par un mélange statistique d'états purs générique.

Slepian-Wolf quantique

Rajoutons $I_{\text{vn}}[A, B][c \rightarrow c]$ aux deux membres de l'inégalité (7.71) et utilisons la téléportation $[qq] + 2[c \rightarrow c] \geq [q \rightarrow q]$ dans le membre de droite pour faire apparaître $I[A, B][q \rightarrow q]/2$. En utilisant l'équation (7.69), nous obtenons une inégalité de ressources dans laquelle toute référence explicite à l'environnement a été supprimée :

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + S_{\text{vn}}[A|B][q \rightarrow q] + I_{\text{vn}}[A, B][c \rightarrow c] \geq \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle. \quad (7.72)$$

Le protocole correspondant s'appelle le *protocole de fusion d'états quantiques* [207] : il permet de transférer un état partagé entre deux parties à une seule en utilisant des ressources de communication classiques et éventuellement quantiques (lorsque $S_{\text{vn}}[A|B] > 0$). C'est le protocole représenté sur la figure 7.12 pour le point $(q)_{1/2}$.

L'inégalité (7.72) permet de discuter l'analogue quantique du problème de Slepian-Wolf classique analysé à la section 2.5.4. Dans le cas quantique, Alice et Bob partagent des corrélations quantiques et cherchent à les transmettre à un tiers en utilisant deux canaux quantiques indépendants aboutissant à un même détecteur (figure 2.11(c)). On cherche alors à borner les taux de communication quantique R_A et R_B que chacun doit utiliser. Tout d'abord, le théorème de Schumacher pour la source (A, B) implique l'analogue de l'équation (2.70c) : $R_A + R_B \geq S_{\text{vn}}[A, B]$.

Les autres inégalités sont obtenues grâce au protocole de fusion d'états quantiques qui décrit ce qui se passe au sein du détecteur. En effet, l'inégalité (7.72) exprime quelles sont les ressources de communication qui doivent être mobilisées d'Alice vers Bob pour transférer l'état quantique ρ_{AB} sur Bob. Ce dernier disposant de l'ensemble des corrélations quantiques entre Alice et Bob, il peut alors transmettre cette information quantique vers une tierce personne. Lorsque $S_{\text{vn}}[A|B] \geq 0$, Alice doit utiliser au minimum $R_A = S_{\text{vn}}[A|B]$

qubits et $I_{\text{vn}}[A, B]$ bits classiques. Sous l'hypothèse que les ressources de communication classiques soient autorisées et gratuites, on retrouve ainsi une borne de Slepian-Wolf (2.70a) mais pour des taux de communication quantiques [207, 208]. Lorsque $S_{\text{vn}}[A|B] < 0$, l'interprétation de l'inégalité (7.72) est que la communication classique de $I_{\text{vn}}[A, B]$ bits a permis de se limiter à $R_B = 0$ et à $R_A = S_{\text{vn}}[A|B]$, la négativité de ce taux signifiant qu'il reste $-S_{\text{vn}}[A|B]$ possibilités de communication quantique ultérieures entre Alice et Bob.

On peut alors discuter l'analogie quantique de la figure 2.12 en considérant une situation où $S_{\text{vn}}[A] > 0$ et $0 < S_{\text{vn}}[B]/S_{\text{vn}}[A] \leq 1$. Comme indiqué sur la figure 7.10, nous avons alors deux régimes possibles :

— le régime dit classique, où

$$1 - \frac{S_{\text{vn}}[B]}{S_{\text{vn}}[A]} \leq \frac{S_{\text{vn}}[A|B]}{S_{\text{vn}}[A]} \leq 1, \quad (7.73)$$

dont la borne inférieure correspond au point (c_λ) de la figure 7.10 avec $\lambda = S_{\text{vn}}[B]/S_{\text{vn}}[A]$. En ce point, les taux (R_A, R_B) sont donnés par

$$R_A \geq S_{\text{vn}}[A] - S_{\text{vn}}[B], \quad (7.74a)$$

$$R_B \geq 0, \quad (7.74b)$$

$$R_A + R_B \geq S_{\text{vn}}[A]; \quad (7.74c)$$

— le régime dit quantique, défini par

$$1 - \frac{2S_{\text{vn}}[B]}{S_{\text{vn}}[A]} \leq \frac{S_{\text{vn}}[A|B]}{S_{\text{vn}}[A]} \leq 1 - \frac{S_{\text{vn}}[B]}{S_{\text{vn}}[A]}, \quad (7.75)$$

et dont la borne inférieure correspond à (q_λ) . En ce point, (R_A, R_B) sont alors donnés par

$$R_A \geq S_{\text{vn}}[A] - 2S_{\text{vn}}[B], \quad (7.76a)$$

$$R_B \geq -S_{\text{vn}}[B], \quad (7.76b)$$

$$R_A + R_B \geq S_{\text{vn}}[A] - S_{\text{vn}}[B]. \quad (7.76c)$$

La figure 7.15 représente les régions autorisées aux points (c_λ) , (q_λ) et pour un point intermédiaire situé entre les deux, mais dans la région quantique sur la figure 7.10. Pour toutes les valeurs de λ , (c_λ) correspond à l'extension maximale de la zone autorisée par la borne de Slepian-Wolf classique : $R_A \geq (1 - \lambda)S[A]$, $R_B \geq 0$ et $R_A + R_B \geq S[A]$. Comme ces taux de communication sont positifs, des ressources de communication quantique devront être consommées. Un point situé dans la région classique de la figure 7.10 donnerait lieu à une zone autorisée dans le plan (R_A, R_B) identique à celle représentée sur la figure 2.12. En revanche, au point (q_λ) ainsi qu'en tous les points de la zone quantique correspondant à un cas où Alice et Bob partagent des corrélations quantiques,

la zone autorisée comporte des régions où R_A ou R_B est négatif. C'est le cas sur la figure 7.15 où nous avons pris $1/2 < \lambda < 1$. Il existe donc des régions avec $R_A < 0$ et des régions avec $R_B < 0$. Pour $0 < \lambda \leq 1/2$, seul R_B pourrait être négatif. Dans ce cas, le sens opérationnel d'un point situé dans une de ces régions est qu'il existe un mode de fonctionnement qui laissera des possibilités ultérieures de communication quantique entre Alice et Bob. Cette discussion du problème de Slepian-Wolf quantique donne donc un sens opérationnel à la notion d'entropie conditionnelle quantique négative qui, contrairement à son analogue classique, ne s'appuie pas sur une notion de probabilité conditionnelle. Comme anticipé à la section 7.6.1, un partage de corrélations quantiques est bien équivalent à des ressources de communication quantique disponibles pour un usage ultérieur.

Enfin, remarquons que les protocoles de transfert et de fusion ont des objectifs très similaires : dans un cas comme dans l'autre, ils permettent à Alice de faire passer à Bob toutes les corrélations quantiques qu'elle partageait avec lui. La différence entre les deux est que le protocole de fusion implique l'utilisation d'un canal de communication classique entre les deux partenaires tandis que le protocole de transfert ne met en œuvre que des ressources quantiques. Pour cette raison, ce dernier est parfois appelé le *protocole de Slepian-Wolf complètement quantique*⁹.

Distillation de paires maximalement intriquées

Parmi les protocoles quantiques élémentaires, la distribution d'intrication permet de partager des paires maximalement intriquées entre Alice et Bob au moyen d'un canal de communication quantique parfait. Cependant dans la réalité, le canal sera bruité et les corrélations quantiques partagées entre Alice et Bob ne correspondront pas à celles de paires maximalement intriquées. Il est donc intéressant de disposer d'un protocole qui permette d'extraire des paires maximalement intriquées à partir de corrélations générales entre Alice et Bob. C'est ce qu'on appelle un *protocole de distillation de paires maximalement intriquées*.

On peut obtenir l'inégalité de ressources correspondante en utilisant l'inégalité de téléportation quantique $[qq] + 2[c \rightarrow c] \geq [q \rightarrow q]$ dans le membre de gauche de l'inégalité (7.71) pour remplacer les utilisations de canaux quantiques parfaits par une utilisation de canaux classiques et de paires intriquées. Pour réécrire le membre de droite de l'inégalité ainsi obtenue

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + \frac{1}{2} I_{\text{vn}}[A, E][qq] + I_{\text{vn}}[A, E][c \rightarrow c] \geq \frac{1}{2} I_{\text{vn}}[A, B][qq] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle, \quad (7.77)$$

9. *Fully Quantum Slepian-Wolf protocol* en anglais.

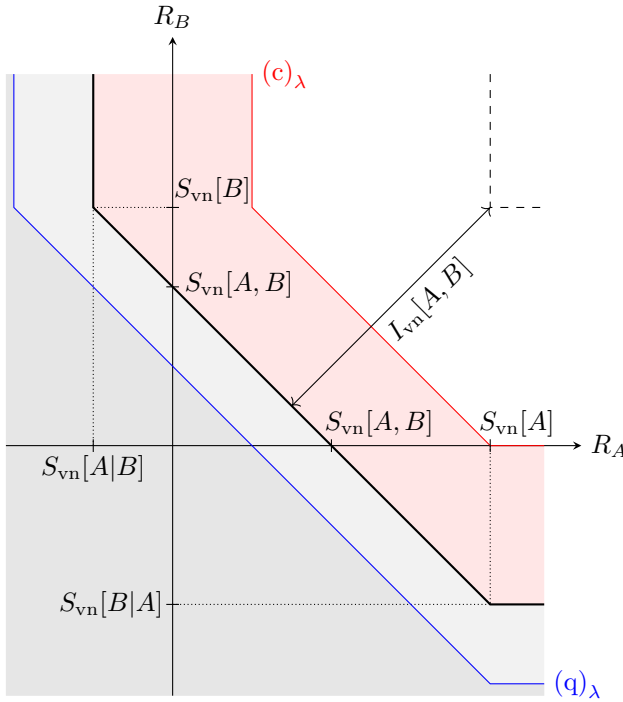


FIGURE 7.15 – Représentation des zones délimitées par la borne de Slepian-Wolf quantique pour $S[B]/S[A] = \lambda = 3/4$ (même valeur que pour la figure 2.12). Ligne continue noire : cas où $S[A, B]$ est plus grand que le minimum donné par la borne d’Araki-Lieb (7.47) mais plus petit que la borne inférieure classique $\max(S[A], S[B])$. Zone gris clair : zone exclue par la borne de Slepian-Wolf quantique dans l’exemple considéré ici. Ligne rouge (point $(c)_{3/4}$ sur la figure 7.10) : extension maximale de la zone autorisée par la borne de Slepian-Wolf dans le cas classique où $S[A, B] \geq \max(S[A], S[B])$. Zone rouge clair : région autorisée par la borne de Slepian-Wolf quantique mais inatteignable classiquement. Ligne bleue (point $(q)_{3/4}$) : cas où $S[A, B]$ est égal au minimum autorisé par la borne d’Araki-Lieb, *i.e.* $S[A] - S[B]$. Zone gris foncé : zone interdite y compris en exploitant les corrélations quantiques maximales possibles entre A et B . Les deux lignes pointillées orthogonales correspondent au cas d’encodages et décodages indépendants ($R_A = S_{\text{vn}}[A]$ et $R_B = S_{\text{vn}}[B]$).

nous utilisons l’équation (7.69), ce qui conduit à l’inégalité de ressources

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + I_{\text{vn}}[A, E] [c \rightarrow c] \geq -S_{\text{vn}}[A|B] [qq] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle. \quad (7.78)$$

Lorsque $S_{\text{vn}}[A|B] < 0$, cette inégalité montre comment à partir d’un mélange d’états quantiques partagés entre Alice et Bob, il est possible, en faisant

passer $I_{\text{vn}}[A, E]$ bits classiques d'Alice à Bob de générer $-S_{\text{vn}}[A|B] > 0$ paires maximalement intriquées. Une entropie conditionnelle quantique négative apparaît donc bien comme le nombre de paires maximalement intriquées disponibles qu'il subsistera après que l'on aura, au moyen du protocole de fusion d'états quantiques, transféré l'information quantique contenue dans ρ_{AB} à Bob [207]. Pour reprendre le mot de Hayden, une entropie quantique conditionnelle négative est une « plus que certitude » distillable en paires maximalement intriquées. Remarquons cependant qu'une telle opération a un coût en termes de communication classique donné par $I_{\text{vn}}[A, E]$.

Ce coût est nul lorsque Alice et Bob partagent un état pur intriqué, vu que dans ce cas $I_{\text{vn}}[A, E] = 0$. Comme $S_{\text{vn}}[A, B] = 0$ dans cette situation, il est possible de distiller au plus $S_{\text{vn}}[A]$ paires maximalement intriquées, un nombre en général inférieur à la valeur maximale $\log_2(\dim(\mathcal{H}_A))$ que l'on pourrait atteindre en partageant un état maximalement intriqué. En termes de ressources, ce cas particulier se résume par l'inégalité de ressources $\langle \Psi_{AB} \rangle \geq S_{\text{vn}}[A] [qq]$ où la ressource $\langle \Psi_{AB} \rangle$ désigne le partage de l'état pur $|\Psi_{AB}\rangle$ entre Alice et Bob. Le point remarquable est que dans ce cas, il existe une inégalité réciproque appelée théorème de distillation inverse [260] qui nous dit que $S_{\text{vn}}[A] [qq] \geq \langle \Psi_{AB} \rangle$. En conséquence, nous avons l'équivalence de ressources

$$\langle \Psi_{AB} \rangle \equiv S_{\text{vn}}[A] [qq]. \quad (7.79)$$

Cette équivalence correspond à une situation que Mark Wilde [379, Partie V] appelle non bruitée c'est-à-dire impliquant des états purs et des protocoles élémentaires. Elle exprime que le nombre $S_{\text{vn}}[A]$ de paires maximalement intriquées distillables caractérise la « valeur » de l'état pur $|\Psi_{AB}\rangle$ partagé entre Alice et Bob. En termes physiques, Alice et Bob peuvent, au moyen d'opérations locales, transformer un grand nombre $N \gg 1$ de copies d'un état intriqué pur qu'ils partagent en $NS_{\text{vn}}[A]$ paires maximalement intriquées et, de manière réciproque, ils peuvent produire n'importe quel état pur intriqué partagé en utilisant comme seule ressource non locale des paires maximalement intriquées [39].

En pratique, la distillation quantique est importante pour les protocoles de communication et de cryptographie quantiques modernes [148, 72] basés sur le partage de paires maximalement intriquées. Comme il est difficile d'assurer la stabilité et la qualité de préparation de telles paires, surtout si elles sont distribuées sur de très grandes distances, la distillation d'intrication est un point important pour les communications quantiques sécurisées.

Téléportation quantique et codage superdense généralisés

Nous allons maintenant discuter la généralisation des protocoles de téléportation quantique et de codage superdense au cas où Alice et Bob partagent des corrélations quantiques imparfaites. Même dans ce cas, le ratio de 2 bits classiques pour 1 bit quantique reste valable en toute généralité.

L'inégalité (7.78) associée à la distillation de paires quantifie combien de paires maximalement intriquées peuvent être générées à partir de l'utilisation d'un canal classique. Or la téléportation quantique nous dit que la communication de deux bits classiques par paire intriquée nous permet de communiquer un qubit. En ajoutant $-S_{\text{vn}}[A|B][c \rightarrow c]$ à l'inégalité (7.78) et en utilisant l'inégalité de téléportation quantique $2[c \rightarrow c] + [qq] \geq [q \rightarrow q]$ dans le membre de droite ainsi que l'équation (7.69) dans le membre de gauche, nous obtenons la généralisation de l'inégalité de ressource associée à la téléportation quantique

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + I_{\text{vn}}[A, B][c \rightarrow c] \geq -S_{\text{vn}}[A|B][q \rightarrow q] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle. \quad (7.80)$$

De même, en combinant l'inégalité (7.71) avec l'inégalité de codage superdense, et en utilisant l'équation (7.70), nous arrivons à l'inégalité de ressources

$$\langle U^{S \rightarrow AB} : \psi_S \rangle + S_{\text{vn}}[A][q \rightarrow q] \geq I_{\text{vn}}[A, B][c \rightarrow c] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle \quad (7.81)$$

qui généralise le codage superdense lorsque Alice et Bob partagent un mélange statistique général (codage superdense bruité). L'information mutuelle quantique $I_{\text{vn}}[A, B]$ est ici le nombre de bits classiques que l'on peut transmettre d'Alice à Bob en utilisant à la fois les corrélations quantiques entre Alice et Bob et un codage optimal des qubits en possession d'Alice. Enfin, comme la borne d'Araki-Lieb (7.47) entraîne que l'information mutuelle quantique est majorée par $2 \min(S_{\text{vn}}[A], S_{\text{vn}}[B])$ au lieu de $\min(S_{\text{vn}}[A], S_{\text{vn}}[B])$ classiquement, on retrouve en toute généralité le facteur 2 du codage super-dense idéal et de la téléportation idéale.

7.6.4 Protocoles duaux

Jusqu'à présent, nous avons discuté du protocole de transfert d'état quantique qui part d'un état pur tripartite entre Alice, Bob et leur environnement et qui relocalise les corrélations sur Bob. Mais un canal de communication quantique fonctionne en étalant une information quantique initialement en possession d'Alice sur Bob et son environnement. Discuter sa capacité quantique conduit donc à rechercher de nouvelles inégalités en ressources qui impliquent un étalement de l'intrication, que ce soit au travers de l'extension du partage d'un état intriqué entre Alice et Bob à l'environnement ou au travers de l'utilisation d'un canal quantique.

Renversement du temps

Le protocole de transfert d'état quantique détruit toutes les corrélations quantiques entre Alice et son environnement. En sortie, les corrélations tripartites sont concentrées entre les mains de Bob et son environnement. En renversant le sens du temps sur la représentation en circuit du protocole de transfert d'état quantique représentée sur la figure 7.14, on obtient un nouveau protocole qui, à partir de corrélations quantiques exclusivement bipartites,

génère des corrélations quantiques tripartites. Concrètement, si Alice et Bob partagent $I[A, B]/2$ paires parfaitement intriquées et que Bob dispose d'un mélange d'états indépendants, ils peuvent générer un état tripartite pur avec leur environnement en utilisant juste des opérations locales et un canal de communication parfait entre eux. Ce protocole se traduit par l'inégalité de ressources

$$\frac{1}{2}I_{\text{vn}}[A, B][qq] + \langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle + \frac{1}{2}I_{\text{vn}}[A, E][q \rightarrow q] \geq \langle U^{S \rightarrow AB} : \psi_S \rangle, \quad (7.82)$$

où dans cette équation $I_{\text{vn}}[A, E]$ correspond à l'information quantique mutuelle qui sera présente dans l'état intriqué final engendré par $\langle U^{S \rightarrow AB} : \psi_S \rangle$.

Par ailleurs, on peut voir la ressource $\langle U^{S \rightarrow AB} : \psi_S \rangle$ comme la composition d'une ressource $\langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle$ et d'une isométrie $U^{B_2 \rightarrow AB}$. Cela s'exprime en termes de ressources par l'équivalence

$$\langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle + \langle U^{B_2 \rightarrow AB} : \rho_{B_2} \rangle \equiv \langle U^{S \rightarrow AB} : \psi_S \rangle \quad (7.83)$$

qui permet, en simplifiant, d'obtenir l'inégalité dite de Shannon inverse quantique :

$$\frac{1}{2}I_{\text{vn}}[A, B][qq] + \frac{1}{2}I_{\text{vn}}[A, E][q \rightarrow q] \geq \langle U^{B_2 \rightarrow AB} : \rho_{B_2} \rangle. \quad (7.84)$$

Celle-ci quantifie les ressources quantiques idéales nécessaires pour simuler un canal de communication quantique réaliste. L'état tripartite $|\Psi_{ABE}\rangle$ est ici obtenu à partir d'un état produit tensoriel $|\psi_{B_2E}\rangle |\psi_{B_1A_1}\rangle$ où $|\psi_{B_2E}\rangle$ est une purification de ρ_{B_2} et où $|\psi_{B_1A_1}\rangle$ contient l'intrication entre Alice et Bob qui sera consommée dans la création de l'état tripartite.

La valeur d'un état mélange partagé

Il est intéressant de rapprocher cette inégalité de ressources de l'inégalité de ressources associée au transfert d'état quantique.

Pour cela, partons du protocole de transfert d'état quantique pour lequel le point de départ est un état pur tripartite $|\Psi_{ABE}\rangle$. Alors que dans l'inégalité (7.71), on le voit comme injecté depuis une source S via un canal imparfait de S vers AB , on peut également considérer qu'on l'obtient par l'action d'un canal quantique $U^{A_1 \rightarrow AB}$ à partir d'un opérateur densité ρ_{A_1} qui provient lui-même de l'injection de $|\psi_S\rangle$ sur A_1 au moyen d'un canal imparfait $\mathbb{1}^{S \rightarrow A_1}$. De même, la ressource relative $\langle \mathbb{1}^{S \rightarrow B_2} : \psi_S \rangle$ peut être vue comme la composition de la ressource $\langle \mathbb{1}^{S \rightarrow A_1} : \psi_S \rangle$ par la ressource $\langle \mathbb{1}^{A_1 \rightarrow B} : \rho_{A_1} \rangle$. En utilisant les règles de simplification, cela amène à l'inégalité de ressources suivante :

$$\langle U^{A_1 \rightarrow AB} : \rho_{A_1} \rangle + \frac{1}{2}I_{\text{vn}}[A, E][q \rightarrow q] \geq \frac{1}{2}I_{\text{vn}}[A, B][qq] + \langle \mathbb{1}^{A_1 \rightarrow B} : \rho_{A_1} \rangle. \quad (7.85)$$

Dans cette inégalité, la ressource $\langle U^{A_1 \rightarrow AB} : \rho_{A_1} \rangle$ représente la transmission d'un état ρ_{A_1} qui possède toutes les corrélations quantiques entre le duo Alice et Bob et l'environnement : on peut le purifier en un état $|\Psi_{A_1, E}\rangle$ qui est équivalent à $|\Psi_{ABE}\rangle$. On a donc réécrit le protocole de transfert en faisant apparaître comme élément de départ des canaux quantiques au lieu d'une source émettant des états tripartites $|\Psi_{ABE}\rangle$. Cette opération est appelée une dualité canal/source.

Par ailleurs, le théorème de Schumacher nous dit que la transmission de l'opérateur densité ρ_{A_1} nécessite $S_{\text{vn}}[A, B] = S_{\text{vn}}[E]$ utilisations d'un canal quantique parfait, ce qui se traduit par l'inégalité

$$\langle \mathbb{1}^{A_1 \rightarrow AB} : \rho_{A_1} \rangle \geq S_{\text{vn}}[E] [q \rightarrow q]. \quad (7.86)$$

Celle-ci permet, en utilisant l'égalité $S_{\text{vn}}[E] - I_{\text{vn}}[A, E]/2 = I_{\text{vn}}[B, E]/2$, qui n'est autre que l'équation (7.70) après permutation de A et E , d'arriver à l'inégalité de ressources

$$\langle U^{A_1 \rightarrow AB} : \rho_{A_1} \rangle \geq \frac{1}{2} I_{\text{vn}}[A, B] [qq] + \frac{1}{2} I_{\text{vn}}[E, B] [q \rightarrow q]. \quad (7.87)$$

Après échange des rôles d'Alice et Bob, cette inégalité, correspondant au protocole père avec rétro-action¹⁰ [142] est exactement la renversée de l'inégalité de Shannon inverse quantique (7.84). En conséquence, nous venons d'obtenir une équivalence entre d'une part une ressource quantique bruitée et d'autre part deux ressources quantiques inéquivalentes¹¹ :

$$\langle U^{A_1 \rightarrow AB} : \rho_{A_1} \rangle \equiv \frac{1}{2} I_{\text{vn}}[B, E] [q \rightarrow q] + \frac{1}{2} I_{\text{vn}}[A, B] [qq]. \quad (7.88)$$

Ce résultat généralise l'équivalence (7.79) qui exprime la valeur d'un état pur intriqué partagé entre Alice et Bob.

Le protocole père

Les différents protocoles étudiés jusqu'ici impliquent, soit une source générant un état tripartite $|\Psi_{ABE}\rangle$ (section 7.6.2), soit une source transportant un état intriqué entre Bob et l'environnement vers Alice et Bob (section 7.6.4). Mais l'objectif de la prochaine section est de discuter des capacités d'un canal quantique $\mathcal{N}^{A \rightarrow B}$. Nous allons donc avoir besoin d'inégalités de ressources qui combinent le canal quantique avec les ressources élémentaires que sont les paires maximales intriquées ou le canal quantique parfait.

Le *protocole père* [131] correspond à l'inégalité de ressources

$$\langle \mathcal{N}^{A \rightarrow B} : \rho_A \rangle + \frac{1}{2} I_{\text{vn}}[A, E] [qq] \geq \frac{1}{2} I_{\text{vn}}[A, B] [q \rightarrow q] \quad (7.89)$$

10. Le terme de rétro-action s'explique par la présence d'un canal quantique entre $A_1 \subset A$ et AB et donc qui permet d'effectuer une rétro-action sur Alice.

11. Rappelons que la ressource canal quantique est strictement plus puissante que la ressource paire intriquée partagée.

qui exprime que l'on peut produire de la communication quantique idéale à partir d'un canal réaliste bruité en y adjoignant des paires intriquées. C'est un premier exemple d'inégalité qui exprime le gain apporté à la communication quantique par l'adjonction de la ressource intrication.

En combinant (7.89) avec les trois protocoles primaires que sont le codage superdense, la téléportation quantique et la distribution d'intrication, nous obtenons respectivement les trois inégalités suivantes :

$$\langle \mathcal{N}^{A \rightarrow B} \rangle + I_{\text{vn}}[A, E] [c \rightarrow c] \geq -S_{\text{vn}}[A|B] [qq], \quad (7.90)$$

$$\langle \mathcal{N}^{A \rightarrow B} : \rho_A \rangle + S_{\text{vn}}[A] [qq] \geq I_{\text{vn}}[A, B] [c \rightarrow c], \quad (7.91)$$

$$\langle \mathcal{N}^{A \rightarrow B} : \rho_A \rangle \geq -S_{\text{vn}}[A|B] [q \rightarrow q]. \quad (7.92)$$

La première nous dit qu'en utilisant conjointement le canal quantique bruité et un canal classique, on peut engendrer des paires maximalement intriquées. La seconde exprime comment l'adjonction de paires maximalement intriquées augmente la capacité de communication classique d'un canal à au moins $I_{\text{vn}}[A, B]$ qui, rappelons-le, peut être strictement plus élevé que la corrélation classique maximale entre les deux sources A et B . Enfin, la dernière inégalité nous apprend qu'un canal quantique est capable de transmettre au moins $-S_{\text{vn}}[A|B]$ qubits. Cette inégalité justifie que la quantité $-S_{\text{vn}}[A|B]$ soit appelée l'information cohérente [325] lorsque $S_{\text{vn}}[A|B] \leq 0$: elle quantifie la capacité de transport idéale d'un canal quantique réaliste. Pour résumer, les inégalités (7.89) et (7.91) expriment le bénéfice apporté par l'utilisation d'intrication pour la communication quantique et pour la communication classique alors que les inégalités (7.90) et (7.92) expriment combien de ressources quantiques idéales on peut extraire d'un canal quantique réaliste.

7.6.5 Capacité d'un canal quantique

Nous allons maintenant aborder la notion de *capacité d'un canal quantique*, faisant ainsi écho à la discussion de la section 2.5 dans le cas classique. Un point important est que plusieurs notions de capacité peuvent être associées à un canal de communication quantique. On peut en effet l'utiliser pour véhiculer une information classique ou bien pour véhiculer une information quantique, ce qui conduit naturellement à discerner capacités classique et quantique d'un canal quantique. De plus, au fil de ce chapitre, nous avons appris trois choses importantes pour les communications quantiques :

- utiliser des états mélanges ou des états non-orthogonaux conduit nécessairement à des performances de communication inférieures à celles autorisées par l'utilisation d'une famille orthogonale d'états purs (section 7.3.2) ;
- il est possible de transférer des états quantiques entre deux parties et, parfois, cela permet de disposer de paires maximalement corrélées supplémentaires (section 7.6.2) ;

- même si l'intrication quantique ne permet pas de communiquer à elle seule, elle modifie les capacités de communication tant classique que quantique (sections 7.2.2 et 7.6.4).

Il est donc naturel de se demander quelles sont les capacités de communication classique et quantique d'un canal quantique. Cependant la question est plus subtile, y compris lorsque l'on ne cherche qu'à caractériser la capacité de communication classique.

En effet, nous avons vu dans la section 4.3.3 qu'effectuer des mesures collectives pouvait significativement améliorer les performances de tomographie. On peut donc considérer des situations dans lesquelles on s'autorise ou non à créer de l'intrication en entrée du canal (côté Alice) et dans lesquelles on s'autorise ou non des mesures collectives en sortie (côté Bob) comme représenté sur la figure 7.16. Le panneau (a) correspond au cas où on ne cherche à utiliser ni l'avantage liée à des corrélations quantiques entre plusieurs envois, ni aux mesures collectives. Le cas (b) est celui où on ne cherche qu'à tirer avantage de mesures collectives lors du décodage. Le cas (c) est exactement l'opposé du cas (a) : on utilise les deux. Nous allons déterminer quelle est la hiérarchie entre les capacités de communication dans les trois modes de communication représentés sur la figure 7.16.

Le troisième enseignement ci-dessus soulève une dernière question importante : celle du gain lorsque l'on s'autorise à utiliser des paires intriquées en plus du canal de communication. Là aussi, nous allons voir les principaux résultats obtenus. Le lecteur intéressé par le détail des preuves et par une revue plus complète de ce domaine consultera l'ouvrage de Mark Wilde [379, Partie VI] ou bien l'article de revue [190].

Capacités classiques d'un canal quantique

Alice cherche à transmettre une information classique x encodée par une variable aléatoire classique x de loi $p_A(x)$ en injectant des états (purs ou mélanges) quantiques $\rho(A|x)$ en entrée d'un canal de transmission quantique $\mathcal{N}$ entre elle et Bob. Pour chaque x , Bob reçoit un état quantique qui est décrit par l'opérateur densité $\rho(B|x)$ obtenu par action du super-opérateur décrivant le canal $\mathcal{N}$ sur $\rho(A|x)$. L'entropie de Shannon $S[A] = S[p_A]$ n'est autre que l'entropie de préparation correspondant à l'information classique qu'elle cherche à transmettre.

De manière générale, l'information classique maximale que l'on peut récupérer en sortie est $\max_R(I[A, R])$ où la minimisation est prise sur R qui est le protocole de mesure en sortie du canal $\mathcal{N}$ et $I[A, R]$ désigne l'information mutuelle classique entre la distribution statistique des choix d'Alice et les résultats de mesure classique de R . Comme dans la théorie classique, la capacité du canal est alors obtenue en maximisant l'information accessible en sortie sur les entrées possibles :

$$C_c^{(\text{seq})}[\mathcal{N}] = \max_A (\text{Acc}[(p_A(x), \rho(B|x))]) . \quad (7.93)$$

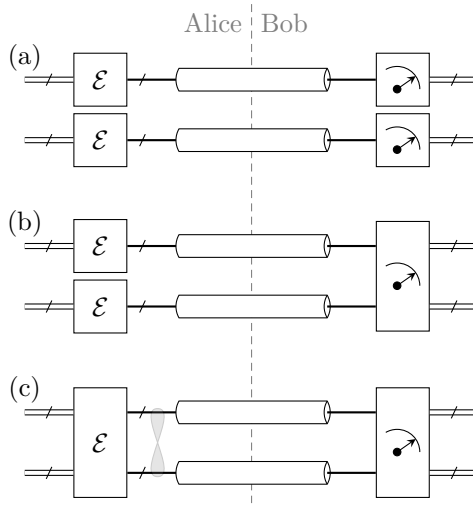


FIGURE 7.16 – Les différentes manières d'utiliser un canal de communication quantique pour véhiculer de l'information classique. (a) Utilisation séquentielle : Alice et Bob ne s'autorisent pas à utiliser l'intrication pour transmettre $N > 1$ copies ni au niveau de l'encodeur, ni au niveau du décodeur. (b) Décodage collectif seulement : Bob s'autorise des mesures collectives en sortie des N canaux en parallèle. (c) Complètement collectif : Alice s'autorise un encodeur qui crée de l'intrication en entrée des N copies de $\mathcal{N}$ et Bob s'autorise un décodage par mesure collective en sortie.

Cette définition suppose que, dans une utilisation multiple du canal quantique, les opérateurs ne créent pas de corrélations quantiques entre les différentes copies $\rho(A|x_j)$ pour $j = 1, \dots, N$ avant transmission ni ne s'autorisent de mesure collective sur ce qui sort du canal. Pour $(x_1, \dots, x_N)$, Alice injecte en entrée du canal $\bigotimes_{j=1}^N \rho(A|x_j)$ et Bob reçoit l'environnement de $\mathcal{N}^{\otimes N}(\rho(A|x_1) \otimes \dots \otimes \rho(A|x_N))$. On parle alors de capacité de Shannon du canal quantique. Elle correspond à la capacité de transmission classique dans la situation représentée sur le panneau (a) de cette figure.

Le second résultat d'importance est dû à Holevo [205], Westmoreland et Schumacher [326]. Appelé *théorème HSW*, il donne la capacité classique du canal quantique lorsque l'on s'autorise des mesures collectives en sortie mais que l'on utilise des états factorisés en entrée, c'est-à-dire dans la situation du panneau (b) de la figure 7.16. Cela définit alors la capacité de Holevo du canal qui est donnée par

$$\chi[\mathcal{N}] = \max_A \left(S_{\text{vn}} \left[\mathcal{N} \left(\sum_x p_A(x) \rho(A|x) \right) \right] - \sum_x p_A(x) S_{\text{vn}} [\mathcal{N}(\rho(A|x))] \right). \quad (7.94)$$

Remarquons que cette quantité est, en vertu de l'inégalité (7.27), supérieure à la capacité de Shannon du canal $C_c^{(\text{seq})}[\mathcal{N}]$. Le théorème HSW montre que l'exploitation d'effets quantiques au niveau du décodeur permet d'aboutir à une capacité classique du canal quantique supérieure à ce qu'on pourrait obtenir dans une théorie purement classique. C'est une expression directe de la supériorité des collectives dans le cas quantique qui fait écho à la discussion de la section 4.3.3.

Cependant, ce n'est pas encore le cas le plus général puisque l'on ne s'autorise à exploiter l'intrication qu'au niveau du décodeur. Considérons la situation représentée sur le panneau (c) de la figure 7.16 avec un encodeur qui génère de l'intrication entre les N entrées de $\mathcal{N}^{\otimes N}$. Dans ce cas, la capacité du canal, lorsque l'on s'autorise des mesures collectives sur la sortie et une entrée intriquée, est définie en considérant la capacité asymptotique correspondant à un très grand nombre d'utilisations du canal quantique :

$$C_c^{(\infty)}[\mathcal{N}] = \lim_{N \rightarrow +\infty} \left(\frac{1}{N} \chi[\mathcal{N}^{\otimes N}] \right), \quad (7.95)$$

où $\chi[\mathcal{N}^{\otimes N}]$ correspond à la capacité de Holevo du canal obtenu en mettant en parallèle N canaux physiques identiques et en s'autorisant un encodeur et un décodeur collectif. La subtilité de cette définition provient du fait que la quantité de Holevo associée à un canal quantique n'est pas additive. Rappelons qu'une quantité Ξ sera dite additive si et seulement si

$$\Xi[\mathcal{N}_1 \otimes \mathcal{N}_2] = \Xi[\mathcal{N}_1] + \Xi[\mathcal{N}_2]. \quad (7.96)$$

Si la capacité de Holevo d'un canal quantique était additive, la capacité $C_c^{(\infty)}[\mathcal{N}]$ coïnciderait avec $\chi[\mathcal{N}]$. Comme l'a montré Shor [338], l'additivité de plusieurs quantités importantes dont la capacité de Holevo du canal quantique est équivalente : soit toutes ces quantités sont additives, soient elles ne le sont pas. Or, Hastings a montré que la capacité de Holevo, que l'on a longtemps considérée comme additive, ne l'était pas [198]. Ceci montre que la capacité classique d'un canal quantique, définie par l'équation (7.95), n'est pas une quantité facile à déterminer. Shor et Beigi ont d'ailleurs montré que le calcul de la capacité de Holevo d'un canal quantique était un problème NP-complet et donc calculer $\chi[\mathcal{N}^{\otimes N}]$ à la limite $N \rightarrow +\infty$ s'avère très coûteux en calcul [35].

En conclusion, nous voyons que, même au niveau de la transmission d'information classique, l'utilisation de canaux de communication quantiques donne lieu à une théorie bien plus riche que celle de Shannon présentée au chapitre 2. Ces différentes capacités classiques sont ordonnées selon l'inégalité

$$C_c^{(\text{seq})}[\mathcal{N}] \leq \chi[\mathcal{N}] \leq C_c^{(\infty)}[\mathcal{N}]. \quad (7.97)$$

Capacité classique privée d'un canal quantique

Un problème intéressant consiste à voir si on peut utiliser un canal quantique pour communiquer de l'information de manière privée. Rappelons déjà le

résultat correspondant dans le cas où on utilise un canal classique $\mathcal{N}^{(\text{cl})}$ entre Alice et Bob avec une interceptrice, Ève : Csizsar et Körner ont montré [112] qu'elle est donnée par la différence d'informations mutuelles classiques

$$P_c \left[\mathcal{N}^{(\text{cl})} \right] = \max_A (I[A, B] - I[B, E]) . \quad (7.98)$$

Ce résultat est intuitif : la capacité privée est obtenue en retirant aux corrélations entre Alice et Bob les corrélations entre Bob et Ève, et en maximisant sur ce que l'on envoie en entrée. On peut montrer que le résultat est positif dans tous les cas et que, dans la théorie de Shannon classique, c'est une quantité additive.

Il se trouve que la capacité classique privée d'un canal quantique $\mathcal{N}^{(\text{qu})}$ est également connue [129, 78]. Lorsque ni Alice, ni Bob, ni l'espionne Ève ne s'autorisent des mesures collectives, elle est donnée par :

$$P_c \left[\mathcal{N}^{(\text{qu})} \right] = \max_A (I_{\text{vn}}[A, B] - I_{\text{vn}}[A, E]) . \quad (7.99)$$

En revanche, si Ève s'autorise ce que l'on appelle des manipulations et mesures collectives sur l'ensemble des messages transmis, dans ce cas, il faut remplacer $I_{\text{vn}}[A, E]$ par la quantité de Holevo $\chi[\mathcal{N}^{A \rightarrow E}]$, ce qui diminue la capacité privée. Ces résultats sont directement pertinents pour quantifier les performances d'un système de distribution de clé privée comme le protocole BB84 (voir section 7.2.1).

Capacité quantique

Passons maintenant à la capacité quantique d'un canal quantique. L'inégalité (7.92) nous donne une borne inférieure pour la capacité d'un canal quantique. Le résultat remarquable, appelé *théorème LSD* du nom de ses auteurs [259, 337, 129], stipule que cette borne est atteignable :

$$C_q^{(\infty)}[\mathcal{N}] = - \max_A (S_{\text{vn}}[A|B]) , \quad (7.100)$$

où la capacité quantique est également définie de manière asymptotique. Récemment, les techniques qui ont permis de démontrer l'inégalité de ressources (7.71) ont permis d'obtenir une preuve plus simple de ce résultat. L'idée consiste à montrer qu'il existe une manière d'encoder les données en entrée du canal de sorte que, à la limite $N \rightarrow +\infty$, les données transmises s'intriquent avec l'environnement avec une probabilité quasi nulle [199]. Remarquons, dans ce cas, on obtient une expression relativement simple pour la capacité quantique, plus facile à calculer en tous cas que dans le cas de la capacité $C_c^{(\infty)}[\mathcal{N}]$.

Capacités assistées par intrication

Le codage superdense nous apprend que, dans le cas d'un canal parfait, lui adjoindre de l'intrication pouvait doubler sa capacité : si un canal quantique

parfait ne peut transmettre qu'un qubit et donc au plus un bit classique, un tel canal associé à une paire intriquée peut en transmettre deux. Remarquons qu'utiliser des paires intriquées supplémentaires ne permet pas de gagner quoi que ce soit vu qu'une telle paire, en soi, ne permet pas de transmettre de l'information.

Il n'est donc pas surprenant que l'on ait cherché très vite à déterminer la capacité de communication classique $C_{c,E}[\mathcal{N}]$ d'un canal quantique imparfait $\mathcal{N}$ assisté par un nombre potentiellement illimité de paires intriquées partagées entre l'émetteur et le récepteur. Le point remarquable est que dans ce cas, on peut évaluer explicitement la capacité asymptotique, alors qu'on ne le pouvait pas pour la capacité $C_c^{(\infty)}[\mathcal{N}]$. Le résultat [43] est l'exact analogue de la capacité d'un canal classique bruité :

$$C_{c,E}^{(\infty)}[\mathcal{N}] = \max_A (I_{\text{vn}}[A, B]) , \quad (7.101)$$

où A désigne ce qui est envoyé en entrée côté Alice et B désigne la sortie correspondant à $B = \mathcal{N}(A)$. C'est l'exact analogue de l'équation (2.67) obtenue dans le chapitre 2 sauf que ici apparaît l'information quantique mutuelle entre l'entrée et la sortie du canal. C'est un résultat étonnamment simple et un des rares pour lequel on ait une expression aussi simple à la limite asymptotique. En fait, cette capacité peut s'obtenir en considérant l'entrée $\rho_A^{\otimes n}$ à la limite $n \rightarrow +\infty$ et en rajoutant des paires intriquées en nombre tel que l'on puisse transmettre deux bits classiques par qubit transmis au moyen du codage superdense. L'évaluation de l'information transmise découle alors de la formule de Holevo pour la capacité classique d'un canal quantique.

7.6.6 Perspectives

En résumé, grâce à une approche de théorie des ressources, nous avons pu répondre aux questions que nous nous posions sur les possibilités de transmission et les équivalences entre ressources quantiques dans la limite asymptotique. Ces résultats s'appuient sur des protocoles permettant la transformation de ressources en d'autres et les inégalités de ressources correspondantes quantifient les ressources qui doivent être mobilisées pour mettre en œuvre ces protocoles. De manière remarquable, les différents protocoles s'organisent en un « arbre généalogique » représenté sur la figure 7.11 [4].

Cependant, cela ne veut pas pour autant dire que la théorie de Shannon quantique soit aussi achevée que son pendant classique présenté au chapitre 2. Suivant que l'on s'autorise ou non à exploiter l'intrication en entrée ou en sortie d'un canal de communication quantique lors d'une utilisation multiple (N fois), on obtient différentes capacités de communication (classiques ou quantiques). Leur évaluation est étroitement liée à l'additivité des quantités qui apparaissent dans la théorie de Shannon quantique. Hélas, il n'y a que pour les capacités assistées par intrication que l'additivité permet de contourner les difficultés. D'un point de vue purement théorique, cela peut amener à se

demander si une stratégie consistant à considérer des problèmes plus proches des applications ne serait pas appropriée.

Une direction de recherche actuelle est ainsi le développement d'algorithmes d'optimisation pour des codes de compression quantique et pour le calcul de quantités pertinentes pour l'information quantique. De tels algorithmes pourraient permettre de calculer efficacement les capacités et les codes optimaux à N fini, c'est-à-dire dans les cas d'importance pratique. Une autre direction de recherche très active est la théorie des réseaux de communication quantiques qui est encore largement en développement.

Chapitre 8

Calculer en utilisant des qubits

Dans le chapitre 3, nous avons vu comment définir ce qu'est une fonction calculable, et comment quantifier la complexité des algorithmes. Pour cela, nous avons introduit des classes de complexité correspondant aux problèmes que l'on peut résoudre efficacement selon que la machine de Turing est déterministe, probabiliste ou non déterministe. Parallèlement à cela, il semble que, d'un point de vue pratique, nous puissions créer des machines physiques déterministes et probabilistes, comme le sont nos ordinateurs. En revanche, ces derniers sont des machines essentiellement classiques, ne tirant pas profit du principe de superposition pour faire les calculs. De même les modèles déterministes et probabilistes sont des modèles classiques, alors que le modèle non déterministe semble éloigné à la fois des lois de la physique classique et de celles de la physique quantique.

L'objectif de ce chapitre va donc être de marier théorie quantique et calcul, à l'instar de ce que nous avons fait dans le chapitre précédent entre théorie quantique et théorie de l'information. Nous allons tout d'abord présenter divers algorithmes qui utilisent le formalisme des circuits quantiques introduit dans le chapitre précédent afin d'illustrer l'apport de la théorie quantique au calcul. Nous verrons ainsi que certains problèmes bénéficient de l'utilisation de ressources quantiques d'un point de vue de la théorie de la complexité. D'un point de vue pratique, un ordinateur quantique voit ainsi certains problèmes comme faciles, alors qu'ils sont pensés comme étant difficiles pour un ordinateur classique.

Nous développerons ensuite le modèle de la machine de Turing quantique, ainsi que la classe des problèmes faciles à résoudre pour cette machine. Cela nous permettra de formaliser la classe des problèmes faciles à résoudre pour un ordinateur quantique et de voir comment celle-ci se place parmi les autres classes de complexité que nous avons introduites.

Le modèle de la machine de Turing quantique n'est cependant pas des plus pratiques pour étudier les algorithmes quantiques. Nous reviendrons donc plus en détail sur le modèle des circuits quantiques. Cela nous permettra de clarifier les notions de complexité en circuit, ainsi que de voir comment il est

possible de réaliser n'importe quel opérateur unitaire à partir d'un ensemble universel de portes quantiques. Enfin, nous verrons un bref aperçu des codes correcteurs d'erreurs quantiques, dont l'importance est primordiale pour le calcul quantique.

Pour conclure cette brève introduction, et avant de nous lancer plus en détail dans l'étude d'algorithmes et systèmes de calculs utilisant des ressources quantiques, il est intéressant de préciser un peu ce que nous recherchons. Pour cela, nous allons utiliser des notions de complexité semblables à celles que nous avons développées dans le chapitre 3. Ainsi, on distinguera deux cas dans lesquels le monde quantique permet « d'aller plus vite » que le monde classique :

- Il existe parfois une borne inférieure sur les ressources nécessaires pour répondre à un problème donné avec un algorithme classique. Dans ce cas, on peut parfois montrer l'existence d'un algorithme quantique qui nécessite moins de ressources que cette borne classique et donc que n'importe quel algorithme classique. On dit alors que l'accélération quantique est *prouvée*.
- Cependant, sur un certain nombre de problèmes, les bornes inférieures connues sont loin d'être atteintes. Dans ce cas, on peut souhaiter non pas comparer à la borne inférieure, mais au meilleur algorithme classique connu. Si on trouve un algorithme quantique meilleur que le meilleur algorithme classique connu, on parlera uniquement d'*accélération quantique*. Notons que cette notion est beaucoup plus fragile que la précédente : il est toujours possible qu'il existe un algorithme classique qui soit meilleur que ceux que l'on connaît actuellement.

À l'heure actuelle, assez peu d'algorithmes quantiques ont été trouvés dans l'une ou l'autre de ces catégories. Nous allons faire un tour d'horizon de la plupart d'entre eux dans les sections 8.1 et 8.2.

Notons que lorsque l'on parle d'accélération quantique, on ne précise pas le type de complexité dont on parle. Dans le chapitre 3, nous avons introduit la notion de complexité en temps, qui consiste à regarder le temps de calcul asymptotique dans le pire des cas. Dans ce chapitre, nous allons introduire deux notions de complexités quelque peu différentes : la *complexité en requêtes* et la *complexité en circuit quantique*. Dans le cas de la complexité en requêtes, nous avons accès à une boîte noire, appelée *oracle*, qui à chaque appel calcule le résultat d'une fonction. Notons que l'on ne place aucune contrainte de calculabilité sur l'oracle. La complexité en requêtes est alors donnée par le nombre d'appels à l'oracle, un algorithme optimal étant un algorithme qui fait un nombre d'appel minimal à l'oracle. La complexité en circuit quantique est, quant à elle, directement reliée au formalisme des circuits quantiques que nous avons développé au chapitre précédent. Ce qui importe est alors le nombre de portes quantiques impliquées pour faire un calcul, une porte quantique étant

n'importe quelle opération agissant sur trois qubits ou moins¹. Un algorithme optimal est alors celui qui a le circuit possédant le moins de portes. Cette définition possède quelques subtilités sur lesquelles nous reviendrons dans la section 8.4. Néanmoins, à ces subtilités près, nous verrons que la complexité en circuit quantique est en réalité équivalente, à un facteur polynomial près, à la complexité en temps pour une machine de Turing quantique. Nous assimilerons donc les deux notions dans cette section.

Nous allons donc étudier les algorithmes selon les deux approches. Nous verrons qu'il est possible de démontrer l'accélération quantique en se basant sur la complexité en requêtes. En revanche, pour la complexité en temps (ou en circuit quantique), il devient très difficile de démontrer des résultats. Les algorithmes que nous allons présenter relèvent donc de la seconde catégorie dans les cas où ceux-ci sont meilleurs que leurs analogues classiques.

Il peut sembler étrange en premier lieu que les résultats donnés par l'approche boîte noire soient si différents de ceux que l'on obtient avec la complexité en temps. Après tout, il est possible de spécifier la boîte noire en prenant par exemple une fonction qui se calcule en temps polynomial. L'hypothèse clé est que l'on n'a pas accès au contenu de la boîte noire ; on ne connaît pas la structure de la fonction calculée par celle-ci. Ainsi, certaines structures sous-jacentes aux fonctions qui se calculent en temps polynomial peuvent être utilisées pour développer un algorithme plus efficace. Notons également que dans bien des cas obtenir une séparation pour des complexités en temps entre le cas classique et le cas quantique aboutirait à montrer que $P \neq NP$, ce qui laisse deviner toute la difficulté de telles preuves.

8.1 Oracles et accélération quantique prouvée

Pour commencer notre tour d'horizon de quelques algorithmes quantiques, nous allons présenter deux algorithmes pour lesquels il est démontré que la complexité en requêtes obtenue dans le cas quantique est meilleure que la meilleure complexité possible dans le cas classique.

8.1.1 Trouver si une fonction est constante

Le problème posé

Le premier problème auquel on s'intéresse vise à déterminer si une fonction est constante ou non. De manière générale, si on considère une fonction définie sur un ensemble fini comportant $N = 2^n$ éléments, déterminer si cette fonction est constante nécessite dans le pire des cas de l'évaluer sur N valeurs.

1. Le nombre de qubits est ici arbitraire : n'importe quelle constante supérieure ou égale à deux convient. Cependant, nous avons choisi de considérer les opérateurs à trois qubits car certaines familles de portes s'expriment bien avec des portes à trois qubits.

Si on impose certaines propriétés à notre fonction, le nombre d'évaluations nécessaires (la complexité en requêtes) peut être inférieure. Dans la suite, nous allons considérer des fonctions à valeurs dans l'ensemble $\{0, 1\}$ qui sont, soit constantes, soit équilibrées. On a donc trois possibilités : soit la fonction est constante égale à 0, soit elle est constante égale à 1, soit elle prend autant de fois la valeur 0 et la valeur 1 sur l'ensemble de ses arguments. Nous allons nous intéresser au *problème de Deutsch-Jozsa* :

Étant donné une fonction f que l'on sait être soit constante, soit équilibrée, f est-elle constante ?

C'est ce qu'on appelle un *problème à promesse*, la promesse ici étant que la fonction est soit constante, soit équilibrée. Dans le cas où la promesse n'est pas satisfaite, on ne demande rien.

Comment comprendre ce problème ? Imaginons que l'on nous donne une boîte noire (l'oracle) qui, si on lui fournit une entrée x , calcule la valeur $f(x)$. De plus, la personne qui nous donne la boîte noire nous fait la promesse que f est soit constante, soit équilibrée. Cette boîte noire est notre oracle. Pour résoudre le problème de Deutsch-Jozsa de manière déterministe, on doit fournir $N/2 + 1$ entrées différentes à l'oracle qui renvoie autant de valeurs de la fonction f . Si toutes les valeurs sont égales, alors on peut conclure que f est constante, sinon elle est équilibrée. Pour cela, on doit interroger l'oracle $N/2 + 1$ fois, et on a donc une complexité en requêtes exponentielle en n .

On va chercher à faire la même chose avec une machine quantique : on va nourrir un oracle quantique qui code le calcul de $f(x)$ sur des objets quantiques et envoyer la sortie de cet oracle quantique dans un dispositif effectuant des mesures. La force est qu'ici l'oracle quantique ne sera utilisé qu'une seule fois.

Pour obtenir une intuition sur la façon de réaliser ce petit miracle, le plus simple est de commencer par le cas $N = 2$.

Le cas $N = 2$

Dans ce cas, il existe quatre fonctions de $\{0, 1\}$ dans $\{0, 1\}$: deux d'entre elles sont constantes et deux d'entre elles sont équilibrées. Pour déterminer si une fonction f est constante, il faut l'évaluer deux fois classiquement. Mais peut-on faire cela quantiquement de manière plus simple ?

De manière intéressante, l'optique nous fournit une situation assez comparable. Imaginons que l'on dispose d'un interféromètre de Mach-Zehnder réglé de telle sorte que toute l'intensité soit concentrée sur la voie de sortie numéro 1 (voir figure 8.1). Dans ce cas, si on introduit un déphasage sur chacune des deux voies de l'interféromètre, la figure d'interférence restera identique si et seulement si la différence des phases introduites sur les deux voies est égale à 2π . Dans le cas contraire, on observera que l'intensité se répartira différemment entre les deux voies de sortie. Dans le cas où la différence de phase est exactement de π , alors toute l'intensité lumineuse sortira par la voie 2 au lieu de la voie 1.

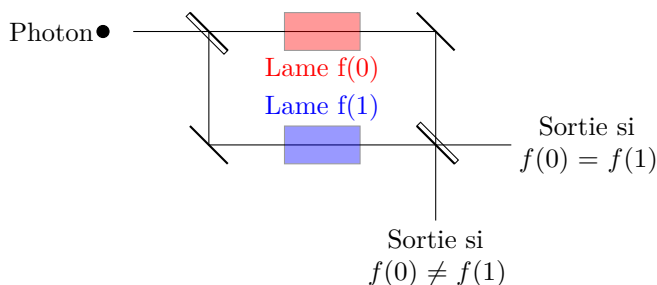


FIGURE 8.1 – Interféromètre de Mach-Zehnder permettant de tester si une fonction est constante ou équilibrée dans le cas $N = 2$. Les lames placées le long des deux chemins optiques ajoutent une phase π si et seulement si $f(x) = 1$. La branche de sortie est donc entièrement déterminée par les valeurs prises par la fonction.

En optique classique, il est donc très facile de mettre au point une machine qui détermine si une fonction sur $\{0, 1\}$ à valeurs dans $\{0, 1\}$ est constante ou équilibrée. Étant donné cette fonction, l'oracle est on ne peut plus concret : ce sont deux lames optiques numérotées 0 et 1 et qui introduisent une phase $(-1)^{f(x)}$ où $x \in \{0, 1\}$ est le numéro de la lame considérée. Pour déterminer si les deux lames sont identiques (fonction constante) ou différentes (fonction équilibrée), il suffit de les introduire dans un interféromètre de Mach-Zehnder réglé comme indiqué au paragraphe précédent, ce qui est représenté sur la figure 8.1. Si les deux lames sont identiques, toute l'intensité sort toujours du côté 1, alors que si les lames sont différentes, toute l'intensité sort du côté 2.

On pourra objecter que c'est une manière bien compliquée de montrer qu'une fonction ayant deux valeurs est constante ou non mais c'est pourtant la clé pour aller plus loin. En effet, dans ce montage, l'intensité lumineuse sort, soit intégralement dans la voie 1, soit intégralement dans la voie 2. Cela veut dire que, si on effectue maintenant l'expérience avec un photon unique, on est capable de décider à coup sûr si les lames sont identiques ou non. En clair, même si nous n'avons le droit d'utiliser les lames qu'avec un seul photon, nous pouvons déterminer si les lames sont identiques ou non. Là où l'oracle classique devait être utilisé deux fois, notre oracle optique n'a besoin d'être utilisé qu'une fois grâce aux interférences quantiques.

Le montage optique que nous avons obtenu dans le cas $N = 2$ nous donne de plus une piste pour généraliser sur des tailles d'entrées plus grandes. En effet, le problème de la détermination du caractère constant d'une fonction sur un ensemble de $N = 2^n$ points peut se reformuler en termes optiques : il s'agit de déterminer si N lames transparentes introduisent le même décalage de phase ou non.

Le cas $N = 2^n$ quantique

Pour pouvoir passer à un plus grand nombre de lames semi-réfléchissantes, on va utiliser plusieurs interféromètres de Mach-Zehnder mis bout à bout. Commençons par le cas $N = 4$ pour comprendre la construction. Tout d'abord, on utilise une première lame semi-réfléchissante qui envoie vers l'entrée de deux interféromètres du cas $N = 2$. On place alors des détecteurs à photons sur les voies de sortie correspondant à des valeurs de f différentes pour chaque interféromètre. Si l'un de ces deux détecteurs « clique », cela signifie que les deux lames correspondantes sont différentes. Sinon, cela signifie que dans chaque interféromètre, les deux valeurs prises par la fonction étaient égales. Dans ce cas, on se sert d'une lame semi-réfléchissante supplémentaire pour comparer à nouveau les deux sorties. Cette fois encore, cela permettra de comparer si les deux blocs correspondaient aux mêmes phases ou non.

Pour passer au cas général, il suffit de mettre ces blocs bout à bout de façon à comparer les lames deux à deux, puis les groupes de deux lames identiques deux à deux, etc. L'interféromètre est construit récursivement : l'interféromètre pour $n + 1$ est l'interféromètre pour n copié deux fois à l'intérieur de l'interféromètre pour 1. Si l'un des détecteurs correspondant à deux lames différentes détecte le photon, alors on sait que la fonction est équilibrée par hypothèse. Un schéma de cette méthode est présenté sur la figure 8.2 pour le cas $N = 8$.

Regardons la probabilité que le détecteur dessiné en vert sur la figure détecte le photon. Pour cela, on va chercher à écrire la probabilité de transition entre la source et ce détecteur. Les règles à retenir sont les suivantes. Chaque réflexion sur un miroir multiplie l'amplitude correspondante par i . Sur une lame semi-réfléchissante, chaque réflexion multiplie l'amplitude par $i/\sqrt{2}$ et chaque transmission la multiplie par $1/\sqrt{2}$. Enfin, chaque traversée de lame dépendant de la fonction f ajoute une phase $f(x)\pi$. Or, on peut vérifier facilement sur le schéma que la construction choisie correspond à exactement 3 réflexions sur des miroirs, 3 sur des lames semi-réfléchissantes et 3 transmissions à travers des lames. Ainsi, l'amplitude totale pour aller de la source à ce détecteur est, dans le cas $N = 8$,

$$\mathcal{A}_{SD} = \sum_{i=0}^7 \mathcal{A}_i = \left(\frac{1}{\sqrt{2}}\right)^3 \left(\frac{i}{\sqrt{2}}\right)^3 i^3 \left(\sum_{i=0}^7 (-1)^{f(i)}\right) = -\frac{1}{8} \left(\sum_{i=0}^7 (-1)^{f(i)}\right). \quad (8.1)$$

Si la fonction est équilibrée, cette somme est nulle, et la probabilité que ce détecteur « clique » est donc nulle également. À l'inverse, si la fonction est constante, on voit que $\mathcal{A}_{SD} = \pm 1$, et la probabilité associée est donc égale à 1. On a bien construit un interféromètre capable de détecter, en un unique photon, si la fonction est constante ou équilibrée. L'extension de cet interféromètre à des valeurs de n plus grandes est directe, puisqu'il suffit d'ajouter des « étages » à celui qui est présenté sur la figure 8.2.

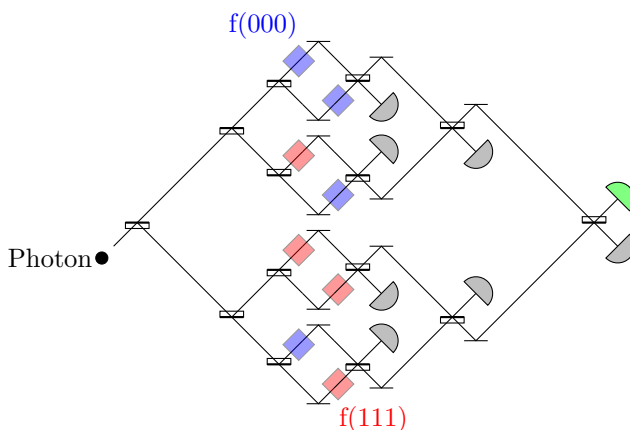


FIGURE 8.2 – Groupe d’interféromètres de Mach-Zehnder permettant de tester si une fonction est constante ou équilibrée dans le cas $N = 8$. Pour chaque interféromètre, le photon est transmis dans la même direction si les deux lames introduisent la même phase, et réfléchi vers un détecteur (en gris) dans le cas contraire. Les lames placées le long des différents chemins optiques x ajoutent une phase π si et seulement si $f(x) = 1$. Les détecteurs dessinés en gris ne peuvent ainsi détecter le photon que si les lames correspondantes sont différentes, ce qui permet de répondre que la fonction est équilibrée. À l’inverse, le dernier détecteur (en vert) ne peut détecter le photon que si la fonction est constante. De plus, il faut bien noter que l’on ne peut pas obtenir plus d’informations que cela avec ce système : dans le cas d’une fonction équilibrée, le détecteur qui va finalement détecter le photon est inconnu *a priori*. Par exemple, dans le cas représenté ici, 4 détecteurs peuvent détecter le photon, chacun avec une probabilité $1/4$.

Circuit quantique de l’algorithme

Pour terminer ce tour d’horizon sur l’algorithme de Deutsch-Jozsa, donnons-en une description en termes de circuits quantiques. Cela nous permettra également de mettre en avant plusieurs éléments qui nous serviront dans tout le reste de ce chapitre pour étudier des algorithmes quantiques.

Le premier élément dont nous avons besoin pour pouvoir décrire notre algorithme est une manière de décrire l’oracle. Le but de l’algorithme dans ce formalisme est un peu différent de celui vu précédemment : le problème ne va pas être d’utiliser un seul qubit, mais de faire appel à cet oracle une unique fois. Dans l’exemple optique précédent, les lames à déphasage que l’on introduisait sur les différents bras jouaient ce rôle d’oracle. Comment décrire cet oracle par une porte de calcul que nous pourrions utiliser dans un circuit quantique ? Puisque la mécanique quantique est unitaire, il est nécessaire que

l'action de l'oracle le soit aussi. On va donc décrire l'oracle lié à la fonction f comme un opérateur U_f agissant sur des états quantiques à n qubits (puisque la fonction f est définie sur $\{0, 1\}^n$ et à valeurs dans $\{0, 1\}$). Pour conserver l'unitarité de cet opérateur, il est nécessaire qu'il agisse également sur un qubit supplémentaire sur lequel le résultat du calcul sera inscrit. Ainsi, pour $x \in \{0, 1\}^n$ et $y \in \{0, 1\}$, on définit l'action de U_f sur les états quantiques $|x\rangle$ et $|y\rangle$ par :

$$|x\rangle \otimes |y\rangle \xrightarrow{U_f} |x\rangle \otimes |y \oplus f(x)\rangle, \quad (8.2)$$

où $\oplus$ dénote l'addition bit-à-bit modulo 2.

Il est possible d'utiliser cet oracle pour encoder l'information sur f dans la phase de l'état. C'est cela qui nous permettra de réaliser des interférences entre les différentes valeurs $f(x)$. Cela s'obtient de la façon suivante :

$$U_f |x\rangle \otimes \left(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle \right) = (-1)^{f(x)} |x\rangle \otimes \left(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle \right). \quad (8.3)$$

Cela se vérifie en regardant pour chacune des deux valeurs possibles de $f(x)$. Ainsi, on se rend compte que, si le dernier qubit est dans l'état $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, alors la valeur de $f(x)$ est entièrement contenue dans la phase de l'état global, tandis que l'état du dernier qubit reste inchangé. Ainsi nous pourrions remplacer l'application de l'opérateur U_f par celle de l'opérateur S_f qui effectue la transformation

$$|x\rangle \xrightarrow{S_f} (-1)^{f(x)} |x\rangle. \quad (8.4)$$

Revenons désormais à l'algorithme de Deutsch-Jozsa, que l'on peut construire à partir de notre opérateur oracle U_f (ou bien de son équivalent S_f) et de l'opérateur de Hadamard. On commence par préparer l'état $|\psi_0\rangle = |00\dots 0\rangle \otimes |1\rangle$ sur nos $n + 1$ qubits, puis on applique l'opérateur de Hadamard à tous ces qubits. L'action des opérateurs de Hadamard sur un groupe de n qubits est la suivante :

$$H_{2^n} |x\rangle = H_2 |x_1\rangle \otimes \dots \otimes H_2 |x_n\rangle = \frac{1}{\sqrt{2^n}} \sum_{z \in \{0,1\}^n} (-1)^{x \cdot z} |z\rangle, \quad (8.5)$$

où l'on a dénoté dans la dernière équation par $|x\rangle$ l'état $|x_1\rangle \otimes \dots \otimes |x_n\rangle$ et par $x \cdot z$ la somme $\sum_i x_i z_i$ modulo 2. Lorsqu'on applique l'opérateur de Hadamard à tous nos qubits dans l'état initial, on crée l'état global

$$H_{2^{n+1}} |\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \otimes \left(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle \right). \quad (8.6)$$

Le signe entre les deux termes provient du fait que le dernier qubit de l'état $|\psi_0\rangle$ est dans l'état $|1\rangle$. On peut voir ici une certaine analogie entre l'action de cet

opérateur et la première partie de l'interféromètre, dans lequel on divisait le photon sur l'ensemble des chemins optiques. L'étape suivante est d'appliquer l'action de notre oracle à cet état :

$$U_f H_{2^{n+1}} |\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle \otimes \left(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle \right), \quad (8.7)$$

puisque le dernier qubit est dans l'état $|-\rangle$. Les divers résultats de l'oracle se retrouvent donc encodés dans la phase des états considérés. On retrouve ici l'idée que l'on avait dans le cas optique avec nos lames à déphasage variable. On applique alors à nouveau l'opérateur de Hadamard à tous nos qubits, à la manière de la recombinaison des chemins optiques,

$$H_{2^{n+1}} U_f H_{2^{n+1}} |\psi_0\rangle = \sum_{y \in \{0,1\}^n} \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x) + x \cdot y} |y\rangle \otimes |1\rangle. \quad (8.8)$$

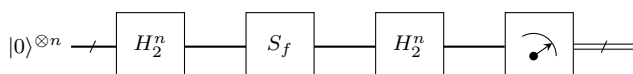
Enfin, on cherche à mesurer l'état des n premiers qubits. La probabilité de trouver l'état $y = |0\rangle^{\otimes n}$, ce qui correspond à la probabilité que le bon détecteur soit activé dans le cas optique, est donnée par

$$\left| \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} \right|^2 = \begin{cases} 0 & \text{si } f \text{ est équilibrée,} \\ 1 & \text{si } f \text{ est constante,} \end{cases} \quad (8.9)$$

ce qui conclut notre algorithme. En faisant appel une seule fois à l'oracle, on a pu répondre à la question qui nous intéressait sur la fonction considérée grâce aux phénomènes d'interférences de la mécanique quantique. Une représentation en circuit quantique de cet algorithme est donnée sur la figure 8.3².

L'algorithme de Deutsch-Jozsa permet donc de passer de 2^n requêtes faites à l'oracle par un algorithme déterministe à une seule requête. On a donc atteint une séparation exponentielle entre une machine déterministe et une machine quantique. Néanmoins, le statut de cette séparation est extrêmement particulier, et insistons sur le fait qu'il est difficile d'en déduire qu'une machine quantique est plus puissante qu'une machine déterministe dans des cas pratiques et notamment en ce qui concerne la complexité en temps. Cela vient à la fois du fait que l'on regarde une complexité en requête ainsi que du fait que l'on regarde des problèmes à promesses. En réalité, pour ce problème, on peut également montrer qu'une machine probabiliste fait également beaucoup mieux qu'une machine déterministe.

2. On pourrait remplacer de manière formelle ce circuit par le suivant, où S_f agit selon l'équation (8.4).



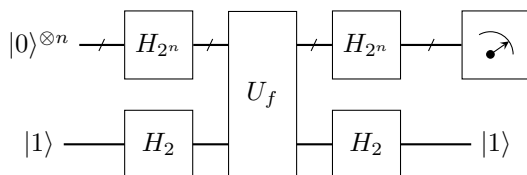


FIGURE 8.3 – Circuit de l’algorithme de Deutsch-Jozsa. On prépare un état initial $|0\rangle^{\otimes n} \otimes |1\rangle$, puis on applique un opérateur de Hadamard à chacun des $n + 1$ qubits, avant d’utiliser un oracle écrivant le résultat de la fonction dans le dernier qubit. Il suffit de mesurer l’état des n premiers qubits après l’application d’un nouvel opérateur de Hadamard pour savoir si la fonction est constante ou équilibrée : seul le résultat correspondant à l’état $|0\rangle^{\otimes n}$ est associé à une fonction constante. Le qubit supplémentaire étant revenu à son état initial à travers cette procédure, il n’y a pas besoin de le mesurer.

Pour cela, considérons l’algorithme suivant : pour k fixé, la fonction est évaluée sur $k + 1$ éléments tirés au sort indépendamment. Si on trouve deux valeurs différentes, on déclare la fonction équilibrée, sinon on décide qu’elle est constante. L’algorithme peut se tromper si la fonction est équilibrée et que les k derniers éléments tirés au sort ont la même image que le premier, ce qui peut se produire avec probabilité au plus $1/2^k$. On a donc une complexité en requête en $\mathcal{O}(1)$. Pour les mêmes raisons que précédemment, cela ne montre pas que les machines probabilistes sont plus puissantes que les machines déterministes pour la complexité en temps. Cela illustre la subtilité des oracles ainsi que celle des problèmes à promesses. Si cela donne des indications sur les possibilités d’une machine par rapport à une autre dans certains cas très précis, il est difficile d’extrapoler cela aux complexités qui nous intéressent véritablement, à savoir les complexités temporelles.

8.1.2 Chercher dans une base de données non triée

Le second problème que nous allons étudier consiste à retrouver, dans une base de données non triée, un élément vérifiant une propriété donnée. Par exemple, étant donné un annuaire classé par ordre alphabétique, comment trouver le nom du détenteur d’un numéro que l’on connaît ? Pour cela, il faut lire un couple nom-numéro dans l’annuaire, et vérifier si le numéro correspond. Ce type d’opération est appelée une requête, et notre but est, tout comme dans l’algorithme de Deutsch-Jozsa, de minimiser le nombre de requêtes que l’on a besoin de faire.

On peut formaliser ce problème de la manière suivante. Étant donné une liste d’éléments indexés par $x \in \{0, 1\}^n$ (les noms se trouvant dans l’annuaire) et une fonction f telle que $f(x) = 1$ si l’élément indexé par x vérifie la propriété voulue (*i.e.* le numéro de téléphone de la personne correspond à celui que

l'on cherche), une solution à notre problème est un entier a tel que $f(a) = 1$ (c'est-à-dire le nom a d'une personne ayant le numéro que l'on cherche). La fonction f joue ici un rôle d'oracle : on ne sait rien de ses propriétés, et le seul accès possible pour obtenir de l'information est d'effectuer une requête. La complexité en requêtes de l'algorithme est le nombre maximal de requêtes dans le pire des cas, c'est-à-dire pour toute fonction f et tout calcul possible de l'algorithme sur cette entrée f . C'est comme toujours une fonction de la taille de la base de données $N = 2^n$. En revenant à l'exemple de l'annuaire, la complexité en requêtes est le nombre maximal de couples nom-numéro que nous devons consulter pour trouver le numéro recherché, et ce nombre est une fonction du nombre total de noms N dans l'annuaire. De façon remarquable, nous disposons d'un algorithme optimal, à la fois dans le cas classique et dans le cas quantique [64, 14].

Penchons-nous tout d'abord sur les résultats classiques pour ce problème, en supposant qu'un seul élément vérifie la propriété recherchée : intuitivement, c'est le cas où trouver cet élément est le plus difficile. Si l'algorithme est classique et déterministe, dans le pire des cas nous devons tester toutes les possibilités : la complexité est N . On peut également autoriser l'utilisation du hasard, en se donnant des algorithmes probabilistes. Par définition, un algorithme résout notre problème avec probabilité p si en un temps fini, pour tout oracle calculant f , et avec probabilité au moins p , il nous donne un entier a tel que $f(a) = 1$. Malheureusement, cela ne permet pas de faire baisser significativement la complexité. Considérons l'algorithme naïf qui consiste à tirer indépendamment et uniformément au hasard k éléments dans $\{0, \dots, N\}$ et à effectuer les requêtes correspondantes pour voir si une solution a été trouvée. La probabilité de succès de cet algorithme est $1 - (1 - \frac{1}{N})^k$, qui est équivalente à k/N pour N grand. Pour avoir une probabilité p de résoudre le problème, il faut donc un nombre de requêtes pN , ce qui donne donc une complexité en $\mathcal{O}(N)$. Nous pouvons montrer que c'est le cas pour tous les algorithmes probabilistes pour ce problème. En revanche, les algorithmes quantiques font mieux, comme nous allons le voir.

Notons que contrairement au cas précédent, ce problème possède des applications pratiques potentielles. En effet, il s'agit de trouver un inverse d'une fonction à valeur dans $\{0, 1\}$. Si l'on pense aux problèmes NP-complets, cela revient donc à trouver le « certificat » qui donne une solution au problème, si on utilise le vérifieur comme fonction. Étant donné la difficulté actuelle pour résoudre les problèmes NP-complets, posséder une méthode générique pour baisser le coût du problème de l'inversion d'une fonction est donc extrêmement intéressant.

L'algorithme de Grover

L'algorithme quantique que nous allons présenter ici a été découvert par Lov Grover en 1996 [187]. Celui-ci permet de réduire le nombre de requêtes

de façon quadratique. L'élément central de cet algorithme est l'opération d'inversion par rapport à la moyenne.

L'inversion par rapport à la moyenne. Imaginons que nous ayons quatre nombres 1, 2, 6 et 3. Quel est leur symétrique par rapport à leur moyenne $m = 3$? Le symétrique x' de x par rapport à m vérifie $(x + x')/2 = m$, c'est-à-dire $x' = 2m - x$. Les symétriques sont donc 5, 4, 0 et 3. Notre but est de trouver la version quantique de cette opération. Considérons l'état $|\varphi\rangle = \frac{1}{\sqrt{50}}(|00\rangle + 2|01\rangle + 6|10\rangle + 3|11\rangle)$. Si nous prenons les symétriques des amplitudes par rapport à leur moyenne, nous devons arriver à l'état $|\varphi'\rangle = \frac{1}{\sqrt{50}}(5|00\rangle + 4|01\rangle + 3|11\rangle)$. L'opération G_N qui à $|\varphi\rangle$ associe $|\varphi'\rangle$ est appelée *inversion par rapport à la moyenne*.

Comment la réaliser par une opération quantique? Le calcul de la moyenne des amplitudes de $|\varphi\rangle$ peut se faire par la projection P sur l'espace engendré par $|\psi_0\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} |x\rangle$. La matrice de $P = |\psi_0\rangle\langle\psi_0|$ est donc la matrice dont tous les éléments sont $1/N$ et, si m est la moyenne des amplitudes de $|\varphi\rangle$, alors $P|\varphi\rangle = \sum_{x \in \{0,1\}^n} m|x\rangle$. Par construction, $2P - \mathbb{1}$ réalise l'inversion par rapport à la moyenne, que l'on note G_N , et nous avons

$$G_N = \begin{pmatrix} \frac{2}{N} - 1 & \frac{2}{N} & \cdots & \frac{2}{N} \\ \frac{2}{N} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{2}{N} \\ \frac{2}{N} & \cdots & \frac{2}{N} & \frac{2}{N} - 1 \end{pmatrix}. \quad (8.10)$$

Nous pouvons vérifier facilement en faisant le produit scalaire de deux lignes de la matrice que celle-ci est unitaire. Mais ce n'est pas le cas de P , et il faut donc trouver une autre façon de réaliser cette opération en pratique. Pour cela, nous allons utiliser l'opérateur de Hadamard, que nous avons déjà utilisé de nombreuses fois, ainsi que l'inversion de phase en 0 : c'est l'opération R_N qui à $|0 \dots 0\rangle$ associe $-|0 \dots 0\rangle$ et laisse invariants les autres éléments de la base canonique. On peut alors construire G_N comme le produit $-H_N R_N H_N$, où $H_N = H_{2^n}$ dénote toujours l'opérateur $H_2^{\otimes n}$. L'opération R_N peut être obtenue avec un nombre de portes polynomial en n (on peut se servir par exemple de la matrice de Pauli σ_z et la porte de Toffoli, voir [287, section 4.3]).

Fonctionnement de l'algorithme On va désormais pouvoir détailler le fonctionnement de l'algorithme de Grover dans le cas où il n'y a qu'un seul élément solution du problème. On utilise à la fois l'opérateur d'inversion par rapport à la moyenne et un oracle quantique de la même forme que celui utilisé dans le cas de Deutsch-Jozsa $S_f : |x\rangle \mapsto (-1)^{f(x)}|x\rangle$, en se rappelant qu'une mise en œuvre physique nécessite d'ajouter un qubit au protocole.

L'idée derrière l'algorithme est représentée sur la figure 8.4. Tout d'abord, on initialise le système en préparant n qubits dans l'état $|0\rangle$, puis on applique

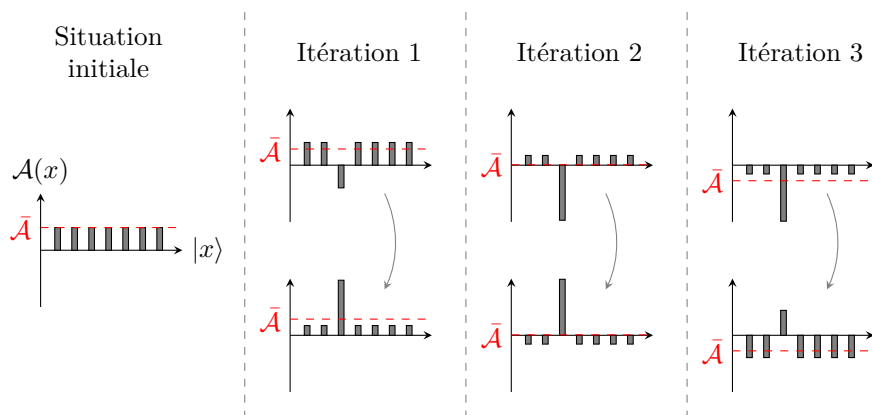


FIGURE 8.4 — Les différentes étapes du protocole de Grover. Dans la situation initiale, on prépare un état superposé tel que la probabilité de le trouver dans n'importe lequel des états $|x\rangle$ de la base de mesure soit la même. Ici, on représente l'amplitude de probabilité et non la probabilité elle-même. Ensuite, à chaque itération du protocole, on retourne la phase de l'amplitude correspondant à la valeur recherchée (en haut sur le schéma), puis on inverse toutes les amplitudes par rapport à la valeur moyenne $\bar{A}$ (en bas). On voit que la probabilité de trouver le bon résultat lors de la mesure de l'état augmente au début, avant de diminuer ensuite. Ici, il aurait fallu s'arrêter à deux itérations du protocole.

un opérateur H_{2^n} à ces qubits pour créer une superposition équilibrée sur tous les états à n qubits possibles, dénotée $|\psi_0\rangle$. Ensuite, on va effectuer un certain nombre de fois la boucle suivante :

1. On applique l'oracle S_f à nos n qubits. Cela a pour effet de « retourner » l'amplitude de l'état $|x\rangle$ tel que $f(x) = 1$, c'est-à-dire celui qui correspond à l'élément que l'on cherche.
2. On applique l'opérateur d'inversion par rapport à la moyenne G_{2^n} sur nos n qubits. Pour 2^n suffisamment grand, la moyenne des amplitudes est très proche de $1/\sqrt{2^n}$, et l'inversion n'a donc que peu d'effet, excepté pour l'état de la superposition qui a une amplitude $-1/\sqrt{2^n}$, qui se retrouve après inversion avec une amplitude de l'ordre de $3/\sqrt{2^n}$.

Ce protocole peut être représenté par le circuit quantique visible sur la figure 8.5. À chaque étape, et tant que la moyenne n'est pas trop modifiée, on va augmenter l'amplitude de l'état solution d'une valeur $2/\sqrt{2^n}$. Enfin, après un certain nombre d'étapes, on mesure l'état obtenu. On a alors une forte probabilité de trouver le résultat correspondant à la bonne solution du problème, puisque l'amplitude correspondante est bien plus élevée que celle des autres états de la superposition. Si ce n'est pas le cas, il suffit de recommencer l'algorithme.

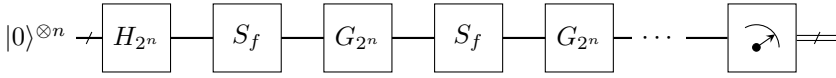


FIGURE 8.5 – Circuit de l’algorithme de Grover. On doit appliquer un certain nombre de fois les opérateurs S_f et G_{2^n} à nos qubits.

Estimer le nombre d’étapes Pour pouvoir arrêter la boucle au bon moment pour nous permettre d’avoir le résultat avec la plus forte probabilité possible, il est nécessaire d’estimer avec soin le nombre d’étapes à faire. En effet, si on effectue trop de boucles, on diminue à nouveau l’amplitude de la solution, comme cela est visible sur la figure 8.4. De plus, l’estimation de ce nombre d’étapes nous permettra de donner la complexité de cet algorithme.

En dénotant par $|s\rangle$ l’état solution du problème, l’état initial peut être réécrit :

$$|\psi_0\rangle = \frac{1}{\sqrt{2^n}} |s\rangle + \frac{1}{\sqrt{2^n}} \sum_F |x\rangle, \quad (8.11)$$

où F correspond à l’ensemble des résultats faux de l’algorithme, c’est-à-dire $\{0, 1\}^n \setminus \{s\}$. On note s_r l’amplitude de $|s\rangle$ après r étapes de l’algorithme. De même, l’amplitude des non-solutions sera notée f_r , les non-solutions possédant la même amplitude. On peut noter :

$$|\psi_0\rangle = s_0 |s\rangle + \sum_F f_0 |x\rangle \text{ avec } s_0 = f_0 = \frac{1}{\sqrt{2^n}}, \quad (8.12a)$$

$$|\psi_r\rangle = s_r |s\rangle + \sum_F f_r |x\rangle \text{ avec } s_r^2 + (2^n - 1)f_r^2 = 1. \quad (8.12b)$$

On peut alors définir un angle θ_r par :

$$\sin \theta_r = s_r \text{ et } \cos \theta_r = \sqrt{2^n - 1} f_r. \quad (8.13)$$

Cela permet de représenter les états quantiques au cours de l’algorithme par des points du cercle de rayon 1, en représentant l’espace engendré par les non-solutions sur l’axe des abscisses et l’espace engendré par la solution sur l’axe des ordonnées. L’état représenté par le point $(\cos \theta, \sin \theta)$ donnera alors la solution avec probabilité $\sin^2 \theta$ si on le mesure. Pour avoir une forte probabilité d’observer la solution, il faut donc se rapprocher autant que possible de l’angle $\pi/2$. Ici, on peut noter en particulier $\sin \theta_0 = 1/\sqrt{2^n}$ et $\cos \theta_0 = \sqrt{2^n - 1}/\sqrt{2^n}$. Mesurer l’état initial donne donc la bonne solution avec probabilité $1/2^n$.

Définissons maintenant les relations de récurrence entre ces quantités. Pour obtenir $|\psi_{r+1}\rangle$, on part de $|\psi_r\rangle$ et on applique l’inversion de phase, puis l’inversion par rapport à la moyenne,

$$|\psi'_r\rangle = S_f |\psi_r\rangle = -s_r |s\rangle + \sum_F f_r |x\rangle. \quad (8.14)$$

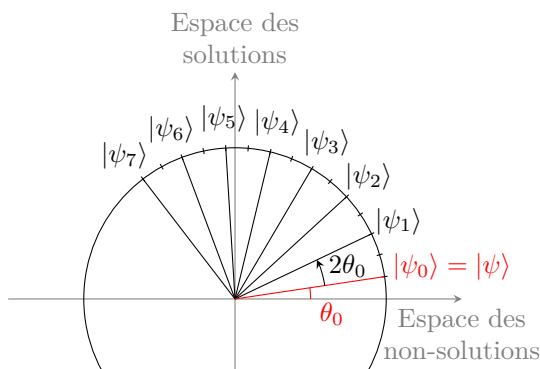


FIGURE 8.6 – Une vision géométrique de l’algorithme de Grover. On peut représenter un état quantique par un point sur le cercle unité de coordonnées $(\sqrt{1-p}, \sqrt{p})$, avec p la probabilité de trouver une solution au problème lors de la mesure de cet état. Chaque étape de l’algorithme de Grover peut alors être vue comme une rotation dans le sens trigonométrique d’un angle $2\theta_0$, où θ_0 correspond à l’angle fait par rapport à l’axe des abscisses par l’état initial. On va donc chercher à arrêter l’algorithme de manière à être le plus près possible de $\pi/2$, c’est-à-dire après environ $(\pi/4\theta_0 - 1/2)$ étapes.

La moyenne des amplitudes de cet état est $m_r = \frac{1}{2^n}(-s_r + (2^n - 1)f_r)$, et on a

$$s_{r+1} = 2m_r + s_r \text{ et } f_{r+1} = 2m_r - f_r, \quad (8.15)$$

soit finalement

$$\begin{pmatrix} \sin \theta_{r+1} \\ \cos \theta_{r+1} \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{2^{n-1}} & \frac{\sqrt{2^n-1}}{2^{n-1}} \\ -\frac{\sqrt{2^n-1}}{2^{n-1}} & 1 - \frac{1}{2^{n-1}} \end{pmatrix} \begin{pmatrix} \sin \theta_r \\ \cos \theta_r \end{pmatrix}. \quad (8.16)$$

On peut reconnaître ici la matrice de rotation d’un angle ω tel que $\cos \omega = 1 - 1/2^{n-1}$ et $\sin \omega = \sqrt{(2^n - 1)}/2^{n-1}$. En remarquant de plus que $\omega = 2\theta_0$, on trouve

$$\theta_{r+1} = \theta_r + 2\theta_0 \text{ et donc } \theta_r = (2r + 1)\theta_0. \quad (8.17)$$

Si l’on réinterprète l’algorithme de Grover géométriquement, on voit que l’inversion par rapport à la moyenne revient à faire une symétrie par rapport à la droite d’angle θ_0 , et que l’application de l’oracle correspond à une symétrie par rapport à l’axe des abscisses. La composée des deux est une rotation d’angle $2\theta_0$. Après r itérations de ces deux opérations, on a donc un état d’angle $(2r + 1)\theta_0$, comme on vient de le montrer par le calcul. C’est ce qui est représenté sur la figure 8.6.

Il reste à choisir r pour que cet angle soit le plus proche possible de $\pi/2$, c’est-à-dire tel que $\frac{\pi}{2} - \theta_0 \leq (2r + 1)\theta_0 \leq \frac{\pi}{2} + \theta_0$. On peut par exemple choisir

$r = \lfloor \pi/4\theta_0 \rfloor$, où $\lfloor x \rfloor$ dénote la partie entière de x . L'angle θ_0 étant tel que $\sin \theta_0 = 1/\sqrt{2^n}$, on peut l'approximer pour n assez grand par la valeur de son sinus. Le nombre d'étapes à effectuer dans cette approximation est donc

$$r \simeq \frac{\pi\sqrt{N}}{4} = \mathcal{O}(\sqrt{N}). \quad (8.18)$$

De plus, la probabilité d'erreur est $1 - \sin^2((2r+1)\theta_0)$, et $\sin((2r+1)\theta_0) \geq \sin(\pi/2 - \theta_0)$ au vu des bornes pour r . La probabilité d'erreur est donc bornée supérieurement par $1 - \cos^2 \theta_0 = 1 - (2^n - 1)/(2^n) = \mathcal{O}(1/2^n)$.

On a donc montré que pour ce problème, l'utilisation de ressources quantiques permettait une amélioration quadratique de la complexité en requêtes. On peut se demander ce que devient la complexité en temps si on prend en compte le temps de calcul de la fonction réalisée par l'oracle. Pour effectuer cette étude, nous prendrons ici une fonction qui se calcule de manière polynomiale en temps. Tester l'ensemble des 2^n solutions donne une complexité en $\mathcal{O}(\text{poly}(n)2^n)$. L'algorithme de Grover permet de trouver la solution en un temps $\mathcal{O}(\text{poly}(n)2^{n/2})$. Même si on a toujours affaire à une complexité exponentielle, l'amélioration donnée par l'algorithme de Grover peut servir à traiter des instances plus grandes d'un problème donné. Évidemment, cela n'a d'intérêt que s'il est plus rapide d'utiliser l'algorithme de Grover que de calculer l'inverse de la fonction.

Même s'il n'est pour l'heure pas démontré que de telles fonctions existent ³, il est toujours possible de se restreindre au cas où aucun algorithme plus rapide que la recherche exhaustive n'est connu. Notamment, pour un cryptosystème à clés symétriques, on considère que l'algorithme est cassé à partir du moment où, étant donné un message en clair et sa contrepartie chiffrée, il est possible de retrouver la clé de manière classique plus rapidement que par une recherche exhaustive ⁴. L'utilisation de l'algorithme de Grover permet donc potentiellement de s'attaquer à des clés deux fois plus longues. Une façon de se protéger de l'algorithme de Grover est donc de doubler la taille des clés [48].

On peut se demander également s'il y a un intérêt à utiliser l'algorithme de Grover pour trouver une solution à des problèmes NP en général. En fait, pour un certain nombre de problèmes NP-complets, on connaît une solution en $\mathcal{O}(\text{poly}(n)2^{an})$ où $a < 1/2$. C'est le cas notamment du problème 3SAT, où il existe un algorithme déterministe capable de trouver la solution en $\mathcal{O}(\text{poly}(n)2^{0,42n})$ [278]. En revanche, il existe certains problèmes pour lesquels on ne connaît pas d'algorithme fondamentalement meilleur que la recherche exhaustive. C'est le cas du problème de la satisfaisabilité d'un circuit. On peut donc imaginer que l'algorithme de Grover puisse bénéficier à ces cas-là. Par

3. Cela impliquerait que $P \neq NP$.

4. En pratique, même un algorithme cassé selon ce critère peut rester solide et utilisé. C'est le cas notamment de l'algorithme AES toujours massivement utilisé, pour lequel il existe une attaque avec de l'ordre de $2^{254,3}$ opérations pour la version 256 bits [60, 356]. Cette amélioration n'est cependant que de deux bits, et ne change donc pas grand chose à la solidité pratique d'AES.

ailleurs, une généralisation de l'algorithme de Grover appelée amplification d'amplitude [65] permet de bénéficier du gain quadratique sur un plus grand nombre de techniques que la simple recherche exhaustive. Il est ainsi possible de transcrire certaines des heuristiques classiques utilisées pour résoudre les problèmes NP-complets, et ainsi de bénéficier d'un gain quadratique par-dessus les bons algorithmes classiques connus. Cette généralisation se fait en remplaçant les opérateurs de Hadamard par d'autres opérateurs quantiques qui utilisent cette heuristique. Bien entendu, cela ne pourra donc être fait que dans les cas où une telle adaptation quantique est possible, ce qui est par exemple le cas du problème 3SAT [275].

✂ Cas de solutions multiples

On peut étendre facilement l'étude que l'on vient de faire au cas où il y a k solutions possibles pour le problème donné et où l'on cherche à trouver l'une d'entre elles. Dans ce cas, on peut réécrire l'état initial de la façon suivante, en appelant T l'ensemble des solutions, qui est de cardinal k , et F l'ensemble des non-solutions, qui est donc de cardinal $2^n - k$:

$$|\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in T} |x\rangle + \frac{1}{\sqrt{2^n}} \sum_{x \in F} |x\rangle. \quad (8.19)$$

Au cours de l'exécution de l'algorithme, toutes les solutions sont traitées de la même manière, ce qui implique qu'à tout moment l'amplitude associée à une solution ne dépend pas de la solution considérée. On note t_r cette amplitude après r étapes de l'algorithme. Comme précédemment, l'amplitude des non-solutions sera notée f_r . On peut noter :

$$|\psi_r\rangle = \sum_{x \in T} t_r |x\rangle + \sum_{x \in F} f_r |x\rangle \text{ avec } kt_r^2 + (2^n - k)f_r^2 = 1. \quad (8.20)$$

On définit donc ici les angles θ_r par :

$$\sin \theta_r = \sqrt{k} t_r \text{ et } \cos \theta_r = \sqrt{2^n - k} f_r. \quad (8.21)$$

Avec la même étude que précédemment, on retrouve exactement le même résultat soit $\theta_r = (2r + 1)\theta_0$. La matrice de rotation permettant de passer d'un angle au suivant est encore une rotation d'angle $2\theta_0$, seul θ_0 étant changé. Le nombre d'étapes à effectuer est donc toujours $\lfloor \pi/4\theta_0 \rfloor$, θ_0 pouvant être obtenu à partir de l'état initial si l'on connaît le nombre k de solutions. Il faudra ainsi $\mathcal{O}(\sqrt{N/k})$ étapes pour trouver une solution avec une probabilité proche de 1.

Comment faire si on ne connaît pas le nombre de solutions ? Malheureusement, il peut arriver que l'on ne connaisse pas le nombre k de solutions *a priori*. Dans ce cas, on va devoir réfléchir un peu plus précisément au nombre d'itérations à effectuer. En effet, si on applique l'algorithme $\frac{\pi}{4}\sqrt{N}$ fois, on

obtiendra une solution avec forte probabilité s'il n'y a qu'une seule solution, mais extrêmement faible s'il y en a plus. Une solution à ce problème a été proposée la même année que l'algorithme de Grover [64], et nous allons présenter ici la version simplifiée proposée dans [202]. L'idée de cet algorithme est d'appliquer l'opérateur de Grover G_N un nombre aléatoire de fois, dans une gamme permettant de trouver un résultat avec une probabilité supérieure à $1/4$. Si on échoue, il suffit de refaire l'algorithme environ 4 fois pour espérer trouver une solution. Bien entendu, on peut commencer par tester sur l'état initial si une réponse au hasard est solution (dans le cas où k est assez grand, une simple recherche au hasard classique a de grandes chances d'aboutir) et n'appliquer l'algorithme que si cette étape n'a pas donné de réponse concluante en quelques essais.

Supposons que la fonction que l'on recherche possède k solutions, et que l'on ne connaisse pas k . On définit, comme précédemment, $\sin^2 \theta_0 = k/2^n$. On appelle m l'entier positif qui va définir notre gamme de tirage aléatoire, c'est-à-dire que l'on va appliquer r fois l'opérateur G_N , où r est tiré aléatoirement avec une distribution uniforme dans $\{0, \dots, m-1\}$. La probabilité de trouver une solution pour r itérations de G_N étant $\sin^2((2r+1)\theta_0)$, la probabilité de trouver une solution pour la gamme donnée par m est donc

$$P_m = \frac{1}{m} \sum_{r=0}^{m-1} \sin^2((2r+1)\theta_0) = \frac{1}{2} - \frac{\sin 4m\theta_0}{4m \sin 2\theta_0}. \quad (8.22)$$

Comment choisir m ? Si $m \geq 1/\sin 2\theta_0$, alors $P_m \geq 1/4$. Peut-on donner une borne à la valeur de $1/\sin 2\theta_0$? On a

$$\frac{1}{\sin 2\theta_0} = \frac{1}{2 \sin \theta_0 \cos \theta_0} = \frac{2^n}{2\sqrt{k}\sqrt{2^n - k}} \leq \sqrt{2^n}, \quad (8.23)$$

la dernière inégalité étant valide tant que k est différent de 0 ou 2^n . Ainsi, si l'on choisit $m = \sqrt{2^n}$, on a nécessairement $m \geq 1/\sin 2\theta_0$, et donc une probabilité moyenne supérieure à $1/4$ de trouver une solution à notre problème. Ainsi, on va trouver une solution en moyenne en quelques itérations du protocole, chaque itération demandant au pire $\sqrt{N}$ requêtes, et donc trouver une solution en $\mathcal{O}(\sqrt{N})$, comme attendu.

Il est possible d'optimiser encore ce protocole en suivant la méthode exacte proposée dans [65], qui augmente m au fur et à mesure si l'on n'a pas obtenu de réponse positive :

1. On initialise $l = 0$ et on se donne un facteur $1 < c < 2$.
2. On augmente l de 1 et on pose $m = \lceil c^l \rceil$.
3. On tire au hasard $r \in [1, m]$.
4. On applique r itérations de l'algorithme de Grover sur un état initialisé $|\psi_0\rangle$.
5. On mesure l'état et on obtient le résultat i .

6. (a) Si i est solution, on arrête l'algorithme.
- (b) Si i n'est pas solution, on teste un autre $i \in \{0, 1\}^n$ au hasard. Si celui-ci n'est toujours pas solution, on repart à l'étape 2.

Avec ce protocole, deux cas sont possibles. Si le nombre réel de solutions k est tel que $k/N > 3/4$, on trouvera suffisamment rapidement une solution en testant des états au hasard dans la dernière étape. Dans le deuxième cas, $0 < k/N < 3/4$, on peut montrer que la probabilité de trouver une solution lors d'une itération de l'algorithme avec une gamme donnée par m est bornée inférieurement par $(1 - \sqrt{N/2m\sqrt{k}})/2$. Cette borne permet ensuite de montrer que le nombre total de requêtes nécessaire pour obtenir une solution est alors d'ordre $\mathcal{O}(\sqrt{N/k})$. Ainsi, trouver une solution se fera avec la même complexité algorithmique, que l'on connaisse ou non le nombre de solutions à l'avance.

8.2 Une accélération exponentielle ?

Dans ce chapitre, nous avons présenté deux problèmes pour lesquels l'utilisation de ressources quantiques permet une accélération. Cependant, les applications de ces problèmes restent limitées. En effet, l'algorithme de Deutsch-Jozsa nous donne une diminution exponentielle du nombre de requêtes par rapport au cas classique, mais il résout un problème *ad hoc* qui n'est pas vraiment utile en pratique (savoir si une fonction est équilibrée ou constante). L'algorithme de Grover, quant à lui, s'attaque au problème très utile de recherche dans une base de données non triée, mais nous donne une amélioration quadratique plus modeste. Les quelques problèmes difficiles qui peuvent en bénéficier restent donc difficiles malgré cette amélioration. Il paraît dès lors naturel de se demander si nous pouvons trouver mieux que ces deux premiers exemples : existe-t-il un algorithme quantique pour un problème connu permettant une amélioration exponentielle de la complexité, c'est-à-dire permettant de passer d'un problème actuellement difficile à un problème soluble en temps polynomial ?

En 1994, avec la découverte de l'algorithme de Simon [340], ce but a été atteint, dans le cas particulier d'une complexité en requêtes dans un algorithme utilisant un oracle. La même année, Shor a trouvé un premier algorithme avec une amélioration exponentielle de la complexité en temps par rapport à tous les algorithmes connus jusqu'à maintenant [335]. Nous allons présenter ces deux algorithmes ici. Cependant, avant de rentrer dans le détail de l'algorithme de Shor dans la section 8.2.4, nous présenterons en détail la partie réellement quantique de cet algorithme : la transformée de Fourier quantique (QFT). C'est son utilisation qui apporte une amélioration exponentielle de complexité en temps par rapport aux algorithmes connus, comme nous le verrons sur un exemple particulier avec l'algorithme d'estimation de phase quantique. Enfin, nous verrons à la fin de cette section que tous ces outils correspondent à des cas particuliers d'un problème plus général, appelé problème du sous-groupe caché.

8.2.1 Le problème de Simon

Tout comme l'algorithme de Deutsch-Jozsa, l'algorithme de Simon se base sur un problème à promesse. Avant de parler de l'algorithme en lui-même, nous allons introduire ce problème. Pour cela, on considère une fonction $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$, que l'on promet être telle qu'il existe $a \in \{0, 1\}^n$ non nul (c'est-à-dire différent de $0 \dots 0$) avec

$$f(x) = f(y) \Leftrightarrow (x = y \vee x = y \oplus a), \quad (8.24)$$

où $\oplus$ désigne l'addition bit à bit modulo deux. Pour chaque x , il existe un unique y différent de x , tel que $f(x) = f(y)$. De plus, l'idempotence de l'opération $\oplus$ assure $x \oplus y = a$. Le problème de Simon consiste, étant donné un oracle pour f , à trouver a .

Si l'on souhaite résoudre ce problème de façon probabiliste, il faut au moins effectuer $\mathcal{O}(\sqrt{N})$ requêtes afin d'obtenir le résultat avec une bonne probabilité. L'algorithme quantique de Simon va permettre de trouver a en $\mathcal{O}(\log N)$ requêtes. On peut de plus montrer que cette complexité en requêtes est optimale pour le cas quantique [238]. Pour obtenir ce résultat, l'algorithme de Simon va utiliser comme oracle une boîte noire quantique similaire à celles utilisées en début du chapitre, sans utiliser cette fois le codage de la fonction dans la phase. Si x et y sont des mots booléens de longueur n , ou autrement dit si $|x\rangle \otimes |y\rangle$ est un vecteur de la base canonique des qubits de longueur $2n$, alors on définit la boîte noire quantique qui calcule f par l'opérateur $U_f : |x\rangle \otimes |y\rangle \mapsto |x\rangle \otimes |y \oplus f(x)\rangle$. La seule différence avec les cas vus précédemment est qu'ici le second état n'est plus un état à un qubit mais un état à n qubits (la somme $\oplus$ est faite bit à bit).

La première étape de l'algorithme de Simon est d'appliquer cet opérateur à l'état $(H_{2^n} |0\rangle) \otimes |0\rangle$ pour obtenir

$$|\psi_{\text{out}}\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \otimes |f(x)\rangle. \quad (8.25)$$

On va alors mesurer les n derniers qubits, contenant l'information sur $f(x)$, et obtenir un résultat que l'on note z , qui est tel qu'il existe x_0 tel que $f(x_0) = z$. On a également, par les hypothèses sur f , $f(x_0 \oplus a) = z$, et l'état correspondant à ce résultat de mesure est donc

$$|\psi|_z\rangle \otimes |z\rangle = \frac{1}{\sqrt{2}} (|x_0\rangle + |x_0 \oplus a\rangle) \otimes |z\rangle. \quad (8.26)$$

On peut désormais oublier les qubits dans l'état $|z\rangle$ et se concentrer sur les n autres qubits. En appliquant l'opérateur de Hadamard à ces qubits, on obtient

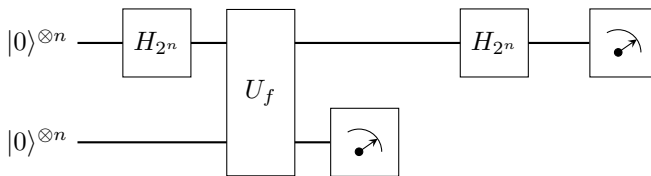


FIGURE 8.7 – Circuit de l’algorithme de Simon. On prépare un état initial $|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes n}$ puis on applique un opérateur de Hadamard aux n premiers qubits. On utilise ensuite un oracle qui écrit le résultat de la fonction f dans les n derniers qubits. On mesure alors ces qubits, obtenant un certain résultat z . Cette mesure a pour effet de projeter les n premiers qubits dans la superposition des deux états $|x_0\rangle$ et $|x_0 \oplus a\rangle$ tels que $f(x_0) = f(x_0 \oplus a) = z$. On applique ensuite l’opérateur de Hadamard à cette superposition de deux états, puis on mesure l’état obtenu, qui sera nécessairement un nombre y tel que $y \cdot a = 0$.

l’état :

$$H_{2^n} |\psi|_z\rangle = \frac{1}{\sqrt{2^{n+1}}} \sum_{y \in \{0,1\}^n} \left((-1)^{x_0 \cdot y} + (-1)^{(x_0 \oplus a) \cdot y} \right) |y\rangle \quad (8.27a)$$

$$= \frac{1}{\sqrt{2^{n-1}}} \sum_{\substack{y \in \{0,1\}^n \\ y \cdot a = 0}} (-1)^{x_0 \cdot y} |y\rangle. \quad (8.27b)$$

On voit que le résultat de cette procédure, représentée en circuit sur la figure 8.7, est de nous donner une superposition quantique de tous les états $|y\rangle$ tels que $y \cdot a = 0$. Si on mesure cet état, on va donc récupérer de manière aléatoire un nombre y tel que $y \cdot a = 0$, c’est-à-dire orthogonal à a , ou autrement dit une équation linéaire vérifiée par a . Obtenir $n - 1$ équations linéairement indépendantes nous permet de déterminer a . On peut montrer qu’en observant $(n-1)$ états orthogonaux à a au hasard, la probabilité d’obtenir $(n-1)$ équations linéairement indépendantes est plus grande que $1/4$ [202, 371]. Puisqu’une seule requête est nécessaire pour l’obtention d’une équation vérifiée par a , on a bien la complexité en $\mathcal{O}(\log N)$ annoncée.

8.2.2 La transformée de Fourier quantique

Les séries de Fourier permettent de décomposer un signal périodique de période T en la somme :

$$s(t) = \sum_{n \in \mathbb{N}} c_n e^{2i\pi n t / T}, \quad (8.28)$$

où c_n est appelé le n -ième coefficient de Fourier qui caractérise l’amplitude et la phase de l’onde à la fréquence n/T . Cela permet aussi de représenter sans

perte d'information un signal de durée finie T : il suffit de le prolonger à tous les temps en le rendant T -périodique.

Cependant, en pratique, numériser un signal implique de le discrétiser. Le signal discret S est alors une suite de N coefficients qui sont des instantanés du signal continu s ,

$$S(k) = s(kT/N), \text{ pour } k \in \mathbb{Z}/N\mathbb{Z}. \quad (8.29)$$

Comme k est un indice périodique, le cercle $\mathbb{R}/T\mathbb{Z}$ sur lequel est défini le signal continu périodique est remplacé par son équivalent discret, le groupe cyclique $G = \mathbb{Z}/N\mathbb{Z}$. Étant donné que les points sont échantillonnés avec une fréquence N/T , le théorème de Nyquist-Shannon nous assure que S ne peut décrire correctement les fréquences plus élevées que $N/2T$. Cela signifie notamment que l'on peut ne garder que N coefficients de Fourier, sans perte d'information.

Cette transformation s'appelle la *transformée de Fourier discrète*. Elle consiste à exprimer S sur la base orthonormée de fonctions

$$\chi_j : k \mapsto \frac{1}{\sqrt{N}} e^{2i\pi jk/N}, \text{ pour } j \in \mathbb{Z}/N\mathbb{Z}, \quad (8.30)$$

pour le produit scalaire

$$\langle f|g \rangle = \sum_{k=0}^{N-1} f^*(k)g(k). \quad (8.31)$$

Les coordonnées $(\beta_0, \dots, \beta_{N-1}) \in \mathbb{C}^N$ d'une fonction S dans cette base sont alors obtenues grâce au produit scalaire entre le signal S et la base χ_j par :

$$\beta_j = \langle \chi_j | S \rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{-2i\pi kj/N} S(k). \quad (8.32)$$

Cette transformée de Fourier discrète peut alors se décliner à de nombreux groupes abéliens, ce qui permet son application dans un grand nombre de cas. Elle est très utile dans la manipulation des signaux échantillonnés, comme les images numériques. De plus, alors que l'algorithme naïf pour calculer la transformée de Fourier discrète d'un groupe de taille N a une complexité en $\mathcal{O}(N^2)$, il existe un algorithme de transformée de Fourier rapide (FFT) qui a une complexité en $\mathcal{O}(N \log N)$ [171, 109, 300]. Cet algorithme très célèbre est à la base des techniques de compression numérique qui ont mené au format JPEG. Il est également utilisé dans des algorithmes de multiplication rapide des grands entiers [322] et des polynômes [300].

Nous allons maintenant nous intéresser à la transformée de Fourier quantique, qui n'est autre que la réécriture de la transformée de Fourier discrète dans un formalisme quantique. Contrairement à l'algorithme classique, le but

ne va pas être de calculer explicitement les coefficients β_j , mais de transformer un état quantique à n qubits $|x\rangle$, avec $x \in \{0, 1\}^n$, en l'état ⁵

$$\text{QFT}_{2^n} |x\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2i\pi xk/2^n} |k\rangle. \quad (8.33)$$

Ainsi, si l'on part d'un état $|\psi\rangle = \sum_k \alpha_k |k\rangle$ et qu'on lui applique l'opérateur QFT, l'état transformé a pour amplitudes les coefficients de Fourier β_k qui sont obtenus à l'aide de la transformée de Fourier discrète de $f : k \mapsto \alpha_k$. Il faut prendre quelques précautions, car nous n'avons pas accès à cet ensemble d'amplitudes de manière directe, puisqu'il est impossible de caractériser l'état quantique obtenu en un nombre fini de mesures. Cependant, comme nous le verrons dans les sections suivantes, cet état peut être utilisé comme étape intermédiaire pour certains algorithmes. Nous allons montrer ici que l'opérateur QFT_{2^n} peut se construire par récurrence et donner l'état final avec une complexité de $\mathcal{O}((\log N)^2)$ pour agir sur n qubits ⁶.

Pour construire cet opérateur, nous avons besoin d'un nouvel opérateur sur deux qubits, que l'on appelle *controlled phase shift* (CPS). Cet opérateur va appliquer l'opérateur de déphasage $R_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}$ au qubit considéré si et seulement si le qubit de contrôle se trouve dans l'état $|1\rangle$, ce qui se traduit par l'équation

$$\text{CPS}_\phi |x_1\rangle |x_0\rangle = e^{i\phi x_1 x_0} |x_1\rangle |x_0\rangle. \quad (8.34)$$

Comme on peut le voir dans cette équation, savoir lequel de ces qubits est le contrôle n'a pas d'importance réelle, on a uniquement besoin de savoir quels sont les deux qubits reliés par cet opérateur.

Notre but est de construire un circuit permettant de créer l'état $\text{QFT}_{2^n} |x\rangle$. Tout d'abord, remarquons que pour le système à un qubit, on a

$$\text{QFT}_2 |x\rangle = \frac{1}{\sqrt{2}} \sum_{k=0}^1 e^{i\pi xk} |k\rangle = \frac{|0\rangle + (-1)^x |1\rangle}{\sqrt{2}}. \quad (8.35)$$

Cela correspond exactement à l'action de l'opérateur de Hadamard sur un qubit, comme vu à l'équation (8.5). Ainsi, $\text{QFT}_2 = H_2$.

Pour construire les opérateurs suivants, nous allons montrer par récurrence la formule suivante :

$$\text{QFT}_{2^n} |x\rangle = \frac{|0\rangle + e^{i\pi x/2^0} |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + e^{i\pi x/2^1} |1\rangle}{\sqrt{2}} \otimes \dots \otimes \frac{|0\rangle + e^{i\pi x/2^{n-1}} |1\rangle}{\sqrt{2}}. \quad (8.36)$$

5. La convention de signe est différente de celle introduite au-dessus, afin de se rapprocher de la littérature du calcul quantique.

6. D'autres algorithmes, plus complexes, permettent même de descendre jusqu'à une complexité de $\mathcal{O}(\log(N) \log(\log N))$.

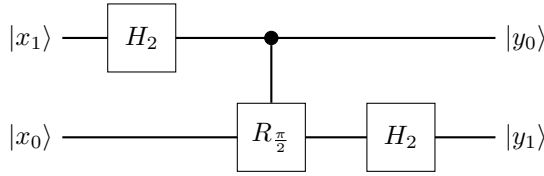


FIGURE 8.8 – Circuit de l'opérateur QFT_4 , permettant de donner les qubits $|y_1 y_0\rangle = \text{QFT}_4 |x_1 x_0\rangle$. On voit que cet opérateur est simple à construire et ne dépend que des opérateurs de Hadamard et d'un opérateur qui change la phase si le qubit de contrôle est dans l'état $|1\rangle$.

Notons qu'ici le premier facteur n'est sensible qu'au bit le moins significatif de l'entier x , car les autres bits génèrent des phases multiples de 2π . De même, le n -ième facteur est sensible aux n bits les moins significatifs de x .

Nous avons montré cette formule pour $n = 1$. Pour $n > 2$, il suffit de diviser la somme sur tous les k en deux, selon que k est pair ou impair :

$$\text{QFT}_{2^n} |x\rangle = \frac{1}{\sqrt{2^n}} \sum_{k'=0}^{2^{n-1}-1} e^{\frac{2i\pi x 2k'}{2^n}} |k'\rangle \otimes |0\rangle + \frac{1}{\sqrt{2^n}} \sum_{k'=0}^{2^{n-1}-1} e^{\frac{2i\pi x (2k'+1)}{2^n}} |k'\rangle \otimes |1\rangle \quad (8.37a)$$

$$= \frac{1}{\sqrt{2^{n-1}}} \sum_{k'=0}^{2^{n-1}-1} e^{2i\pi x k' / 2^{n-1}} |k'\rangle \otimes \frac{|0\rangle + e^{i\pi x / 2^{n-1}} |1\rangle}{\sqrt{2}}, \quad (8.37b)$$

ce qui finit la récurrence et permet de démontrer l'équation (8.36). Peut-on désormais se servir de cette équation pour proposer un circuit quantique permettant de mettre en œuvre l'opérateur de transformée de Fourier quantique ? Pour répondre à cette question, le plus simple est de commencer par l'opérateur QFT_4 , qui transforme l'état à deux qubits $|x\rangle = |x_1 x_0\rangle$ en

$$\text{QFT}_4 |x_1 x_0\rangle = \frac{|0\rangle + e^{i\pi x_0} |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + e^{i\pi x_1} e^{i\pi x_0/2} |1\rangle}{\sqrt{2}}. \quad (8.38)$$

Il est très facile de démontrer que cet état peut être obtenu en utilisant le circuit représenté sur la figure 8.8, à un détail près : les qubits en sortie de circuit sont dans l'ordre inverse de l'ordre naturel. En effet, ce circuit applique l'opérateur $(\mathbb{1} \otimes H_2) \text{CPS}_{\frac{\pi}{2}} (H_2 \otimes \mathbb{1})$ à l'état $|x_1 x_0\rangle$. Un calcul direct de l'état en sortie de circuit nous donne alors l'état

$$\frac{|0\rangle + e^{i\pi x_1} e^{i\pi x_0/2} |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + e^{i\pi x_0} |1\rangle}{\sqrt{2}}, \quad (8.39)$$

qui est bien l'état $\text{QFT}_4 |x\rangle$ à une inversion des qubits près.

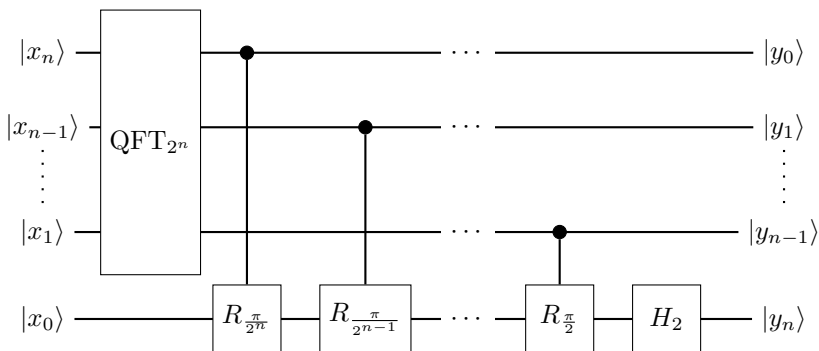


FIGURE 8.9 – Circuit de l'opérateur $\text{QFT}_{2^{n+1}}$, agissant sur $n + 1$ qubits. On le construit par récurrence à partir de l'opérateur agissant sur n qubits et de n opérateurs de déphasages contrôlés par les n premiers qubits. On peut vérifier facilement que l'opérateur QFT_4 vérifiait cette forme, avec $\text{QFT}_2 = H_2$.

Dans le cas général, l'opérateur de transformée de Fourier quantique peut s'exprimer à l'aide du circuit représenté sur la figure 8.9, qui utilise à nouveau l'idée d'une construction par récurrence. On peut démontrer que ce circuit redonne bien l'état donné dans l'équation (8.36) à une inversion de l'ordre des qubits près⁷.

Du point de vue de la complexité du circuit ainsi obtenu, on voit que pour passer de l'opérateur QFT_{2^n} à l'opérateur $\text{QFT}_{2^{n+1}}$, on va devoir effectuer $n + 1$ opérations : n changements contrôlés de phase, et une porte de Hadamard sur le qubit supplémentaire. QFT_{2^n} nécessite donc $\sum_{k=0}^n k = n(n+1)/2$ portes, ce qui donne bien une complexité en $\mathcal{O}((\log N)^2)$ comme nous l'avons affirmé au début de cette section. Mais, comme nous l'avons indiqué précédemment, cette complexité réduite ne nous apporte rien si notre but est de calculer la valeur de tous les coefficients de Fourier, car nous avons ici obtenu un état quantique et non un résultat qui serait la liste de ces coefficients. Nous allons montrer dans les deux prochaines sections qu'il est cependant possible d'utiliser l'état quantique obtenu pour effectuer certains calculs qui auraient nécessité une transformation de Fourier discrète, sans avoir besoin de connaître la valeur de tous les coefficients. Ce sont ces algorithmes complets, utilisant la transformée de Fourier quantique comme outil intermédiaire qui peuvent alors bénéficier d'une accélération quantique.

8.2.3 L'estimation de phase quantique

Un premier algorithme utilisant la transformée de Fourier quantique est celui d'estimation de phase quantique [335, 228], dont le but est de trouver la

7. Cette inversion d'ordre est également présente dans l'algorithme classique de la transformée de Fourier rapide.

phase ajoutée par un opérateur unitaire lorsqu'il agit sur un de ses vecteurs propres. Plus précisément, on appelle U l'opérateur unitaire agissant sur m qubits, et $|\psi\rangle$ le vecteur propre de U tel que

$$U |\psi\rangle = e^{2i\pi\theta} |\psi\rangle \text{ avec } 0 \leq \theta < 1. \quad (8.40)$$

Le but de cet algorithme est de trouver θ , avec une bonne précision si possible.

En termes physiques, l'estimation de phase quantique permet de trouver l'énergie d'un état propre d'une dynamique. Au-delà de ça, c'est un outil qui peut servir pour de nombreux usages. Il est notamment utilisé pour les algorithmes de chimie quantique, ainsi qu'en métrologie quantique [118]. C'est également la brique de base pour d'autres algorithmes quantiques, tels que l'algorithme de Shor dont nous allons parler ensuite.

L'algorithme quantique

L'algorithme quantique va nous permettre de trouver avec une bonne probabilité l'approximation de la valeur θ sur n bits. Par ailleurs, la distribution de probabilité est concentrée autour des valeurs de θ proches de l'optimale. Ainsi, en répétant l'algorithme plusieurs fois et en donnant la valeur qui apparaît le plus souvent, on verra que l'on peut obtenir une bonne approximation de θ avec une probabilité arbitrairement grande.

L'opérateur U agissant sur m qubits, l'algorithme opérera sur un ensemble de $n + m$ qubits. L'idée générale consiste à générer une superposition d'états dans laquelle chaque état apparaît avec une phase multiple de $2\pi\theta$, puis d'utiliser la transformée de Fourier quantique afin d'extraire cette période. Pour cela, on commence dans l'état initial $|0\rangle^{\otimes n} \otimes |\psi\rangle$. Comme pour tous les autres algorithmes, on va commencer par appliquer l'opérateur de Hadamard H_{2^n} aux n premiers qubits, pour créer l'état $\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right)^{\otimes n} |\psi\rangle$. On va ensuite appliquer n opérations utilisant U sur l'état $|\psi\rangle$, chacune contrôlée par la valeur d'un des qubits du premier registre. Plus précisément, le qubit i va contrôler l'application de l'opérateur $U^{2^{n-i}}$, comme cela est représenté sur le circuit quantique 8.10.

Après cette étape, l'état est donné par

$$\begin{aligned} \frac{1}{\sqrt{2^n}} (|0\rangle + e^{2i\pi 2^{n-1}\theta} |1\rangle) \otimes (|0\rangle + e^{2i\pi 2^{n-2}\theta} |1\rangle) \otimes \cdots \otimes (|0\rangle + e^{2i\pi\theta} |1\rangle) \otimes |\psi\rangle \\ = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2i\pi k\theta} |k\rangle \otimes |\psi\rangle \end{aligned} \quad (8.41)$$

et on peut oublier désormais la partie correspondant aux m qubits dans l'état $|\psi\rangle$ et se concentrer sur les n premiers qubits, qui sont dans une superposition dépendante de θ , auxquels on va appliquer une opération de transformée de

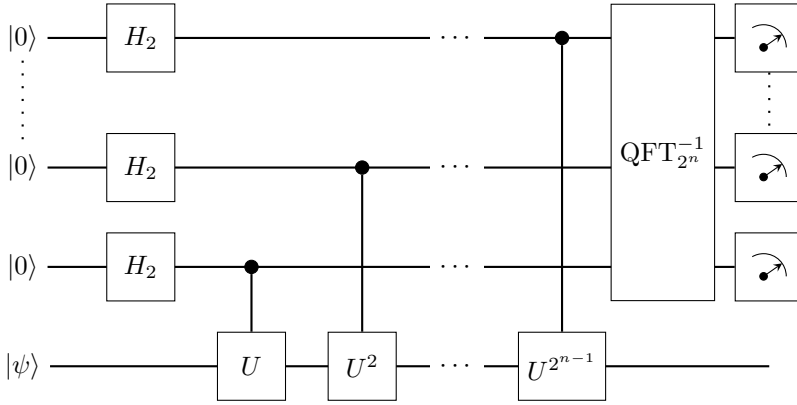


FIGURE 8.10 – Circuit quantique de l’algorithme d’estimation de phase. L’opérateur U agit sur son vecteur $|\psi\rangle$ par $U|\psi\rangle = e^{i\theta}|\psi\rangle$, et on l’applique un certain nombre de fois en fonction de la valeur des qubits de contrôle. La mesure finale donne une bonne estimation de la valeur de θ .

Fourier quantique inverse⁸. L’état après cette opération est

$$\frac{1}{2^n} \sum_{k=0}^{2^n-1} \sum_{j=0}^{2^n-1} e^{2i\pi k(\theta-j/2^n)} |j\rangle. \quad (8.42)$$

On mesure ensuite l’ensemble des qubits. Cela va nous donner le nombre j qui correspond à estimer θ par $j/2^n$. Afin que l’estimation soit bonne, il est nécessaire que la probabilité de mesurer j tel que $j/2^n \approx \theta$ soit élevée. La probabilité de mesurer un j particulier est donnée par

$$p(j) = \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{2i\pi k(\theta-j/2^n)} \right|^2. \quad (8.43)$$

Si l’on écrit $\theta = j_{\text{opt}}/2^n$, la question est alors de savoir avec quelle probabilité on trouve un entier j qui est une bonne approximation du réel α . On peut alors distinguer deux cas.

Cas où $2^n\theta$ est entier Dans ce cas, on peut réécrire la probabilité de mesure sous la forme

$$p(j) = \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{2i\pi k(2^n\theta-j)/2^n} \right|^2 = \begin{cases} 1 & \text{si } j = 2^n\theta \\ 0 & \text{sinon,} \end{cases} \quad (8.44)$$

8. Elle se construit comme l’opération de QFT mais en appliquant des rotations d’angle opposé.

où la dernière égalité s'obtient en observant qu'il s'agit alors de la somme des racines n -ièmes de l'unité. Dans ce cas, le j que l'on obtient comme résultat de mesure est égal avec probabilité 1 à $2^n\theta$, et on a donc pu trouver la phase que l'on recherchait avec une précision de n bits.

Cas où $2^n\theta$ n'est pas entier. Dans ce cas, on peut utiliser les formules habituelles pour la somme des termes d'une suite géométrique et obtenir la forme suivante pour la probabilité de mesure

$$p(j) = \frac{1}{2^{2n}} \left| \frac{e^{2i\pi 2^{2n}(\theta - j/2^n)} - 1}{e^{2i\pi(\theta - j/2^n)} - 1} \right|^2. \quad (8.45)$$

Ce que l'on va chercher à montrer désormais, c'est que cette probabilité est élevée pour les j proches de $2^n\theta$ et faible ailleurs [101, 371, 202].

Si l'on mesure j , on va décomposer $\theta = j/2^n + \epsilon$. Ainsi, on peut étudier les différents cas en mettant une contrainte sur l'erreur ϵ . Commençons par regarder la probabilité d'obtenir un résultat j_0 éloigné de la vraie valeur de $2^n\theta$. Pour cela, on introduit un entier α qui minore l'erreur lorsque l'on approxime $2^n\theta$ par j_0

$$\frac{\alpha}{2^n} \leq |\epsilon| < \frac{1}{2}. \quad (8.46)$$

On a donc

$$p(j_0) = \frac{1}{2^{2n}} \left| \frac{e^{2i\pi 2^{2n}\epsilon} - 1}{e^{2i\pi\epsilon} - 1} \right|^2 \leq \frac{1}{2^{2n}} \left| \frac{2}{4\epsilon} \right|^2, \quad (8.47)$$

où l'on a utilisé le fait que le numérateur est nécessairement plus petit que 2, tandis que le dénominateur vérifie l'inégalité $|e^{2i\pi\epsilon} - 1| \geq 4\epsilon$, comme cela est visible sur la figure 8.11. Ainsi, la probabilité d'obtenir un résultat j_0 éloigné d'au moins α de $2^n\theta$ est

$$p(j_0) \leq \frac{1}{4\alpha^2}. \quad (8.48)$$

On a donc bien une très faible probabilité de trouver un résultat très éloigné de la valeur recherchée. Cela signifie qu'il n'y a pas de phénomène de concentration autour des j_0 éloignés.

À l'inverse, quelle est la probabilité de mesurer le résultat optimal j_1 , avec la précision maximale de n bits ? Cela impose de prendre

$$|\epsilon| \leq \frac{1}{2^{n+1}}. \quad (8.49)$$

La probabilité associée peut donc se réécrire

$$p(j_1) = \frac{1}{2^{2n}} \left| \frac{e^{2i\pi 2^{2n}\epsilon} - 1}{e^{2i\pi\epsilon} - 1} \right|^2. \quad (8.50)$$

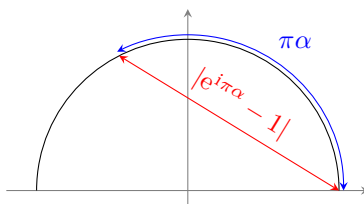


FIGURE 8.11 – Représentation des bornes utilisées ici. On ne regarde que des nombres de la forme $e^{i\pi\alpha} - 1$ avec $\alpha < 1$. Le module d'un tel nombre est donné par la longueur du segment allant du point $(1, 0)$ au point $(\cos(\pi\alpha), \sin(\pi\alpha))$. Comme on le voit sur cette figure, cette longueur (en rouge) est plus petite que la longueur de l'arc de cercle de rayon unité joignant ces deux points (en bleu). Mais le ratio de ces deux longueurs est aussi borné par la valeur prise pour $\alpha = 1$. On a donc les inégalités suivantes $1 \leq \frac{|e^{i\pi\alpha} - 1|}{\pi\alpha} \leq \frac{2}{\pi}$.

En se servant toujours de l'argument présenté sur la figure 8.11, on sait que le numérateur est plus grand que $2^n 4\epsilon$, tandis que le dénominateur est lui plus petit que $2\pi\epsilon$. En mettant ces bornes ensemble, on a donc

$$p(j_1) \geq \frac{4}{\pi^2} \simeq 0,405, \quad (8.51)$$

ce qui montre bien que l'on a une forte probabilité de trouver un résultat optimal.

Précision du résultat et complexité

Deux questions se posent naturellement à partir de cet algorithme : comment obtenir une précision arbitraire sur θ , et quel est le coût en temps d'une telle procédure ?

Imaginons que l'on veuille obtenir θ avec une précision d'au moins k bits, c'est-à-dire avec une erreur plus faible que $1/2^k$. Le plus simple dans ce cas est d'appliquer l'algorithme ci-dessus avec $n > k$. Par exemple, pour $n = k + 2$, on sait qu'au moins un résultat précis à $k + 2$ bits, a une probabilité supérieure à $4/\pi^2$ d'être obtenu, tandis que ceux éloignés avec une erreur supérieure à $1/2^k$ sont obtenus avec une probabilité inférieure à $1/64$. Cela assure que si l'on répète l'algorithme complet et que l'on donne la réponse majoritaire, celle-ci a une très faible chance d'avoir une précision inférieure à k bits. En fait, si on arrondit la valeur de θ à k bits, on peut montrer que la probabilité de réussite tend vers 1 exponentiellement vite avec le nombre de répétitions de l'algorithme. Plus quantitativement, on montre que l'on peut obtenir une estimation de θ précise à $k + 1$ bits avec une probabilité d'au moins $1 - \epsilon$ en appliquant l'algorithme $n = k + \lceil \log(1/2\epsilon + 1/2) \rceil$ fois [101]. De plus, on n'a besoin que d'une copie de l'état $|\psi\rangle$ pour effectuer plusieurs itérations de l'algorithme, puisque celui-ci est laissé inchangé dans notre procédure.

8.2.4 Décomposer un nombre en facteurs premiers

Nous sommes désormais armés pour étudier ce qui constitue l'algorithme quantique le plus connu, y compris du grand public : l'algorithme de factorisation d'entiers découvert en 1994 par Peter Shor [335, 336]. Il a eu un impact retentissant sur le développement de la recherche en calcul quantique car il offre une solution polynomiale à un problème qui est encore actuellement considéré comme difficile pour un ordinateur classique, et qui est de plus utilisé dans la vie de tous les jours. En effet, un des protocoles de cryptographie asymétrique les plus utilisés, le protocole RSA, est construit sur la difficulté présumée à retrouver les facteurs premiers des grands nombres entiers. Le résultat principal de l'algorithme de Shor est que l'on va passer d'une complexité classique de $2^{\mathcal{O}(\log(N)^{1/3} \log(\log(N))^{2/3})}$ pour le meilleur algorithme de factorisation connu actuellement à une complexité en $\mathcal{O}(\log(N)^3)$.

L'algorithme de Shor est construit autour d'un algorithme classique, mais utilise un procédé équivalent à l'estimation de phase quantique, et donc la transformée de Fourier quantique, pour effectuer une des étapes de manière exponentiellement plus rapide. Cela est possible car l'opération à mettre en œuvre dans l'estimation de phase peut l'être de manière efficace. Dans cette section, nous allons commencer par présenter toutes les étapes de l'algorithme, ainsi qu'une justification de pourquoi cette méthode fonctionne, avant de montrer plus en détail la partie quantique.

Les étapes de l'algorithme de Shor

On cherche à trouver un facteur premier d'un nombre N que l'on sait n'être pas premier, ce qui peut se tester avec un algorithme probabiliste polynomial en $\log N$ [271, 306].

On commence par deux étapes rapides permettant de résoudre ce problème dans certains cas :

1. si N est pair, il suffit de répondre 2, et l'algorithme est terminé ;
2. on teste si N est de la forme a^b . Pour cela, il suffit de regarder si la racine k -ième de N est entière, pour $k \leq \log N$. L'extraction de racine peut se faire en temps polynomial à l'aide d'une recherche dichotomique sur l'intervalle $\{0, \dots, N\}$. On peut donc tester l'ensemble des racines k -ième en temps polynomial. Si on trouve un résultat positif, deux cas sont possibles. Soit a est premier, auquel cas il suffit de renvoyer a , soit a n'est pas premier et on recommence alors l'algorithme avec a au lieu de N .

Si ces deux étapes ne sont pas concluantes, alors on va utiliser l'algorithme suivant.

1. On choisit au hasard un nombre a dans $\{2, \dots, N - 1\}$.

2. On calcule $z = \text{pgcd}(a, N)$, par exemple avec l'algorithme d'Euclide, qui est polynomial. Si $z \neq 1$, alors z est un diviseur de N et on a une réponse.
3. On cherche l'ordre de a , noté r , qui est le plus petit entier tel que $a^r = 1 \pmod{N}$. C'est cette étape qui va pouvoir être réalisée sur une architecture quantique.
4. Si r est impair, l'algorithme a échoué, et on peut repartir de l'étape 1.
5. On calcule $a^{r/2}$. Si $a^{r/2} = -1$, alors l'algorithme a échoué et on repart de l'étape 1.
6. Sinon, $z_{\pm} = \text{pgcd}(a^{r/2} \pm 1, N)$ sont deux facteurs non triviaux de N , et on a notre réponse.

Pour montrer que cet algorithme fonctionne, nous devons montrer que les deux nombres $z_{\pm}$ définis ci-dessus sont bien des facteurs non triviaux de N . L'entier r est défini de telle sorte que $a^r = 1 \pmod{N}$, donc $a^r - 1$ est divisible par N . Or, on peut écrire $a^r - 1 = (a^{r/2} - 1)(a^{r/2} + 1)$ ⁹, ce qui signifie que l'un de ces deux nombres au moins possède un facteur en commun avec N . On peut alors l'extraire en calculant les pgcd correspondants. De plus, la définition de l'ordre r nous affirme que N n'est pas diviseur de $a^{r/2} - 1$, car $r/2$ est inférieur à r . Enfin la condition de l'étape 5 nous affirme que N n'est pas diviseur de $a^{r/2} + 1$. Ainsi, les deux nombres $z_{\pm}$ sont nécessairement différents de N .

Par ailleurs, la probabilité d'échouer doit être assez petite. Si N contient l facteurs premiers impairs distincts, alors celle-ci est $1/2^{l-1}$. En éliminant les facteurs pairs ainsi que les cas où $N = a^b$ avant de faire la recherche de période, on s'assure que la probabilité d'échec est au plus $1/2$. Il faut donc tirer au hasard peu d'entiers a afin d'avoir un facteur non trivial pour N .

À titre d'exemple, prenons le nombre $N = 143$. Après un premier tirage, nous obtenons $a = 62$, premier avec N , dont la période est $r = 30$. Cependant, on a $a^{r/2} = -1 \pmod{N}$ et l'algorithme échoue. Sur un second tirage, nous obtenons $a = 74$, dont la période est également $r = 30$. On obtient $a^{r/2} = 131 \pmod{N}$. En regardant le plus grand dénominateur commun de 143 respectivement avec 130 et 132, on obtient les deux facteurs premiers $13 \times 11 = 143$.

Le but de l'algorithme de Shor est donc de permettre le calcul de l'ordre de a . Pour cela, définissons la fonction $f : x \mapsto a^x \pmod{N}$. Au vu de sa définition, r est le plus petit entier tel que $f(x) = f(x + r)$. Ainsi, trouver r revient à trouver la période de la fonction f , et c'est cela que l'on va chercher à réaliser en quantique.

Trouver la période de f

La vraie avancée de l'algorithme de Shor repose donc sur le calcul de la période de f à l'aide de ressources quantiques. Comme on va le voir, il y a

9. r est pair par l'étape 4.

un grand nombre de similarités entre ce problème et celui de l'estimation de phase que l'on a déjà présenté. Pour cet algorithme, on va devoir utiliser $2q$ qubits, avec q tel que $N^2 \leq 2^q < 2N^2$ pour pouvoir trouver l'ordre avec forte probabilité. On va ensuite placer les q premiers qubits dans une superposition en appliquant l'opérateur de Hadamard à ces qubits dans l'état $|0\rangle$. On obtient l'état

$$|\psi\rangle = \frac{1}{\sqrt{2^q}} \sum_{k \in \{0,1\}^q} |k\rangle |0\rangle. \quad (8.52)$$

Puis on applique un opérateur qui va réaliser f quantiquement, c'est-à-dire effectuer l'action suivante

$$|k\rangle |0\rangle \rightarrow |k\rangle |a^k \pmod{N}\rangle. \quad (8.53)$$

Comment mettre en œuvre un tel opérateur ? En classique, la meilleure méthode pour calculer $a^k \pmod{N}$ consiste à calculer tous les $a^{2^i} \pmod{N}$ avec $i \leq \log k$ en mettant le résultat obtenu pour $i-1$ au carré, puis à multiplier les résultats correspondant à l'écriture binaire de k (toujours $\pmod{N}$). Ici, la méthode est équivalente et nous rappelle tout à fait celle mise en place dans le cas de l'algorithme d'estimation de phase : pour pouvoir appliquer notre opérateur d'exponentiation modulaire avec toutes les puissances k possibles, il suffit d'appliquer des opérateurs calculant a^{2^i} contrôlés par le $(n-i)$ -ième qubit. On a donc exactement le même début de circuit quantique que celui de la figure 8.10, en changeant les opérateurs U^{2^i} par des opérateurs multipliant l'état du qubit par $a^{2^i} \pmod{N}$. La question est donc celle du circuit permettant une telle opération, qui est en fait assez compliquée, et nous emmènerait trop loin pour ce livre. Néanmoins, de nombreux exemples de mises en œuvre de tels opérateurs à l'aide d'opérations quantiques usuelles existent [336, 364, 264]. Le point important pour la complexité totale de cet algorithme est que cette étape peut être effectuée en $\mathcal{O}(q^3)$ opérations pour $2q$ qubits¹⁰.

En appliquant cet opérateur à notre état initial, on obtient donc l'état

$$|\psi'\rangle = \frac{1}{\sqrt{2^q}} \sum_{k \in \{0,1\}^q} |k\rangle |a^k \pmod{N}\rangle \quad (8.54a)$$

$$= \frac{1}{\sqrt{2^q}} \sum_{l=0}^{r-1} \sum_{m=0}^{s_l-1} |mr+l\rangle |a^l \pmod{N}\rangle, \quad (8.54b)$$

où l'on est passé à la deuxième égalité en séparant les états en fonction de la valeur a^l prise dans le deuxième registre de qubits, r valeurs différentes

10. Il est également possible d'utiliser l'algorithme de multiplication de Schönhage-Strassen [322] afin de réaliser cette opération en $\mathcal{O}(q^2(\log q)(\log \log q))$. Cependant, les nombres pour lesquels cela est intéressant sont bien plus grands que les nombres utilisés en pratique dans RSA.

pouvant exister. Pour chacune de ces valeurs, s_l est le nombre d'éléments x tels que $f(x) = a^l$.

On va alors appliquer l'opération de transformée de Fourier quantique à notre premier registre de q qubits, et obtenir l'état

$$\frac{1}{2^q} \sum_{l=0}^{r-1} \sum_{m=0}^{s_l-1} \sum_{p=0}^{2^q-1} e^{\frac{2i\pi}{2^q} p(mr+l)} |p\rangle |a^l \pmod{N}\rangle. \quad (8.55)$$

On mesure ensuite l'état de nos deux registres. La probabilité de mesurer l'état $|p_0\rangle |a^{l_0}\rangle$ est donnée par

$$p(|p_0\rangle |a^{l_0}\rangle) = \left| \frac{1}{2^q} \sum_{m=0}^{s_l-1} e^{\frac{2i\pi}{2^q} p_0(mr+l_0)} \right|^2 = \left| \frac{1}{2^q} \sum_{m=0}^{s_l-1} e^{\frac{2i\pi}{2^q} p_0 mr} \right|^2. \quad (8.56)$$

Et on peut montrer, à l'aide du même type d'arguments que ceux utilisés pour l'estimation de phase, que cette probabilité est plus grande que $4/(\pi^2 r)$ [202] dès que le reste de la division de $p_0 r$ par 2^q est dans $[-r/2, +r/2]$, c'est-à-dire dès qu'il existe un entier d tel que

$$\left| \frac{p_0}{2^q} - \frac{d}{r} \right| \leq \frac{1}{2^{q+1}}. \quad (8.57)$$

On a donc une forte probabilité de trouver comme résultat p_0 une valeur qui soit très proche d'une fraction de la forme d/r . De plus, puisque $2^q > N^2$, il y a au plus une seule fraction de la forme d/r satisfaisant l'inégalité (8.57) avec $r < N$. Il existe alors des méthodes de calcul classiques pour trouver cette fraction en temps polynomial, et ainsi r . Enfin, il y a r valeurs possibles pour le nombre d , et donc pour les p_0 avec une forte probabilité de mesure. Au total, on peut montrer que l'algorithme de Shor permet de remonter à une fraction de la forme d/r avec une probabilité supérieure à $4/\pi^2$ [202], et donc que quelques itérations au plus seront nécessaires pour trouver la période r qui nous intéresse. Pour améliorer la précision, on peut toujours utiliser les mêmes méthodes que celles présentées pour l'estimation de phase.

Complexité de l'algorithme quantique Comme on l'a vu, l'algorithme quantique nécessite de faire appel un certain nombre de fois à une boucle calculant la période d'une fonction spécifique. Dans cette boucle, il est nécessaire de faire quelques appels à une transformée de Fourier quantique dont on sait que la complexité est en $\mathcal{O}(\log N)$, mais surtout de faire appel à l'opérateur permettant d'effectuer l'opération d'exponentiation modulaire, c'est-à-dire de mettre en œuvre le calcul de la fonction f . La meilleure mise en œuvre connue actuelle étant en $\mathcal{O}((\log N)^3)$, c'est cette complexité qui domine l'algorithme total.

8.2.5 Au-delà de l'algorithme de Shor

La question naturelle à se poser maintenant que nous avons vu plusieurs exemples d'algorithmes quantiques est celle du cas général : quelles sont les catégories de problème qui peuvent bénéficier d'une amélioration lorsque l'on se donne des ressources quantiques ? À l'heure actuelle, deux grandes catégories de problèmes sont ressorties : les problèmes de recherche, comme celui de Grover ou de Deutsch-Jozsa (qui peut être vu comme le moyen de chercher les fonctions constantes parmi toutes les fonctions équilibrées), et les problèmes de sous-groupe caché que nous allons discuter ici.

Le problème du sous-groupe caché

Les deux autres exemples principaux que nous avons présentés ici, l'algorithme de Simon et celui de Shor, sont en fait des cas particuliers d'un problème plus générique appelé problème du sous-groupe caché. Ce problème se formalise de la manière suivante : pour une fonction d'un groupe fini $(G, \cdot)$, à valeurs dans un ensemble X , on cherche un sous-groupe $K \subseteq G$ tel que $f(x) = f(y)$ si et seulement s'il existe $k \in K$ et $y = x \cdot k$. Le groupe K est alors appelé sous-groupe caché par f . Commençons par montrer que nos deux exemples rentrent bien dans cette définition.

- Pour l'algorithme de Simon, l'analogie est directe. f est une fonction de $G = (\mathbb{Z}/2\mathbb{Z})^n$ dans $X = \{0, 1\}^n$, la loi considérée sur G est l'addition bit à bit modulo 2 dénotée $\oplus$, et on a la relation $f(x) = f(y) \Leftrightarrow (y = x \oplus s \vee y = x)$, ce qui se traduit par $K = \{0, s\}$.
- Pour l'algorithme de Shor, on peut se ramener à ce problème en considérant le groupe multiplicatif $\mathbb{Z}/\varphi(N)\mathbb{Z}$ des entiers modulo $\varphi(N)$, avec $\varphi(N)$ le nombre d'entiers plus petits que N et premiers avec N . La loi de composition est alors l'addition modulo $\varphi(N)$, et on cherche bien les éléments r tels que $f(x+r) = f(x)$ avec f l'opération d'exponentiation modulaire. Le sous-groupe K est alors $\{0, r, 2r, \dots\}$. Cette méthode fonctionne car on peut montrer que pour tout a premier avec N , $a^{\varphi(N)} = 1 \pmod{N}$.

Le premier résultat important de ce problème du sous-groupe caché est que l'on peut montrer que la complexité en requêtes de ce problème est polynomiale, et plus précisément en $\mathcal{O}((\log |G|)^2)$, où $|G|$ dénote le cardinal de G [150].

✂ L'idée est la même que celle vue jusqu'ici. Si l'on dispose d'un opérateur U agissant comme $U|x\rangle|y\rangle = |x\rangle|y \cdot f(x)\rangle$, on commence par créer l'état superposé sur tous les éléments du groupe, $\psi = 1/\sqrt{|G|} \sum_{g \in G} |g\rangle|0\rangle$, avec $|G|$ le cardinal de G , puis on applique l'opérateur U à cet état. En observant le deuxième registre, on obtient alors un état de la forme

$$|xK\rangle|f(x)\rangle = \frac{1}{\sqrt{|K|}} \sum_{k \in K} |x \cdot k\rangle|f(x)\rangle. \quad (8.58)$$

On sait de plus que le nombre maximal de sous-groupes distincts de G , noté r , est d'ordre $2^{\mathcal{O}((\log |G|)^2)}$, car un sous-groupe de G peut être généré par au plus $\log |G|$ éléments. On peut alors ordonner nos r sous-groupes en les notant $K_1, K_2, \dots, K_r$, et on prend généralement $|K_1| \geq |K_2| \geq \dots \geq |K_r|$. Notre problème est de savoir lequel de ces sous-groupes est le sous-groupe K caché par notre fonction f . Pour cela, on a besoin de savoir quelle est la probabilité de se tromper de sous-groupe étant donné un état de la forme $|xK\rangle$, c'est-à-dire la valeur de $|\langle xK_i | xK_j \rangle|^2$,

$$\langle xK_i | xK_j \rangle = \frac{1}{\sqrt{|K_i||K_j|}} \sum_{k_i \in K_i} \sum_{k_j \in K_j} \langle x \cdot k_i | x \cdot k_j \rangle \quad (8.59a)$$

$$= \frac{|K_i \cap K_j|}{\sqrt{|K_i||K_j|}}, \quad (8.59b)$$

$$\langle xK_i | xK_j \rangle \leq \frac{1}{\sqrt{2}} \text{ si } i \neq j. \quad (8.59c)$$

Cette borne s'obtient car $K_i \cap K_j$ est un sous-groupe à la fois de K_i et de K_j . Son cardinal est donc un diviseur du cardinal de K_i et de K_j . Comme $K_i \neq K_j$, le cardinal de $K_i \cap K_j$ est nécessairement un diviseur strict de $|K_i|$ ou de $|K_j|$, d'où le facteur $1/\sqrt{2}$. L'idée est ensuite très simple : on va répéter la partie de l'algorithme aboutissant à la création de l'état $|xK\rangle$ un certain nombre de fois pour obtenir une famille d'états $(|x_1K\rangle, |x_2K\rangle, \dots)$. On va ensuite tester si tous ces états correspondent au sous-groupe K_1 en regardant s'ils appartiennent au sous-espace généré par K_1 $(|x_1K_1\rangle, |x_2K_1\rangle, \dots)$. Si on mesure que l'un de ces états est en dehors de ce sous-espace, on peut conclure que $K \neq K_1$. Inversement, la probabilité de trouver tous ces états dans le sous-espace généré par K_1 peut être rendue aussi petite que voulue si $K \neq K_1$, puisque pour chaque état la probabilité $|\langle xK_1 | xK \rangle|^2$ est une constante strictement plus petite que 1. On peut alors tester si $K = K_2$ avec la même méthode, et ainsi de suite. Avec $\mathcal{O}((\log |G|)^2)$ états, on peut trouver le bon sous-groupe avec forte probabilité [150].

Un deuxième résultat important est que l'on peut montrer que pour tout groupe abélien, l'oracle peut également être calculé en un temps polynomial en $\log |G|$, et donc qu'il existe un algorithme quantique efficace pour résoudre le problème du sous-groupe caché, comme démontré dans les travaux originaux de Shor [336] et de Kitaev [228].

Cette généralité vient notamment des propriétés de la transformée de Fourier discrète sur les groupes abéliens finis. Pour comprendre cela, nous allons généraliser la transformée de Fourier à n'importe quel groupe G . Considérons une fonction f d'un groupe fixé G dans $\mathbb{C}$ invariante par conjugaison $g \in G \mapsto hgh^{-1}$ ($h \in G$). Nous pouvons la décomposer dans une base orthogonale de fonctions de G dans $\mathbb{C}$: ces fonctions sont les caractères des représentations irréductibles de G . Dans le cas où G est abélien, toutes les fonctions sont invariantes par conjugaison, les représentations irréductibles

sont unidimensionnelles et leurs caractères sont des morphismes de G dans le groupe multiplicatif $\mathbb{C}^*$. L'ensemble des caractères de G constitue un groupe $\hat{G}$ appelé groupe dual de G , le produit de deux caractères correspondant au produit tensoriel des représentations. Si G est un groupe abélien fini, il est isomorphe à son dual et a donc en particulier le même nombre d'éléments. C'est cette propriété qui permet d'obtenir un oracle polynomial. Par exemple, pour $G = \mathbb{Z}/N\mathbb{Z}$, les représentations sont indexées par un entier j modulo N et les fonctions suivantes

$$\chi_j : k \mapsto e^{2i\pi jk/N}, \text{ pour } j \in \mathbb{Z}/N\mathbb{Z}. \quad (8.60)$$

sont les caractères correspondants. Remarquons que la fonction qui à j associe χ_j est bien un isomorphisme du groupe G dans le groupe dual $\hat{G}$ muni de la multiplication usuelle. Ce n'est pas le seul, et une autre numérotation des caractères peut nous conduire à une autre transformée de Fourier discrète.

Retour sur la cryptographie

Comme on l'a expliqué dans la section 7.2.1, les protocoles actuels de cryptographie asymétrique utilisent divers problèmes mathématiques difficiles à résoudre *a priori*, mais dont la vérification d'une solution est simple, pour chiffrer des messages. Dans le cas de RSA, il est ainsi difficile de factoriser un nombre en ses facteurs premiers, mais très simple de vérifier si l'on a ces facteurs que leur produit donne bien le nombre initial. La question posée par l'existence du calcul quantique est donc celle de la protection de tels protocoles asymétriques. En effet, on a vu que l'algorithme de Shor pouvait factoriser un nombre premier en temps polynomial, et donc compromettrait totalement la sécurité du protocole RSA. Même si sa mise en œuvre expérimentale sur des grands nombres tels que ceux utilisés dans RSA semble encore éloignée, il est dès maintenant nécessaire de réfléchir à d'autres solutions de cryptographie qui ne seraient pas compromises par l'existence d'un ordinateur quantique. Malheureusement, comme on vient de le voir dans la section précédente, cela signifie qu'aucun problème basé sur un groupe abélien ne peut prendre la relève, puisque tous ces problèmes sont résolus de manière équivalente. C'est le cas par exemple d'un autre protocole actuel de cryptographie, basé sur l'étude des courbes elliptiques [272, 236], qui a l'intérêt d'utiliser des clés plus courtes que RSA, mais qui repose également sur un groupe abélien. Pour se protéger contre un ordinateur quantique, il est donc nécessaire de trouver un cryptosystème qui se base sur un groupe non abélien. L'étude de certains de ces protocoles date d'avant l'algorithme de Shor [268, 368]. Depuis, le domaine a fortement évolué et d'autres algorithmes sont apparus. Ceux-ci sont notamment basés sur différents problèmes mathématiques, comme la recherche du plus court vecteur sur un réseau [203], ou bien d'autres problèmes plus complexes basés sur les isomorphismes des courbes elliptiques supersingulières [218]. Ces problèmes correspondent à des groupes non abéliens, et sont pensés résistants

TABLE 8.1 – Un résumé des différents types d’accélération des algorithmes quantiques présentés dans ce chapitre. Savoir si un algorithme permet une diminution de la complexité en temps est dépendant du problème exact sur lequel il est considéré. Comme on le voit, tous ces algorithmes donnent lieu à des types d’accélération quantique différents.

Algorithme	Accélération	Prouvée
<i>Avec oracle</i>		
Deutsch-Jozsa ¹	$\exp n$	Oui
Grover	n^2	Oui
Simon	$\exp n$	Oui
<i>Sans oracle</i>		
Amplification d’amplitude ²	n^2	Non
Shor	$\exp n$	Non

¹ Le problème de Deutsch-Jozsa est un problème à promesse, donc le statut de la séparation est un peu particulier.

² On entend ici l’amplification d’amplitude appliquée à certains problèmes comme 3SAT.

à un ordinateur quantique. On parle alors de *cryptosystème post-quantique*. Le NIST a lancé un concours en 2017 afin de désigner un cryptosystème post-quantique standardisé [283].

Cette résistance au quantique pourrait bien entendu changer si un algorithme quantique efficace fonctionnant sur les groupes non abéliens était découvert. À noter qu’un algorithme quantique efficace pour les sous-groupes cachés du groupe des permutations donnerait un algorithme quantique efficace pour le problème de l’isomorphisme de graphes, un problème NP qui n’est cependant pas NP-complet.

Enfin, soulignons que la cryptographie quantique dont nous avons parlé dans la section 7.2.1 n’apporte en rien une solution pour remplacer RSA. En effet, les protocoles de cryptographie quantique sont des protocoles symétriques contrairement à RSA et aux protocoles de cryptographie post-quantique.

8.3 Complexité et théorie quantique

Jusqu’ici, nous avons présenté plusieurs algorithmes permettant « d’aller plus vite » que leurs homologues classiques. Cependant, on se rappelle que cette notion dépend de la manière dont nous définissons cette accélération, ainsi que des quantités que l’on regarde exactement. Avant d’aller plus loin, faisons un bilan des résultats que nous avons obtenus à l’aide du tableau 8.1.

Comme on le voit, il existe à peu près autant de cas différents d'accélération quantique que d'algorithmes que nous avons étudiés. Il paraît alors important de chercher à prendre un peu de recul et à trouver un meilleur moyen de classer les problèmes que nous étudions en fonction de leur capacité à être résolus efficacement ou non par des ordinateurs quantiques. Pour cela, et comme nous l'avons vu dans la section 3.4, la bonne notion à mettre en avant est la notion de complexité. C'est elle qui permet de définir la difficulté d'un calcul, selon le type de ressources qu'il faut se donner pour le résoudre. À partir de cette notion, nous avons pu définir un certain nombre de classes de complexité, dont les classes P, BPP et NP. L'intérêt de ces trois classes est qu'elles donnent l'ensemble des problèmes qui sont faciles pour une machine donnée, respectivement déterministe, probabiliste et non déterministe. Bien qu'il n'y ait pas pour l'heure de preuve de cela, l'opinion majoritaire est que les machines de Turing probabilistes sont équivalentes aux machines de Turing déterministes, et que les machines de Turing non déterministes sont plus puissantes que les deux autres. Par ailleurs, dans l'état actuel de nos connaissances, il ne semble pas possible de construire la machine physique ayant les capacités d'une machine non déterministe.

Néanmoins, les machines que nous avons introduites jusqu'alors n'utilisent pas la mécanique quantique de manière essentielle. Dans le début de ce chapitre, nous avons vu un ensemble de problèmes pour lesquels l'utilisation de ressources quantiques semble simplifier grandement leur résolution. Le but de cette section va être de définir des modèles de calcul pertinents pour le calcul quantique, et de reprendre les résultats que nous avons énoncés sur la complexité. Nous allons donc commencer par introduire un modèle de calcul quantique basé sur la machine de Turing. Cependant, contrairement au cas classique, le modèle de calcul privilégié n'est pas la machine de Turing quantique mais celui des circuits quantiques. Ces derniers sont plus proches de ce qui est accessible expérimentalement, et il s'agit également de la plateforme de choix pour mettre en œuvre la correction d'erreurs, sujet extrêmement important dans le cas quantique, et que nous aborderons par la suite.

8.3.1 Machine de Turing quantique

Le premier modèle que nous allons définir est une machine de Turing probabiliste transposée au monde quantique. Nous allons voir que ce modèle possède néanmoins quelques subtilités par rapport à la machine de Turing classique. Deux questions importantes sont celles de l'unitarité et de la mesure.

Tout d'abord, dans le cas probabiliste, nous avons autorisé la machine de Turing à associer une probabilité à chaque transition possible. Ces probabilités doivent être calculables efficacement : on doit pouvoir en calculer la n -ième décimale en temps polynomial en n . De plus, l'ensemble de ces nombres doit satisfaire les conditions nécessaires à une distribution de probabilité : la somme

des probabilités étiquetant les transitions partant du même état initial et lisant la même lettre sur le ruban doit faire un.

La machine de Turing quantique est la transposition de ce modèle au cas quantique [128, 49]. Dans ce modèle, l'espace des états est l'espace de Hilbert $\mathcal{H}_{\text{TM}} = \mathcal{H}_Q \otimes \mathcal{H}_{\text{tape}}$, où $\mathcal{H}_Q$ est l'espace de Hilbert généré par les états de la machine de Turing et $\mathcal{H}_{\text{tape}}$ est l'espace de Hilbert du ruban. Ce dernier contient à la fois l'information sur le contenu du ruban, ainsi que sur la position de la tête de lecture.

Transposer le modèle probabiliste revient alors à étiqueter les transitions par des nombres complexes dont on peut calculer efficacement la partie réelle et la partie imaginaire. De plus, les transitions sont contraintes par le fait que l'évolution à chaque pas de temps doit être unitaire. La contrainte d'unitarité offre une première difficulté, car l'unitarité impose la réversibilité temporelle. Néanmoins, cela n'est pas une difficulté majeure, car il est tout à fait possible de rendre réversible une machine qui ne l'est pas [38]. Par ailleurs, dans le cas probabiliste, il est possible de se contenter uniquement des transitions ayant une probabilité 0, $1/2$ ou 1. De même, dans les cas qui vont nous intéresser, nous pourrions nous contenter de l'ensemble d'amplitudes $\{0, \pm 3/5, \pm 4/5, \pm 1\}$ [7, 345] ou bien $\{0, \pm 1/\sqrt{2}, \pm 1\}$ [383, 288]. On notera que les nombres complexes ne sont pas nécessaires au calcul quantique, car il suffit de transformer chaque amplitude complexe en deux amplitudes réelles.

La difficulté majeure consiste alors à savoir quand et quoi mesurer. Il est nécessaire de mesurer la totalité du ruban à la fin, afin de récupérer le résultat. Étant donné que la machine ainsi que le ruban sont dans des états quantiques avant la mesure, le résultat de la mesure, qui n'est rien d'autre que le résultat du calcul, est probabiliste. En fait, on peut se contenter de mesurer le ruban une seule fois, à la toute fin du calcul. Mais pour cela, il faut pouvoir déterminer si la machine de Turing quantique est arrêtée. Il est tout à fait possible que certaines branches de la superposition soient arrêtées tandis que d'autres branches continuent leur exécution. Pour résoudre cette difficulté, il est possible d'utiliser une case du ruban qui notifie si la machine a terminé. On peut montrer que ce modèle permet de simuler un modèle de machine de Turing quantique dans lequel on ne mesure pas le ruban [293]. Nous considérerons donc que l'arrêt de la machine se fait à partir d'une case que l'on observe, et l'on obtient le résultat en mesurant l'ensemble du ruban lorsque cette case indique l'arrêt.

8.3.2 La classe BQP

Tout cela définit donc le modèle de la machine de Turing quantique. Nous pouvons alors définir la classe BQP, qui est l'analogue quantique de BPP.

Définition 8.1 (BQP). Un problème appartient à la classe BQP (pour *bounded-error quantum polynomial*) s'il existe une machine de Turing quantique dont le temps d'exécution est polynomial en la taille de l'entrée et qui accepte

avec une probabilité strictement supérieure à $2/3$ si la réponse est « oui » et avec une probabilité strictement inférieure à $1/3$ si la réponse est « non »¹¹.

Se pose alors la question de savoir où se situe BQP par rapport aux autres classes de complexité que nous avons vues. Tout d'abord, on peut se servir d'une machine de Turing quantique pour générer des nombres aléatoires. De plus, il est possible de simuler efficacement une machine de Turing déterministe avec une machine de Turing quantique. Cela signifie qu'un problème que l'on peut résoudre efficacement à l'aide d'une machine de Turing probabiliste peut aussi se résoudre efficacement à l'aide d'une machine de Turing quantique. On a donc

$$\text{BPP} \subseteq \text{BQP}. \quad (8.61)$$

Peut-on trouver une classe de complexité dans laquelle BQP est incluse ? Nous allons prouver que BQP est une sous-classe de PP, que nous avons définie dans la section 3.4.4. Pour cela, considérons un problème L dans BQP et la machine de Turing quantique associée M_Q qui décide si une entrée x de taille n est dans L . Nous considérerons que les amplitudes de transition de la machine de Turing M_Q sont dans l'ensemble $\{0, \pm 1/\sqrt{2}, \pm 1\}$.

Nous allons simuler classiquement un calcul de la machine de Turing quantique. À chaque fois que nous rencontrons une transition menant à une superposition (donc d'amplitude $\pm 1/\sqrt{2}$), nous tirons la transition au hasard. Nous effectuons donc une simulation de la machine de Turing quantique « branche par branche », de façon probabiliste. Sur chacune des exécutions probabilistes, nous notons l'amplitude $(-1)^z/\sqrt{2}^m$ où $z \in \{0, 1\}$ indique le signe de l'amplitude et $m \in \mathbb{N}$ compte le nombre de transitions de type $\pm 1/\sqrt{2}$. À chaque exécution, on note également l'état final de la machine y , qui peut être acceptant ou rejetant.

L'algorithme dans PP qui permet de résoudre un problème dans BQP est le suivant. Sur l'entrée x , nous effectuons deux simulations classiques, qui nous donnent les triplets (m, z, y) et (m', z', y') . Dans le cas où les deux exécutions donnent des résultats différents ($y \neq y'$), nous tirons le résultat au hasard avec une probabilité $1/2$. Le cas intéressant est donc lorsque les deux exécutions donnent le même résultat ($y = y'$). Nous allons d'abord considérer le sous-cas où les probabilités associées à chacune des exécutions sont les mêmes ($m = m'$). Dans ce cas, on regarde le signe des amplitudes de probabilité. Si les signes sont les mêmes ($z = z'$), on répond y . Sinon, on répond la négation de y . L'idée est la suivante : le cas $z = z'$ correspond à des interférences positives, ce qui augmente la probabilité du résultat alors que le cas $z \neq z'$ correspond à des interférences négatives et c'est pourquoi nous choisissons le résultat inverse. Si on a $m \neq m'$, on peut supposer $m > m'$. On utilise alors la même méthode

11. Tout comme pour la classe BPP, la probabilité $2/3$ est arbitraire. N'importe quelle probabilité strictement supérieure à $1/2$ convient.

que dans le cas $m = m'$ avec probabilité $1/2^{m-m'}$ et on répond au hasard le reste du temps.

Pour cet algorithme, la différence entre la probabilité d'accepter et la probabilité de refuser dépend de la taille de l'entrée. Dans le cas où les branches d'exécution ont toutes la même probabilité (et donc toutes le même m), cette différence est égale à 2^{-m} fois la différence entre la probabilité d'accepter et la probabilité de refuser de la machine de Turing quantique, ce qui explique pourquoi la probabilité d'erreur n'est pas bornée dans notre algorithme probabiliste. Le lecteur intéressé par le détail des calculs pourra se référer à l'article original [7]. On a donc

$$\text{BQP} \subseteq \text{PP}. \quad (8.62)$$

Nous sommes donc capables d'encadrer la classe BQP entre BPP et PP. Cependant la question la plus intéressante si l'on cherche à construire un ordinateur quantique, est de savoir s'il existe des problèmes pour lesquels l'ordinateur quantique est plus efficace que l'ordinateur classique. En termes de théorie de la complexité, cela revient à savoir s'il existe des problèmes dans BQP qui ne sont pas dans BPP. Comme il est habituel en théorie de la complexité, il n'y a aucun résultat démontré sur ce sujet. En revanche, nous pouvons nous servir des exemples d'algorithmes que nous avons vus pour développer une intuition. En ce début de chapitre, nous avons vu des exemples d'algorithmes quantiques dont les performances surpassent tous les algorithmes classiques connus.

On peut catégoriser ces algorithmes en deux familles. Nous avons d'abord vu des algorithmes qui utilisent une boîte noire, comme les algorithmes de Deutsch-Jozsa ou de Simon. Dans ces cas-là, on a une borne inférieure démontrée sur la complexité asymptotique en requêtes qui est meilleure dans le cas quantique que dans le cas classique. Notamment, l'algorithme de Simon semble particulièrement intéressant, car il produit une séparation exponentielle entre le cas quantique et le cas classique. Naïvement, on pourrait penser que cela implique que BQP est strictement plus grand que BPP. En fait, on touche ici du doigt les limitations de la complexité en requêtes. Si l'on souhaite transformer la complexité en requêtes en complexité en temps, il est nécessaire de spécifier la boîte noire et de la remplacer par le calcul d'une fonction d'une certaine complexité, par exemple une fonction que l'on peut calculer en temps polynomial. Or, les fonctions d'une classe de complexité donnée possèdent une structure, que l'on s'interdit d'exploiter lorsque l'on traite ces fonctions comme des boîtes noires. Il n'est alors pas dit que la complexité optimale en requêtes donne la complexité optimale en temps.

D'un autre côté, nous avons vu certains algorithmes quantiques n'utilisant pas de boîte noire. Un exemple particulièrement impressionnant est l'algorithme de Shor qui permet de résoudre le problème de la factorisation en temps polynomial alors que les meilleures méthodes connues classiquement sont en

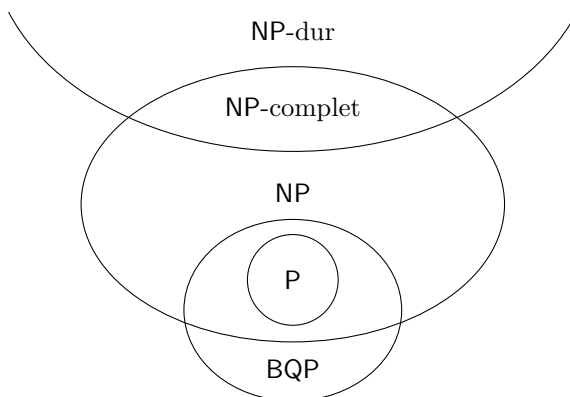


FIGURE 8.12 – Les relations supposées entre la classe BQP et les classes P, NP et NP-complet.

temps exponentiel. Cependant si la factorisation est un problème NP, on ne sait pas si c'est un problème NP-complet.

Actuellement, la classe BQP est pensée strictement plus grande que BPP, ce qui signifie que l'on pourrait tirer parti de la puissance d'un ordinateur quantique sur certains problèmes. Cependant, il est fort probable que les problèmes NP-complets soient inaccessibles à un ordinateur quantique. De même, il n'est pas impossible que certains problèmes qui ne sont pas dans la classe NP appartiennent à BQP. Cela vient du fait que la machine de Turing non déterministe trouve, par définition, un état acceptant. C'est cette propriété qui est responsable des subtilités de la classe NP vues dans la section 3.4.2. Notamment, une machine déterministe appelant une machine non déterministe n'a, *a priori*, pas la même puissance de calcul. En revanche, les amplitudes de probabilité de la machine de Turing quantique, elles, n'ont pas cette contrainte de converger nécessairement vers un état acceptant lorsqu'il est disponible. C'est pour cela qu'il n'y a pas de relation d'inclusion connue de BQP dans NP. L'ensemble des relations supposées entre ces classes est résumé par la figure 8.12.

8.3.3 Thèse de Church-Turing étendue quantique

Nous pouvons à présent reprendre la discussion introduite à la section 3.4.5. En effet, nous avons posé la question de savoir quels étaient les problèmes accessibles dans le monde physique. En d'autres termes, quelle est la classe de problèmes pour lesquels il existe une mise en œuvre physique réaliste capable de résoudre les problèmes de cette classe en temps polynomial.

En utilisant des ressources classiques, il semble que la plus grande classe de complexité accessible soit la classe BPP. Il est en effet possible d'utiliser de l'aléatoire avec un ordinateur classique, ce qui permet d'exécuter les problèmes

de cette classe. En revanche, l'utilisation de ressources quantiques remettrait en cause cette version de la thèse de Church-Turing étendue. Un ordinateur quantique donnerait immédiatement accès aux problèmes de la classe BQP. On aurait alors la formulation suivante, pour la thèse de Church-Turing étendue quantique :

Une machine de Turing quantique peut simuler efficacement n'importe quel modèle de calcul réaliste.

Savoir laquelle de ces deux thèses privilégier dépend directement de la réalisation d'un ordinateur quantique à grand nombre de qubits. Cette dernière rendrait effectivement fausse la thèse de Church-Turing étendue probabiliste, au profit de sa version quantique.

8.3.4 Les problèmes dans BQP ont-ils un intérêt ?

La construction d'un ordinateur quantique n'a d'intérêt que dans le cas où l'ordinateur quantique est bien plus rapide que l'ordinateur classique. Nous avons vu dans la section 8.1 que l'amplification d'amplitude permet une accélération quadratique assez générique, qui touche donc un certain nombre de problèmes. D'autres techniques, basées sur des marches aléatoires quantiques, permettent le même type d'accélération. Si, sur des problèmes compliqués, cela permet d'augmenter la taille des instances d'un facteur 2, ce gain n'est intéressant qu'au cas par cas. Il faut à la fois des instances des problèmes trop grosses pour un ordinateur classique, mais pas trop grosses pour un ordinateur quantique.

Le véritable intérêt d'un ordinateur quantique réside dans les problèmes qui sont dans BQP mais que l'on ne pense pas être dans BPP. Pour ces problèmes, avoir un ordinateur quantique permettrait de traiter des tailles d'instance radicalement différentes. Trouver un problème intéressant pourrait bouleverser certains domaines techniques et scientifiques. Toute la question est alors de savoir s'il existe un problème intéressant possédant ce potentiel.

En section 8.2, nous avons vu l'algorithme de Shor qui permet de factoriser un nombre en ses facteurs premiers, en temps polynomial. Une variation de l'algorithme de Shor permet également de résoudre le problème du logarithme discret en temps polynomial. La difficulté de ces problèmes est massivement utilisée dans la cryptographie asymétrique actuelle, et changerait donc complètement le paysage algorithmique. Hormis cela, l'impact serait relativement mineur, et devrait de toute façon passer relativement inaperçu, en dehors des communautés techniques.

Il existe cependant d'autres algorithmes quantiques que nous n'avons pas abordés dans ce chapitre. L'un d'entre eux est l'échantillonnage de bosons [3] (*boson sampling*). Celui-ci consiste à échantillonner la distribution des photons qui sortent d'un interféromètre linéaire à N voies. L'intérêt de ce problème est pour l'heure, purement académique : il s'agit de résoudre un problème pensé difficile en classique, à l'aide uniquement de sources de photons uniques

et de composants optiques linéaires. En dehors de ce point remarquable, ce problème ne se rattache à aucun autre problème pratique. Plus récemment Harrow, Hassidim et Lloyd (HHL) [197] ont montré que l'on pouvait utiliser un ordinateur quantique pour évaluer rapidement certaines quantités statistiques de certains systèmes d'équations linéaires. Il s'agit là plutôt d'une technique que d'un algorithme en soi [1] : le système d'équation doit pouvoir être généré efficacement algorithmiquement. Cela implique notamment que le système d'équation est creux. De plus, celui-ci doit être bien conditionné, ce qui signifie que l'écart entre les valeurs propres de la matrice qui définit le système d'équation doit être suffisamment petit. Pour transformer cela en algorithme utilisable, il faut donc trouver un problème qui satisfait l'ensemble de ces contraintes. L'une des applications pourrait notamment être dans le domaine de la simulation en éléments finis de systèmes d'équations aux dérivées partielles. Dans ce domaine, on peut parfois manipuler le système linéaire obtenu de façon à ce que celui-ci satisfasse l'ensemble de ces critères. Cela a notamment été montré pour le problème de la diffusion des ondes électromagnétiques sur un objet [97]. Cependant, si l'on souhaite garder un contrôle sur l'erreur numérique finale de la quantité évaluée (la section efficace, par exemple), le gain apporté n'est plus exponentiel mais polynomial [276], et est d'autant plus élevé que la dimension du système d'équations différentielles est élevée. Dans le cas d'un système de haute dimension pour lequel les autres techniques (par exemple Monte-Carlo) fonctionnent mal, il peut donc y avoir un gros intérêt à utiliser un algorithme quantique.

Enfin, l'application la plus évidente d'un ordinateur quantique serait la simulation de systèmes quantiques. Cet usage a été proposé dès 1982 par Feynman [158]. Actuellement on ne connaît pas d'algorithme permettant la simulation efficace de systèmes quantiques possédant un grand nombre de sous-systèmes, la taille de l'espace de Hilbert croissant exponentiellement avec le nombre de sous-systèmes. Il est ainsi considéré comme impossible de simuler un ordinateur quantique ayant plus d'une cinquantaine de qubits sur un ordinateur actuel. Utiliser des ressources quantiques pour simuler d'autres systèmes quantiques permettrait donc d'aller vers des systèmes autrement plus grands, avec des applications allant de l'optimisation de processus chimiques à l'exploration numérique des supraconducteurs haute température. En ce sens, le plus grand impact d'un ordinateur quantique serait les perspectives de simulation de tels processus. Nous reviendrons sur les questions de simulation à la section 10.2.7 car dans certains cas, il n'est pas besoin d'avoir un ordinateur quantique au sens où nous l'entendons dans ce chapitre. Il est possible de se contenter de systèmes moins bien contrôlés.

8.4 Circuits quantiques

Si la machine de Turing quantique est un modèle élégant de calcul, car celui-ci étend naturellement le modèle de la machine de Turing au monde

quantique, il n'est cependant pas le plus utilisé à l'heure actuelle. D'une part, les architectures quantiques qui sont développées actuellement en laboratoire sont extrêmement éloignées d'une machine de Turing quantique. Elles sont plus proches d'un circuit à faible nombre de qubits et de portes que d'un ordinateur avec des mémoires de masse. De la même manière, le modèle de calcul que nous avons introduit à la section 7.2.2 et utilisé pour présenter les algorithmes quantiques était un modèle de circuit. Un point également important est que la correction d'erreurs est beaucoup plus facile à mettre en œuvre sur un circuit que sur une machine de Turing. Enfin, d'un point de vue théorique, les circuits offrent de nouvelles techniques pour démontrer des séparations entre différentes classes de complexité, et sont donc intéressants à étudier en soi.

Nous allons maintenant détailler ces modèles, en faisant un retour sur les circuits classiques, afin d'illustrer les difficultés conceptuelles des approches basées sur les circuits.

8.4.1 Circuits classiques et complexité

Nous avons vu au chapitre 3 qu'il était possible de trouver un circuit, constitué de portes logiques classiques, à n entrées et une sortie pour toutes les fonctions de $\{0, 1\}^n$ dans $\{0, 1\}$. Afin d'avoir un modèle complet qui agisse sur l'ensemble des entrées dans $\{0, 1\}^*$, il est donc nécessaire de pouvoir associer un circuit C_n pour chaque taille d'entrée n . On appelle cela un modèle non uniforme, pour lequel chaque taille d'entrée correspond à un circuit différent.

Une première subtilité apparaît alors. Étant donné que toute fonction de $\{0, 1\}^n \rightarrow \{0, 1\}$ est calculable par un circuit, le modèle de circuit est, si l'on ne rajoute pas de restriction, infiniment plus puissant que les machines de Turing. Notamment, il existe un circuit capable de déterminer si un programme de taille n s'arrête sur l'ensemble de ses entrées. Le modèle de circuit permet donc, si l'on n'impose pas de restriction, de calculer le problème de l'arrêt. Cette puissance vient du fait que la seule chose qui compte dans ce modèle est l'existence d'un circuit permettant de faire le calcul. Notamment, cela n'implique pas qu'il existe une procédure permettant, pour un entier n donné, de construire le circuit C_n associé en pratique. Afin d'obtenir un modèle de calcul raisonnable, il est donc nécessaire d'y adjoindre la procédure qui calcule chacun des circuits. Habituellement, cela consiste à faire calculer la famille de circuits C_n par une machine de Turing.

On peut alors se poser les questions de complexité en termes de circuits. Un premier résultat est que l'on peut associer à une machine de Turing déterministe de complexité $\mathcal{O}(t(n))$ une famille de circuits avec $\mathcal{O}(t(n)^2)$ portes. Cela démontre que les modèles de circuits sont au moins aussi puissants que les machines de Turing. En revanche, on s'attend à ce que, sans les restrictions dont nous avons parlé au paragraphe précédent, les circuits soient plus puissants. Afin d'obtenir un modèle raisonnable, nous allons donc imposer le fait que

l'on puisse calculer en temps polynomial la famille de circuit (C_n) . Il existe une machine de Turing qui sur l'entrée 1^n calcule le circuit C_n en temps polynomial. On appelle cette condition la P-uniformité. On peut alors montrer qu'un problème est dans P si et seulement s'il existe une famille de circuit P-uniforme résolvant ce problème. Étant donné que le circuit est calculé en temps polynomial, sa taille est elle-même polynomiale.

Les modèles de circuits raisonnables faisant appel à une machine de Turing, leur intérêt peut sembler limité. Il y a, en fait, plusieurs intérêts à ces modèles. Tout d'abord, on peut se demander comment transcrire la notion de parallélisme dans les classes de complexité. Une machine de Turing étant uniquement séquentielle, si l'on veut pouvoir classer les problèmes entre ceux qui se parallélisent bien sur une machine massivement parallèle et les autres, il faut un modèle différent. Pour cela on peut s'intéresser, en plus de la taille des circuits, à leur profondeur. Si on a un circuit dont la profondeur est logarithmique en la taille de l'entrée et dont le nombre de portes est polynomial, alors la parallélisation permet d'obtenir des gains exponentiels sur le temps de calcul, si l'on n'est pas limité par le nombre de processeurs. Chaque porte peut être vue comme un processeur indépendant, et les connexions entre les portes sont des échanges de messages entre les processeurs. Étant donné que les processeurs peuvent tourner simultanément, le temps de calcul est dans ce cas donné par la profondeur du circuit et non plus sa taille [254]. Un autre intérêt des circuits est que cela donne des pistes pour montrer des bornes sur les classes de complexité [319, 310, 2].

8.4.2 Circuits quantiques et complexité

Nous allons maintenant nous intéresser aux propriétés des circuits quantiques. Cela permettra de formaliser de façon plus rigoureuse la complexité en temps sur les algorithmes que nous avons vus dans la section 8.2. Tout comme dans le cas classique, nous allons définir une complexité sur la taille du circuit quantique en fonction de la taille de l'entrée. Afin de définir des classes de complexité, nous allons également imposer une condition d'uniformité.

La taille d'un circuit étant définie par le nombre de portes, tout comme dans le cas classique, il est nécessaire de définir une notion de porte quantique utile pour la complexité. Dans la suite, nous allons considérer qu'une opération est une porte quantique si elle agit sur trois qubits ou moins. Contrairement au cas classique, le nombre de portes à trois qubits ou moins est infini. Cela pose deux difficultés. Dans le cas classique, nous avons pu réécrire n'importe quelle opération booléenne comme un assemblage fini de trois portes élémentaires (**not**, **or** et **and**). Il serait pratique, dans le cas quantique, d'avoir accès à un ensemble fini de portes qui permette d'approximer n'importe quelle opération quantique sur n qubits, avec une précision arbitraire.

Nous pouvons donc séparer la tâche en deux parties. Tout d'abord, nous allons décrire n'importe quelle opération unitaire par un arrangement de portes

à deux qubits. Nous verrons ensuite comment approximer n'importe quelle porte à deux qubits. Ces deux étapes constitueront donc l'analogie quantique de la construction que nous avons développée au chapitre 3. Le lecteur intéressé par ces questions pourra se référer à [229, 231, 287].

Décomposer une matrice unitaire en portes

Pour décomposer une matrice unitaire en portes quantiques, nous allons devoir décomposer cette matrice en un ensemble d'opérations plus simples. Pour cela, nous allons utiliser d'une part la vision que nous avons vue précédemment, en termes d'opérations logiques agissant sur n qubits, et d'autre part la vision opératorielle habituelle de la mécanique quantique qui agit sur l'espace de Hilbert de dimension 2^n . La raison de cela est que pour exprimer un opérateur sur $n + 1$ qubits en termes d'opérateurs à n qubits, il va d'abord nous falloir une façon de couper en deux l'espace de Hilbert de dimension 2^{n+1} en deux espaces de Hilbert de dimension 2^n .

L'état quantique à n qubits peut être vu comme une mémoire à 2^n cases, où chaque case adressée par $j \in \{0, \dots, 2^n - 1\}$ contient l'amplitude associée à $|j\rangle$. Nous appellerons « qubit de poids fort » le qubit correspondant au premier chiffre dans l'écriture binaire de l'adresse mémoire représenté par les n qubits, celui qui lorsqu'il est changé, entraîne les plus fortes variations sur les adresses. Réciproquement, les qubits de poids faibles sont les derniers chiffres dans l'écriture binaire de l'adresse. Si l'état quantique est vu comme une mémoire, les opérateurs unitaires changent le contenu de la mémoire en effectuant des rotations sur les amplitudes. Que se passe-t-il si on a un opérateur unitaire qui n'agit que sur la première ou la seconde moitié des amplitudes ? Chacune de ces opérations agit sur les $n - 1$ qubits de poids faible, tout en étant contrôlée par le qubit de poids fort (voir figure 8.13).

Deux questions naturelles se posent alors. D'une part, est-il possible d'écrire n'importe quel opérateur comme un produit d'opérateurs agissant sur des sous-espaces distincts de l'espace de Hilbert total ? En appliquant cela de manière récursive, cela permettrait d'écrire l'opérateur initial comme une suite d'opérations agissant sur un seul qubit de manière conditionnée à l'état de l'ensemble des autres qubits. À partir d'une telle description, pouvons-nous réécrire chacune des opérations comme une suite de portes agissant sur un nombre limité de qubits ? En répondant à ces deux questions nous obtiendrons ainsi une procédure permettant de passer de n'importe quel opérateur agissant sur n qubits, à un produit d'opérateurs agissant sur un nombre fixe de qubits.

Pour répondre à la première question, nous allons chercher à couper l'espace de Hilbert en deux, selon le qubit de poids fort. L'espace de Hilbert total peut donc s'écrire $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, où $\mathcal{H}_0$ est le sous-espace de l'espace de Hilbert dans lequel le qubit de poids fort vaut 0 et $\mathcal{H}_1$ est le sous-espace dans lequel le qubit de poids fort vaut 1. Il s'agit alors de décomposer n'importe quel opérateur unitaire en un produit d'opérateurs plus simples dans ces sous-espaces. Pour cela, nous introduisons la décomposition suivante, dans la base $(\mathcal{H}_0, \mathcal{H}_1)$ [315,

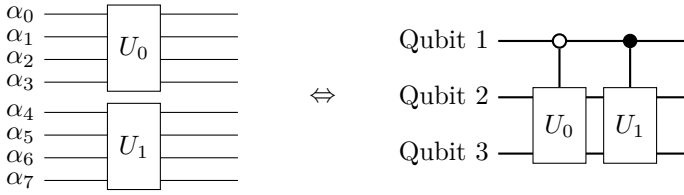


FIGURE 8.13 – Action des opérations contrôlées sur l'espace de Hilbert. À gauche : les opérateurs U_0 et U_1 agissent sur les sous-espaces donnés par $\text{Vect}(|0\rangle, |1\rangle, |2\rangle, |3\rangle)$ et $\text{Vect}(|4\rangle, |5\rangle, |6\rangle, |7\rangle)$ respectivement. Ils transforment indépendamment les amplitudes correspondantes. À droite : pour obtenir la même chose en termes d'opérations agissant sur des qubits, les opérateurs U_0 et U_1 doivent être contrôlés par le qubit de poids fort (qubit 1, à droite). L'opérateur U_0 agit sur les qubits de poids faibles (qubits 2 et 3) seulement si le qubit 1 vaut 0 (ce que l'on note avec un contrôle dont l'intérieur est blanc), et U_1 seulement si le qubit 1 vaut 1.

Annexe C] :

$$U = \left(\begin{array}{c|c} V_0 & 0 \\ \hline 0 & V_1 \end{array} \right) W \left(\begin{array}{c|c} U_0 & 0 \\ \hline 0 & U_1 \end{array} \right) \text{ avec } W = \left(\begin{array}{c|c} u & -v \\ \hline v & u \end{array} \right). \quad (8.63)$$

Les opérateurs U_0 et V_0 sont des opérateurs unitaires qui n'agissent que sur le sous-espace $\mathcal{H}_0$, de même que les opérateurs U_1 et V_1 n'agissent que sur le sous-espace $\mathcal{H}_1$. L'opérateur W mélange les sous-espaces $\mathcal{H}_0$ et $\mathcal{H}_1$. Cependant, il existe un choix des opérateurs (U_0, U_1) et (V_0, V_1) pour lequel l'expression de W est particulièrement simple et s'exprime en fonction de matrices u et v qui sont diagonales réelles.

Si l'on revient à notre vision en termes d'amplitudes, les opérateurs U_0 et V_0 agissent sur les amplitudes qui concernent la première moitié des adresses, allant de 0 à $2^{n-1} - 1$. Les opérateurs U_1 et V_1 agissent sur la seconde moitié des adresses, allant de 2^{n-1} à $2^n - 1$. Il s'agit donc de réorganisations internes aux sous-espaces $\mathcal{H}_0$ et $\mathcal{H}_1$. Entre ces deux réorganisations se trouve l'opérateur W qui mélange les deux sous-espaces. Cependant, comme les matrices u et v sont diagonales, cet opérateur met en relation les couples d'amplitudes se situant aux adresses i et $2^{n-1} + i$ pour tout $i \in \{0, \dots, 2^{n-1} - 1\}$, et ce de manière indépendante. Cela est résumé par la figure 8.14.

Si l'on part du point de vue « qubit », les transformations U_0 , U_1 , V_0 et V_1 sont des opérations agissant sur les qubits de poids faibles, contrôlées par le qubit de poids fort, qui indique le sous-espace sur lequel chacun de ces opérateurs agit. Si l'on reprend la terminologie introduite plus haut, les opérateurs U_0 et V_0 agissent sur l'ensemble des adresses qui commencent par 0, et donc n'agissent que lorsque le qubit de poids fort est à 0. Il en va de même pour U_1 et V_1 qui n'agissent que lorsque le qubit de poids fort est à 1. Entre les couples d'opérateurs (U_0, U_1) et (V_0, V_1) , on a un ensemble d'opérations

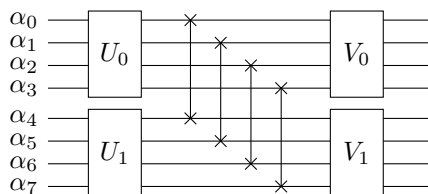


FIGURE 8.14 – Décomposition d’une matrice unitaires en opérateurs agissant sur des sous-espaces. Ici, on considère le cas à 3 qubits, et on sépare l’espace de Hilbert en deux sous-espaces, $\mathcal{H}_0$ associé à $(|0\rangle, |1\rangle, |2\rangle, |3\rangle)$ et $\mathcal{H}_1$ associé à $(|4\rangle, |5\rangle, |6\rangle, |7\rangle)$. On peut ainsi décomposer n’importe quelle transformation unitaire comme deux opérateurs U_0, U_1 agissant indépendamment sur chacun de ces sous-espaces, suivi d’un ensemble de transformations mettant en correspondance les amplitudes deux à deux. Enfin, on change de nouveau de base à l’aide des opérateurs V_0 et V_1 . Le point de vue en termes de circuit est illustré par la figure 8.15.

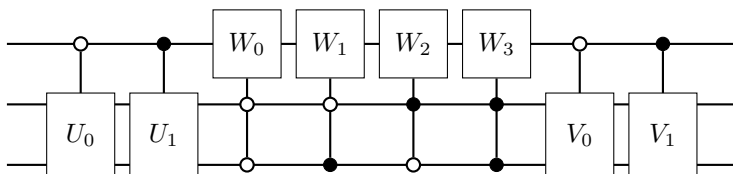


FIGURE 8.15 – Décomposition d’une matrice unitaire en opérations contrôlées. Une matrice unitaire sur n qubits (ici 3) peut être écrite comme quatre opérations, U_0, U_1, V_0, V_1 , à $n - 1$ qubits contrôlées par un qubit, ainsi que 2^{n-1} opérations à un qubit contrôlées par $n - 1$ qubits. Cela correspond au développement illustré par la figure 8.14.

qui mélangent des amplitudes partageant les mêmes qubits de poids faibles mais dont le qubit de poids fort diffère. Étant donné qu’il y a $n - 1$ qubits de poids faibles, il y a donc 2^{n-1} opérations. Ce point de vue est résumé sur la figure 8.15.

Afin d’obtenir la décomposition en termes d’opérations à un seul qubit contrôlées par tous les autres, il est possible de réappliquer cette décomposition, récursivement, aux opérations U_0, U_1, V_0 et V_1 . À chaque itération, le nombre de qubits de contrôle pour chaque opération augmente d’au moins un, jusqu’à arriver à n’avoir que des opérations qui agissent sur un seul qubit conditionnellement à l’ensemble des autres. Étant donné qu’à chaque étage n on a quatre opérations contrôlées portant sur $n - 1$ qubits, ainsi que 2^{n-1} opérations contrôlées à un qubit, le nombre total d’opérations contrôlées est donné par la relation de récurrence $G_{n+1} = 4G_n + 2^n$. On a donc

$$G_n = \frac{2^n(2^n - 1)}{2} \quad (8.64)$$

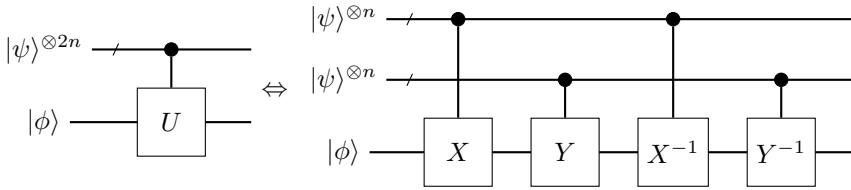


FIGURE 8.16 – Décomposition d’une opération sur un qubit contrôlée par $2n$ qubits. L’opération U est appliquée si et seulement si les $2n$ premiers qubits sont à 1. En utilisant une décomposition de la forme $U = XYX^{-1}Y^{-1}$, il est possible de transformer cette opération en quatre opérations contrôlées par n qubits. En appliquant ce procédé récursivement, on divise par deux le nombre de qubits de contrôle à chaque itération.

opérations à un qubit contrôlées par $n - 1$ qubits nécessaires pour décrire un opérateur unitaire agissant sur n qubits.

Il nous faut maintenant une méthode pour passer des opérations contrôlées par plusieurs qubits à un ensemble d’opérations contrôlées par un seul qubit. Prenons un cas à trois qubits. Imaginons que l’opération unitaire (que l’on considérera tout d’abord de déterminant unité) soit appliquée au troisième qubit si et seulement si les deux premiers qubits valent 1. Étant donné que U est unitaire de déterminant 1, il est possible de décomposer U en le produit $XYX^{-1}Y^{-1}$, où X et Y sont des opérations unitaires. C’est ce que l’on appelle la décomposition en commutateur de groupe [231]. On applique alors successivement X , Y , X^{-1} et Y^{-1} , où X et X^{-1} sont conditionnées au premier qubit, Y et Y^{-1} sont conditionnées au second qubit. Ainsi, si les deux premiers qubits sont à 1, on applique U . Si un seul des deux qubits est nul, alors on applique l’une des transformations et ensuite son inverse. Enfin, si les deux sont nuls, on n’applique rien. On passe donc d’une porte contrôlée par deux qubits à quatre opérations contrôlées par un seul qubit.

Il est possible d’appliquer de nouveau ce procédé récursivement, comme on peut le voir sur la figure 8.16. Ainsi, si l’on a une opération contrôlée par $2n$ qubits, il est possible de transformer cela en quatre opérations contrôlées par n qubits. Il faut donc de l’ordre de n^2 portes à deux qubits pour exprimer une opération à un qubit possédant n qubits de contrôle.

Si U n’est pas de déterminant 1, il est toujours possible d’écrire U comme le produit d’une phase et d’une matrice de déterminant 1. Nous pouvons donc décomposer la matrice de déterminant 1 à l’aide de la décomposition en commutateurs de groupe. D’un autre côté, appliquer la phase φ de façon conditionnée à $n - 1$ qubit revient à appliquer l’opération

$$R_\varphi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{pmatrix} \quad (8.65)$$

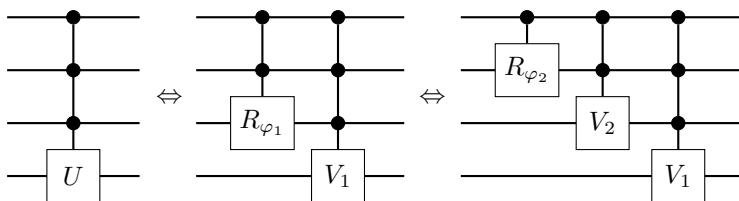


FIGURE 8.17 – Décomposition d’une opération contrôlée de déterminant différent de 1. Si l’opération U n’est pas de déterminant 1, on peut la décomposer comme $e^{i\varphi_1} V_1$, où V_1 est une opération unitaire de déterminant 1. On peut alors exprimer la phase φ_1 comme l’opération R_{φ_1} , conditionnée aux autres qubits. En réitérant ce procédé, on arrive à n’avoir plus qu’une seule opération de déterminant différent de 1, ici le déphasage R_{φ_2} contrôlé par le premier qubit. Les opérations V_1 et V_2 sont de déterminant 1, et on peut donc utiliser la décomposition en commutateurs de groupe.

de manière conditionnée au $n - 2$ premiers qubits. On peut réitérer ce procédé, de façon à réduire le nombre de qubits de contrôle sur la partie phase, jusqu’à n’avoir plus que des portes à deux qubits. Cela est illustré sur la figure 8.17.

À l’aide de ces procédés nous sommes donc capables de passer de n’importe quelle opération unitaire à une série de portes à deux qubits. Les circuits produits ont de l’ordre de $\mathcal{O}(n^3 \times 4^n)$ portes, et sont donc extrêmement longs. Cela n’est pas surprenant, car il s’agit là d’une méthode générique pour produire des circuits. Toute la difficulté des algorithmes quantiques est donc de trouver une famille de circuits qui sont à la fois courts et rapidement calculables.

Un nombre fini de portes

L’un des problèmes de notre construction est qu’elle nécessite une famille de portes infinie (l’ensemble des portes à un qubit contrôlées par un autre). De manière analogue au calcul classique, on aimerait bien avoir un ensemble fini et universel de portes, qui puisse générer n’importe quelle opération. Il est évidemment impossible de générer n’importe quelle opération unitaire de façon exacte, mais est-il possible d’obtenir une approximation aussi bonne que l’on veut ? Et si oui, à quelle vitesse approche-t-on de l’opérateur cible ? Enfin, on peut se demander s’il y a une raison de privilégier un ensemble de portes ou un autre.

Universalité La première question revient à se demander si l’on est capable de générer, à partir d’un ensemble de portes fini, un ensemble dense dans $SU(n)$. Étant donné que l’on a montré que l’on pouvait générer l’ensemble des portes dans $SU(n)$ à partir de portes à un qubit, contrôlées par un second

qubit, on peut considérablement simplifier le problème et se concentrer sur la génération de portes à un seul qubit. Le groupe $SU(2)$ est un groupe de rotation, et afin de générer l'ensemble des rotations il suffit d'en trouver deux ayant des axes différents et dont les angles sont incommensurables à 2π . Il est ainsi assez intuitif de voir que chaque rotation génère un ensemble dense dans le cercle. Les cercles étant distincts, on peut ainsi montrer que l'on obtient l'ensemble des rotations. Il faut donc trouver deux opérateurs générés par les portes de base, qui ne commutent pas, et dont les valeurs propres sont des phases incommensurables avec 2π . Lorsque l'on s'intéresse aux portes à deux ou trois qubits, le plus facile est de partir d'un sous-espace à deux dimensions, dans lequel on montre l'universalité, et d'itérer à partir de là. L'idée est alors d'étendre cet espace, en trouvant des transformations qui permettent de peupler davantage de dimensions. Pour les détails formels, le lecteur pourra se référer à [231, 287].

Il est possible d'utiliser cet argument pour montrer que l'on peut reconstruire l'ensemble des portes à un qubit à partir de la porte d'Hadamard que nous avons vue précédemment, ainsi que de la porte $R_{\pi/4}$, nommée porte $\pi/8$. Ainsi, si l'on a la version contrôlée de ces deux portes, on peut générer n'importe quelle porte à n qubits.

Comme dans le cas classique, il existe plusieurs ensembles universels de portes. Dans le cas classique, ce qui a amené à décider l'un plutôt que l'autre des ensembles était l'architecture CMOS, dans laquelle il est naturel de construire les portes **not** et **nand**. On peut se demander si c'est également sur cette base que l'on doit choisir les portes quantiques. L'une des difficultés spécifiques du cas quantique, que nous aborderons brièvement dans la section 8.4.3, vient de la correction d'erreurs. Contrairement au cas classique, dans laquelle la correction d'erreurs est déjà présente au niveau matériel par la redondance de l'information classique (chaque case mémoire a un état déterminé par un grand nombre d'électrons, par exemple), dans le cas quantique le théorème de non-clonage nous empêche d'avoir une telle approche. Dans l'idéal, on aimerait avoir une approche dans laquelle chaque porte peut être mise en œuvre avec correction d'erreurs, ce qui nécessite également que les erreurs effectuées sur chacune des portes soient faibles (qu'il s'agisse de bruit dû à l'environnement ou d'erreur systématique à cause de la conception des portes).

En gardant en tête ces considérations, nous allons discuter de deux autres ensembles universels de portes d'intérêt. Le premier est la base dite standard [287]. L'immense intérêt de celle-ci est qu'il est relativement aisé de mettre en œuvre ces portes avec certains codes correcteurs d'erreurs. Celle-ci se compose des portes d'Hadamard, **cnot**, de la porte de phase $R_{\pi/2}$, ainsi que de la porte $\pi/8$. Le second ensemble de portes consiste en la porte d'Hadamard, la porte de phase et la porte de Toffoli. L'intérêt de cet ensemble est qu'il permet de marquer la différence entre calcul classique et calcul quantique. La porte de Toffoli est une porte classique universelle. On voit que pour reproduire n'importe quel opérateur unitaire, il faut y ajouter la porte d'Hadamard qui

engendre des superpositions, ainsi que la porte de phase qui, elle, engendre les complexes. En réalité, si on cherche un ensemble de portes qui est capable d'effectuer n'importe quel calcul quantique sans perdre d'efficacité plutôt que n'importe quel unitaire, il est tout à fait possible de se passer des complexes, et seules les portes de Toffoli et d'Hadamard sont nécessaires [332, 9]. Cela vient du fait que l'on peut transformer une amplitude complexe en deux amplitudes réelles facilement avec ces jeux de portes. Ceci indique que le « seul » ajout nécessaire pour obtenir toute la puissance de calcul de la mécanique quantique est le principe de superposition.

Algorithme de Solovay-Kitaev Une fois que nous avons une famille que nous savons universelle, la question est alors de savoir s'il est possible d'approximer rapidement n'importe quel opérateur à partir de cette famille. Il convient de noter que cette question n'a rien d'évident au premier abord, car les démonstrations que nous avons données de l'universalité ne sont pas constructives. Nous avons utilisé le fait qu'une rotation autour d'un axe dont l'angle est irrationnel à 2π génère un ensemble dense dans les rotations autour de cet axe. Mais il se pourrait très bien que les rotations générées s'accumulent autour de certains points d'abord, et qu'obtenir de bonnes approximations pour certaines rotations nécessite de nombreuses itérations de la rotation initiale.

Cette question est résolue par le théorème de Solovay-Kitaev, qui énonce que si l'on a une famille universelle de portes, alors il est possible de générer rapidement une bonne approximation de n'importe quelle porte quantique [231, 122]. Plus précisément, si l'on fixe le nombre de qubits, il est possible d'approximer n'importe quel opérateur unitaire à une erreur δ près avec au plus $\mathcal{O}(\log^c(1/\delta))$ portes de notre ensemble universel, où c est une constante valant environ 4. Si l'on considère un circuit avec L portes à k qubits, k étant fixé, et que l'on souhaite l'approximer à δ près, il faut approximer chaque porte avec une précision δ/L , les erreurs étant linéaires en la longueur du circuit. Il est donc possible d'approximer le circuit à δ près à l'aide de $\mathcal{O}(L \log^c(L/\delta))$ portes issues de l'ensemble discret, car il est nécessaire d'approximer chaque porte avec une précision δ/L . Ainsi, on voit que la simulation est extrêmement efficace, étant donné qu'elle ne rajoute qu'un facteur $\log^c(L/\delta)$. Cela justifie notamment notre définition de BQP, qui n'impose pas que l'ensemble de portes soit discret.

L'algorithme de Solovay-Kitaev permet de construire, de manière itérative, de meilleures approximations de n'importe quel opérateur unitaire. Pour cela, l'algorithme part d'un maillage de $SU(n)$ suffisamment fin, qui donne donc une approximation suffisamment précise de n'importe quel opérateur unitaire. À partir de là, il est possible de générer des maillages de plus en plus fins en effectuant des opérations sur les maillages précédents. Néanmoins, pour faire ces opérations, il est nécessaire de partir d'un réseau initial suffisamment fin. Cela nécessite une étape de génération de ce premier réseau, ce que l'on peut

faire par recherche exhaustive. Il faut noter que l'algorithme de Solovay-Kitaev est impraticable au-delà des portes à quelques qubits. En effet, sa complexité en le nombre de qubits n est $\mathcal{O}(n^2 4^n \log^c(n^2 4^n / \delta))$. Il est donc difficile de générer un circuit à partir d'un opérateur unitaire.

8.4.3 Codes correcteurs d'erreurs

Dans un ordinateur classique, la probabilité d'une erreur est suffisamment faible pour qu'en pratique on considère que le processeur opère sans erreur¹². Néanmoins, il reste important de pouvoir détecter et corriger les erreurs à la volée, notamment depuis les périphériques de stockage de masse, qui stockent une grande quantité d'information sur de longues périodes. Ainsi, les codes correcteurs d'erreurs sont omniprésents dans les disques durs et les SSD. Néanmoins, l'information présente dans chacun de ces circuits est déjà hautement redondante. Chaque transistor d'un processeur voit passer quelques centaines à quelques milliers d'électrons par cycle, et chaque cellule de DRAM contient quelques dizaines de milliers d'électrons. C'est pour cela que les taux d'erreurs sont extrêmement bas : la redondance de l'information est présente à tous les étages d'un ordinateur classique, agissant comme le plus simple des codes correcteurs d'erreurs. De plus, même lorsque l'on utilise des codes correcteurs d'erreurs, il est rare que les ressources allouées à la correction d'erreurs représentent une fraction importante de celles allouées au message.

Nous avons vu dans la section 6.2 qu'il n'est pas possible de copier de l'information quantique comme on peut le faire avec l'information classique. Cette différence fondamentale fait que l'on ne peut pas, contrairement au classique, corriger les erreurs simplement en faisant de nombreuses copies de l'état quantique. De prime abord, il semble donc qu'un qubit est limité à un seul degré de liberté ce qui implique une grande sensibilité aux erreurs. De plus, les qubits évoluant sur un espace continu, il faut pouvoir corriger les erreurs provenant des incertitudes que l'on a sur les portes quantiques. Idéalement, il nous faut donc un code correcteur d'erreurs quantiques à même de diminuer l'impact du bruit provenant de l'environnement, ainsi qu'une façon de mettre en œuvre les portes de base au-dessus de ce code correcteur qui puisse limiter les erreurs systématiques provenant de chaque porte.

Le sujet des codes correcteurs d'erreurs quantiques est vaste, et nous allons donner ici une approche simple qui synthétise les idées essentielles derrière ces codes. Le lecteur souhaitant approfondir ces notions pourra se référer à [231, 287, 304].

12. Les études sur la DRAM font état d'un taux d'erreur allant d'une erreur par heure par gigaoctet à une erreur par siècle par gigaoctet [347, 323, 257]. Si l'on trouve de la DRAM avec correction d'erreurs sur les serveurs haut de gamme, ces taux sont suffisamment faibles pour s'en passer sur des machines grand public.

Les types d'erreurs

Avant de parler de codes correcteurs d'erreurs, il nous faut bien sûr parler des erreurs qu'ils devront corriger. Dans le cas classique, la seule erreur possible sur un circuit est l'erreur de type *bit flip*, qui effectue un **not** aléatoirement sur les fils du circuit. En général on suppose que les erreurs ne sont pas corrélées d'un fil à l'autre. Quelles sont les erreurs possibles dans le cas quantique ? Il y a là un continuum de possibilités, mais nous pouvons identifier trois types d'erreurs, correspondant chacune à des rotations sur la sphère de Bloch. Il y a les erreurs de type *bit flip*, tout comme dans le cas classique qui transforment $|0\rangle$ en $|1\rangle$ et inversement, et dont l'action est représentée par l'opérateur σ_x . Le second type d'erreurs sont les erreurs de type *phase flip* qui transforment $|1\rangle$ en $-|1\rangle$ tout en laissant $|0\rangle$ invariant, et dont l'action est représentée par σ_z . Enfin, si on sait résoudre ces deux erreurs, on peut également résoudre les erreurs de type σ_y , car on peut obtenir ce dernier opérateur à partir du commutateur entre σ_x et σ_z . En corrigeant simultanément les erreurs de type *bit flip* et de type *phase flip*, on peut donc corriger n'importe quelle erreur.

Le code *bit flip*

Nous allons d'abord nous intéresser aux erreurs de type *bit flip*. Pour cela nous allons nous intéresser au code *bit flip*, qui consiste simplement à encoder le qubit logique $|\psi_L\rangle = \alpha|0\rangle + \beta|1\rangle$, au moyen de trois qubits physiques :

$$|\psi_P\rangle = \alpha|000\rangle + \beta|111\rangle . \quad (8.66)$$

Ce code n'est pas très éloigné du code par répétition utilisé dans les codes correcteurs classiques. Afin de corriger une éventuelle erreur, la stratégie est d'effectuer une mesure sur les qubits physiques qui donne une information sur l'erreur sans donner d'information sur le contenu du qubit. Pour cela on utilise les mesures suivantes :

$$P_0 = |000\rangle\langle 000| + |111\rangle\langle 111| , \quad (8.67)$$

$$P_1 = |100\rangle\langle 100| + |011\rangle\langle 011| , \quad (8.68)$$

$$P_2 = |010\rangle\langle 010| + |101\rangle\langle 101| , \quad (8.69)$$

$$P_3 = |001\rangle\langle 001| + |110\rangle\langle 110| . \quad (8.70)$$

Le projecteur P_0 correspond au cas où il n'y a pas eu d'erreur. Les projecteurs P_1 , P_2 et P_3 correspondent aux cas où seul le qubit 1, 2 ou 3 s'est retourné. Le résultat de la mesure est appelé *syndrome d'erreur*. Ainsi dans le cas où seul un qubit physique change, on obtient une information sur l'erreur sans obtenir d'information sur le qubit logique. Notons que les opérations de projections nous font passer d'un spectre continu d'erreurs à un spectre discret. Dans le cas où un seul qubit physique s'est retourné, il est possible de retrouver l'état initial. Il suffit pour cela d'appliquer l'opérateur σ_x au qubit physique fautif si nécessaire.

Si l'un des qubits physiques fait une rotation d'angle θ sur un méridien de la sphère de Bloch alors que les autres restent constants, après l'opération de projection on trouvera que l'un des qubits s'est retourné avec une probabilité $\sin^2 \theta$ et que l'état n'a pas changé avec une probabilité $\cos^2 \theta$. On peut donc ne considérer que des erreurs qui retournent complètement les qubits avec une probabilité p . Si on suppose que les erreurs sont indépendantes et que leur probabilité ne dépend pas du qubit physique, la probabilité d'une erreur non corrigée est celle de retourner au moins deux qubits. Cela arrive avec une probabilité $3p^2 - 2p^3$.

En revanche les erreurs de type *phase flip* sur chacun des qubits physiques s'accumulent au niveau du qubit logique. On va donc s'intéresser au code *phase flip* qui lui ne corrige que ce type d'erreur.

Le code *phase flip*

L'idée derrière le code *phase flip* n'est pas très différente de celle du code *bit flip*. La version discrète de l'erreur *phase flip* est la transformation de l'état $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ en l'état $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ et inversement. C'est l'analogue d'une erreur de type *bit flip* dans la base $(|-\rangle, |+\rangle)$. Le code *phase flip* utilise naturellement l'encodage suivant pour l'état logique $|\psi_L\rangle$:

$$|\psi_P\rangle = \alpha |+++\rangle + \beta |--\rangle . \quad (8.71)$$

Les mesures sont données par les projecteurs suivants :

$$P_0 = |+++\rangle \langle +++| + |--\rangle \langle --| , \quad (8.72)$$

$$P_1 = |-\rangle \langle +| + |+\rangle \langle -| + |+-\rangle \langle +-| , \quad (8.73)$$

$$P_2 = |+-\rangle \langle +-| + |-+\rangle \langle -+| , \quad (8.74)$$

$$P_3 = |++-\rangle \langle ++-| + |--+\rangle \langle --+| . \quad (8.75)$$

Le syndrome d'erreur permet de déduire quel qubit a subi un déphasage si tel est le cas, et permet donc de corriger le déphasage. Les propriétés de ce code sont, sans surprises, similaires au code *bit flip*, mis à part le fait que les erreurs corrigées sont de type *phase flip*. On peut donc corriger n'importe quelle erreur de type *phase flip* affectant un seul qubit.

Le code de Shor

Afin d'avoir un schéma de correction d'erreurs complet, il nous faut pouvoir corriger à la fois les erreurs de type *bit flip* et les erreurs de type *phase flip*. Pour cela, nous allons utiliser le code de Shor, qui concatène les deux codes correcteurs vus précédemment. Le qubit logique est d'abord encodé selon le code *phase flip*. Chaque qubit du codage *phase flip* est lui-même un qubit logique qui est encodé sur un code *bit flip*. Un qubit logique est donc encodé

sur neuf qubits physiques. Concrètement, on a :

$$|0_L\rangle = |+\varphi + \varphi + \varphi\rangle = \frac{(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)}{2\sqrt{2}}, \quad (8.76)$$

$$|1_L\rangle = |-\varphi - \varphi - \varphi\rangle = \frac{(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)}{2\sqrt{2}}. \quad (8.77)$$

La correction d'erreurs s'effectue alors en deux étapes. Tout d'abord on identifie et on corrige les erreurs de type *bit flip* sur chacun des trois groupes de trois qubits physiques. Ensuite, on identifie et on corrige les erreurs de type *phase flip* au niveau de l'étage logique intermédiaire.

Une propriété remarquable du code de Shor est qu'il permet de corriger n'importe quelle erreur à un seul qubit. Ainsi, même si l'un des qubits est remis à zéro, il est possible de rétablir l'état initial. Ce résultat est remarquable car, naïvement, on pourrait s'attendre à ce que le fait de pouvoir observer une partie du système altère systématiquement l'information utile sur le système complet. Il se trouve qu'ici un seul qubit ne contient aucune information sur l'état $|\psi_L\rangle$. L'ensemble de l'information sur $|\psi_L\rangle$ est encodé de manière globale sur les neuf qubits physiques, l'intrication jouant un rôle essentiel dans la correction d'erreurs. Notons également que l'hypothèse selon laquelle les erreurs arrivent de manière indépendante sur chacun des qubits est cruciale, car c'est elle qui assure que l'on ne modifie l'état physique que localement. Il n'est également pas évident que cette hypothèse soit vérifiée dans les systèmes physiques. Les systèmes quantiques étant en général petits, le fait de garder des erreurs localisées est un réel défi expérimental.

Quelques bornes sur les codes correcteurs d'erreurs

Avec le code de Shor, nous avons vu qu'il est possible, en principe, de corriger n'importe quelle erreur affectant un des neuf qubits physiques. Cependant, le code de Shor est extrêmement simple et il est possible de faire mieux. Il existe cependant certaines limites à ce qu'il est possible de faire et ce sont ces limites que nous allons voir ici. Pour cela, nous allons devoir introduire quelques notions utiles pour étudier les codes correcteurs d'erreurs dans leur ensemble.

Nous allons ici considérer que les erreurs peuvent arriver sur un nombre non déterminé de qubits. Une erreur est représentée par un opérateur E qui est un produit tensoriel d'opérateurs de Pauli et de l'identité. À première vue, l'absence de corrélation quantique dans les erreurs semble limitant, mais il est possible de se ramener à de telles erreurs par la mesure, tout comme nous sommes passés d'un nombre infini d'erreurs possibles à un nombre fini. Nous appellerons *poids de l'opérateur* E le nombre de qubits sur lequel il agit de manière non triviale. Étant donné que E est un produit tensoriel d'opérateurs de Pauli, nous avons $E^\dagger = E = E^{-1}$.

Une dernière notion que nous avons besoin d'introduire est la *distance* d'un code. Un code correcteur d'erreurs possède les opérations de correction

d'erreurs, mais également des opérations logiques, qui permettent de changer l'état du qubit logique. Ces opérations logiques doivent agir sur un certain nombre de qubits. Le nombre minimum de qubits concernés par une opération logique est appelé la distance du code. Dans le code de Shor par exemple, il est nécessaire d'effectuer des opérations sur au moins trois qubits physiques pour réaliser une opération sur le qubit logique, on a donc une distance $d = 3$. En revanche, pour le code *bit flip*, on peut réaliser l'opération σ_z avec un seul qubit, ce qui implique $d = 1$.

La distance est une donnée extrêmement intéressante pour savoir le poids maximal des erreurs que l'on peut corriger. Les erreurs peuvent être comprises comme des mesures effectuées par l'environnement de notre circuit quantique. Afin de préserver la cohérence quantique à l'échelle des qubits logiques, les erreurs ne doivent pas emporter d'information sur les qubits logiques. Cela implique que, si l'on part d'états logiques $|i\rangle$ et $|j\rangle$ différents et que l'on applique une erreur E_a sur l'état $|i\rangle$ et une erreur E_b sur l'état $|j\rangle$, les états obtenus doivent rester orthogonaux. Dans le cas contraire, on ne pourra pas corriger l'erreur, car on ne saura pas discerner entre ces deux états altérés. De même, si on part de deux états identiques et que l'on applique des erreurs différentes, le produit scalaire ne doit pas dépendre des états de départ. Sinon, cela signifie que les erreurs emportent l'information quantique à l'échelle des qubits logiques vers l'environnement. On a donc

$$\langle j|E_b^\dagger E_a|i\rangle = C_{ab}\delta_{ij}. \quad (8.78)$$

L'ensemble des erreurs qui obéissent à cette condition peuvent être corrigées par le code. À première vue, on peut chercher à imposer la condition plus forte $C_{ab} = \delta_{ab}$. Dans ce cas on a un code non dégénéré dans lequel chaque erreur peut être corrigée de façon unique. Cependant, il est possible de corriger certaines erreurs de plusieurs façons, comme c'est le cas dans le code de Shor, pour lequel on peut corriger les erreurs de phases de multiples façons. Par simplicité, on va considérer que l'ensemble des erreurs possède un poids t ou moins. On voit ainsi qu'un code qui possède une distance $d = 2t + 1$ peut corriger des erreurs de poids t (on dira également que l'on peut corriger t erreurs, sous-entendu agissant sur un qubit isolé). En effet, la correction d'une erreur E_a de poids t se fait en appliquant un opérateur E_b tel que l'élément C_{ab} soit non nul. Après correction d'erreurs, on a donc l'état $E_b^\dagger E_a|i\rangle$. Étant donné que les opérations logiques sur le code sont de poids au moins $2t + 1$ et que l'opération $E_b^\dagger E_a$ est de poids au plus $2t$, l'état logique après correction est équivalent à l'état initial. Enfin, si l'on sait sur quel qubit physique l'erreur se situe, il suffit que le code ait une distance $t + 1$ pour corriger t erreurs. Ainsi, le code de Shor a une distance de 3 et peut corriger deux erreurs localisées ou bien une erreur non localisée.

On peut synthétiser les informations utiles sur un code en trois nombres, ce que l'on notera $[[n, k, d]]$. Dans cette notation, n est le nombre de qubits physiques utilisés pour encoder k qubits logiques et d est la distance du code.

On peut alors se demander quelles sont les relations entre n , k et d , c'est-à-dire quelles sont les limites à la correction d'erreurs quantiques.

Une première borne, valide pour les codes non dégénérés est la borne de Hamming quantique. Il s'agit de compter le nombre d'erreurs à corriger et à le comparer au nombre de degrés de liberté disponibles dans l'espace du code. Du côté des erreurs, si l'on considère les erreurs de poids j , on a $3^j \binom{n}{j}$ possibilités, car les qubits peuvent avoir une erreur selon σ_x , σ_y ou σ_z et que les erreurs ne sont pas localisées. D'un autre côté, si l'on code k qubits logiques avec n physiques, on a un espace de dimension 2^{n-k} pour corriger les erreurs. On a donc la borne suivante [180], où $d = 2t + 1$:

$$\sum_{j=0}^t 3^j \binom{n}{j} \leq 2^{n-k}. \quad (8.79)$$

On notera que pour $k = 1$ et $t = 1$, cette borne est satisfaite pour $n \geq 5$. Si l'on regarde les codes non dégénérés capables de corriger un seul qubit, il faut alors au moins $n = 5$ qubits. En fait, il existe un code $[[5, 1, 3]]$, qui sature la borne de Hamming quantique. Insistons sur le fait que cette borne n'est valide que pour les codes non dégénérés. Cependant, on ne connaît actuellement pas de code dégénéré dépassant cette borne.

Une autre contrainte, valable pour l'ensemble des codes, vient du théorème de non-clonage que nous avons vu au chapitre 6. En effet, si on a un code $[[n, k, 2t + 1]]$, on peut reconstruire $2t$ erreurs localisées. Il est alors tout à fait possible d'oublier $2t$ qubits. Supposons que l'on puisse avoir $n \leq 4t$. On pourrait alors décider d'envoyer la moitié des qubits à Alice et l'autre à Bob. Alice et Bob pourraient alors chacun reconstruire l'état logique complet, dupliquant ainsi l'état quantique. On a donc nécessairement $n > 4t$. En fait on peut donner une version plus forte de cette borne, nommée borne de Singleton quantique [235] :

$$n - k \geq 4t. \quad (8.80)$$

Cette borne, ainsi que la borne de non-clonage est saturée par le code $[[5, 1, 3]]$. Enfin, il existe une borne obtenue par Rains, plus précise dans certains cas, valide pour $k \geq 1$ [307] :

$$t \leq \left\lfloor \frac{n+1}{6} \right\rfloor. \quad (8.81)$$

Les familles de codes correcteurs d'erreurs

Le code de Shor a l'avantage d'être extrêmement simple à construire. En revanche, étant donné qu'il s'agit d'un code $[[9, 1, 3]]$, il est loin d'être optimal. Cependant il existe des outils plus puissants pour construire des codes correcteurs. Nous allons ici donner seulement les idées derrière certains ensembles de codes correcteurs d'erreurs.

Une grande famille de codes correcteurs d'erreurs sont les codes stabilisateurs [181]. Pour construire un de ces codes, il est nécessaire de partir d'un

groupe d'opérateurs qui commutent ne contenant pas $-\mathbb{1}$. Ces opérateurs sont choisis de façon à former un sous-groupe $\mathcal{S}$ du groupe de Pauli qui est généré par l'ensemble des opérateurs de Pauli appliqués indépendamment à chaque qubits ainsi que $i\mathbb{1}$. L'ensemble des états correspondant à des qubits logiques définit un sous-espace $\mathcal{H}_{\mathcal{S}}$ de l'espace de Hilbert des états des qubits physiques. Cet espace est appelé *espace du code*, et pour un code stabilisateur il correspond à l'ensemble des vecteurs propres laissés invariants par le groupe stabilisateur $\mathcal{S}$:

$$M|\psi\rangle = |\psi\rangle, \forall |\psi\rangle \in \mathcal{H}_{\mathcal{S}}, \forall M \in \mathcal{S}. \quad (8.82)$$

Le stabilisateur peut se caractériser par l'ensemble de ses générateurs. Pour obtenir le syndrome d'erreur, il suffit alors d'effectuer la mesure associée à chacun des générateurs. Par définition de $\mathcal{H}_{\mathcal{S}}$, il y a une erreur à corriger si au moins un des générateurs du stabilisateur mesure une valeur -1 .

La plupart des codes construits jusqu'alors relèvent du formalisme des codes stabilisateurs. C'est notamment le cas du code de Shor que nous avons vu. Le code $[[5, 1, 3]]$ en fait partie, ainsi qu'un certain nombre d'autres codes. Une sous-famille des codes stabilisateurs sont les codes de Calderbank-Shor-Steane (CSS) [79, 349], qui sont construits à partir de codes correcteurs classiques, ce qui donne une façon constructive de construire des codes correcteurs quantique, à partir de codes correcteurs classiques. Un code CSS simple est le code $[[7, 1, 3]]$ de Steane.

Une autre famille de codes de grand intérêt sont les *codes topologiques* [164]. Avec les codes correcteurs que nous avons vus, nous n'avons mis aucune contrainte sur l'architecture sous-jacente. Cependant il peut y avoir des difficultés particulières, notamment s'il faut réaliser des portes entre des qubits physiquement distants afin d'assurer la correction d'erreurs. Les codes correcteurs d'erreurs topologiques ont l'avantage d'être définis sur des réseaux de qubits bidimensionnels, dans lesquels les portes sont locales. Tout d'abord inventés par Kitaev [230] pour des géométries toroïdales, ils ont été généralisés à des surfaces planes par la suite. Ces codes sont également des codes stabilisateurs. Dans un code topologique de surface, il est possible de créer des défauts en éteignant sélectivement des qubits physiques. Les paires de défauts jouent alors le rôle de qubits logiques. On peut déplacer ces défauts, et les opérations logiques sont réalisées en tressant les défauts les uns autour des autres. Il est ainsi possible de faire du calcul en effectuant des opérations topologiques au niveau du réseau bidimensionnel de façon résistante aux erreurs. Cette idée d'effectuer les calculs au niveau des qubits logiques n'est cependant pas limitée aux seuls correcteurs d'erreurs topologiques, et c'est ce dont nous allons discuter à présent.

Portes tolérantes aux erreurs

Les codes correcteurs permettent d'assurer la transmission sans erreur de l'état d'un qubit, moyennant certaines hypothèses sur les erreurs. En revanche

la transmission n'est pas la seule source d'erreur possible. Chaque porte d'un circuit quantique introduit des erreurs, à la fois statistiques, à cause du bruit environnemental, et systématiques car la mise en œuvre de la transformation unitaire désirée n'est qu'approximative. Ces erreurs se propagent et s'accumulent le long du circuit et, pour peu que le circuit soit suffisamment long, rendent le calcul inopérant. La logique tolérante aux erreurs permet de calculer directement sur des qubits logiques contenant de la correction d'erreurs. La difficulté ici est d'avoir une mise en œuvre de l'ensemble de la chaîne qui soit tolérante aux erreurs. Il faut donc, d'une part, mettre en œuvre l'ensemble des portes de base agissant sur les qubits logiques à partir des portes de base agissant sur les qubits physiques, et cela de façon tolérante aux erreurs. Cela permettra de limiter les erreurs générées par les portes de base à chaque étape de calcul. De même, il est nécessaire de mettre en œuvre les étapes d'encodage et de correction d'erreurs à partir des portes et des mesures de base, sans que cela n'ajoute plus d'erreurs que cela n'en corrige.

Nous n'entrerons pas dans les détails de telles constructions. Il convient de noter qu'elles dépendent fortement de l'ensemble universel de portes logiques que l'on choisit, ainsi que du code correcteur d'erreurs. Certains ensembles universels de portes logiques peuvent être plus aisés à réaliser avec certains codes correcteurs d'erreurs et certains ensembles universels de portes physiques. Ce seront probablement sur ces considérations que se fera le choix de ces différents ensembles de portes ainsi que du code correcteur. La taille de chaque porte logique avec correction d'erreurs demande de très nombreuses portes physiques. D'une part, expérimentalement, il est vraisemblable que certains types de porte soient plus faciles à réaliser, avec moins d'erreurs, que d'autres. D'autre part, le nombre de portes physiques pour chaque porte tolérante aux erreurs dépend à la fois des ensembles de portes que l'on choisit au départ ainsi que du code correcteur. Il y a donc un compromis à trouver entre la taille du circuit et les contraintes expérimentales afin de garantir la meilleure fidélité du circuit final. Il faut également noter que la logique tolérante aux erreurs nécessite d'effectuer des mesures, ce qui en pratique revient à assurer un approvisionnement régulier en qubits auxiliaires initialisés dans l'état $|0\rangle$ [10].

Enfin, pour que le code correcteur ait un intérêt, il est nécessaire que la fidélité de chaque porte physique soit suffisamment élevée. En effet, si on suppose que chaque porte introduit des erreurs indépendantes sur chaque qubit avec une probabilité p , la correction d'erreurs va systématiquement corriger les erreurs n'affectant qu'un seul qubit, soit au premier ordre en p ¹³. Ainsi, une porte logique tolérante aux erreurs possède elle-même une erreur résiduelle, quadratique en p . Chaque porte logique possède donc une erreur cp^2 où c est une constante liée à la taille du circuit. Afin de corriger plus d'erreurs que

13. En réalité, on peut également être intéressé par les erreurs affectant deux ou trois qubits de façon corrélée, notamment pour les portes à deux ou trois qubits. Il est néanmoins possible de raffiner l'analyse ci-dessus afin que celle-ci reste valide. Il faut également noter que, selon le type d'erreur, il faille des codes correcteurs plus sophistiqués que ceux que nous avons développés jusqu'à présent.

l'on en introduit, il est donc impératif que $p < 1/c$. Pour le code de Steane sur la base standard, c est de l'ordre de 10^5 – 10^6 , ce qui impose d'avoir des portes fiables à 10^{-6} – 10^{-5} près. En revanche les codes de surface ont des seuils beaucoup plus faibles et nécessitent des portes fiables à 10^{-2} près seulement, au prix d'un codage sur un grand nombre de qubits physiques.

Théorème du seuil

Ainsi, s'il est possible de réaliser chaque porte physique avec une probabilité d'erreur p_0 inférieure à une valeur seuil $p_{\text{th}} = 1/c$, il est également possible de fabriquer une porte logique dont la probabilité d'erreur est $p_1 = p_0^2/p_{\text{th}} < p_0$. On peut alors itérer le processus en utilisant un code correcteur par-dessus ces qubits logiques, de façon à réduire la probabilité d'erreur à $p_2 = p_1^2/p_{\text{th}}$. C'est la même stratégie de concaténation que nous avons utilisée pour obtenir le code de Shor. Ainsi, à chaque étage, la probabilité d'obtenir une erreur est plus petite, la probabilité à l'étape n étant donnée par :

$$p_n = p_{\text{th}} \left(\frac{p_0}{p_{\text{th}}} \right)^{2^n}. \quad (8.83)$$

Ainsi, on voit que la probabilité d'erreur décroît extrêmement rapidement avec le nombre d'enchâssements que l'on réalise sur le code correcteur d'erreurs. Notons que le nombre de portes, lui, croît en d^n , où d est le nombre de portes maximal nécessaire pour réaliser n'importe quelle porte de l'étage $n + 1$ avec des portes de l'étage n . Ainsi, si l'on souhaite réaliser des portes avec une précision ϵ , chaque porte logique sera constituée de l'ordre de

$$\left(\frac{\log(p_{\text{th}}/\epsilon)}{\log(p_{\text{th}}/p_0)} \right)^{\log d} = \text{poly}(\log(1/\epsilon)). \quad (8.84)$$

En d'autres termes, pour peu que l'on ait $p_0 < p_{\text{th}}$, en enchâssant les codes correcteurs d'erreurs, il est possible de réduire l'erreur avec un surcoût polynomial en le logarithme de l'erreur cible. Si l'on souhaite obtenir un circuit ayant $g(n)$ portes, où n est la taille de l'entrée, avec une précision ϵ , chaque porte doit alors avoir une erreur $\epsilon/g(n)$. Cela nous amène au théorème du seuil.

Théorème 8.2 (Théorème du seuil). *Un circuit quantique contenant $g(n)$ portes peut être simulé avec une probabilité d'erreur au plus ϵ , avec de l'ordre de $\mathcal{O}(\text{poly}(\log(g(n)/\epsilon))g(n))$ portes, pourvu que chaque porte physique ait une probabilité d'erreur $p_0 < p_{\text{th}}$.*

Le théorème du seuil implique donc que, si le taux d'erreur est suffisamment bas, il est possible de corriger les erreurs plus vite qu'elles ne s'accumulent avec un coût relativement faible en termes de nombre de portes. L'importance conceptuelle de cela est qu'il est possible d'effectuer la correction d'erreurs sans changer de classe de complexité. Cependant, le théorème du seuil s'appuie sur quelques points qu'il est absolument nécessaire de préciser.

Tout d'abord, la correction d'erreurs doit être effectuée en parallèle des calculs. Cela convient tout à fait aux modèles basés sur des circuits, dans lesquels les portes travaillent de manière indépendante, et en parallèle. Il est nécessaire que l'architecture physique possède ce parallélisme. De plus les circuits de correction d'erreurs contiennent une composante classique, qui traite le syndrome, et qui implique potentiellement une communication entre plusieurs parties du circuit. Notamment, lorsque les codes correcteurs sont enchâssés, cela peut impliquer des communications entre des parties physiquement éloignées du circuit. Ces calculs classiques et ces communications doivent également se faire suffisamment rapidement pour assurer la correction d'erreurs. L'architecture physique doit également fournir au circuit des qubits auxiliaires de façon régulière, correctement initialisés.

Enfin, une hypothèse importante concerne le modèle de bruit que nous avons utilisé. Dans le cas présent, nous avons utilisé un modèle de bruit extrêmement simple, qui agit indépendamment sur chacun des qubits. Il est néanmoins possible d'utiliser les mêmes méthodes dans des cas impliquant des erreurs à plusieurs qubits, à la condition que les corrélations des erreurs entre les différents qubits ne soient pas trop fortes. La mise en place de codes correcteurs d'erreurs nécessite donc une modélisation des erreurs à partir de la physique de l'architecture considérée. C'est l'architecture physique, à partir des portes physiques que l'on peut réaliser, ainsi que les types et les taux d'erreurs en jeu, qui va permettre de décider quel code correcteur d'erreurs employer. L'objet des prochains chapitres sera de voir l'origine physique de ces erreurs.

8.5 ☪ La simulation quantique

L'idée de la simulation quantique, due à Feynman [158], consiste à étudier un système physique donné non pas directement mais réalisant un autre système, mieux contrôlé, dont on peut montrer que la dynamique est exactement celle du modèle dont on souhaite confronter les prédictions avec les résultats expérimentaux.

Ce thème de recherche s'est grandement développé ces dernières années. À peu près tous les systèmes permettant la manipulation contrôlée d'états quantiques ayant été envisagés comme plate-forme pour la simulation quantique dans tous les domaines de la physique comme exposé dans la revue [173].

8.5.1 ☪ Enjeux et difficultés de la simulation quantique

☪ Pourquoi la simulation quantique ?

L'intérêt de la simulation quantique vient du fait qu'obtenir des résultats *via* une modélisation d'un système quantique sur un ordinateur classique s'avère rapidement gourmand en ressources. En effet, la modélisation complète

de la dynamique d'un système composé de n unités élémentaires (qubits, spins ou états dans un puits anharmonique) oblige à travailler dans un espace dont la dimension croît exponentiellement avec n . Les méthodes dites exactes se trouvent donc rapidement limitées. À l'opposé, pour des échantillons de grande taille, les méthodes de la physique statistique permettent d'obtenir des résultats sur le comportement des grands systèmes [251].

Les polymères sont constitués de molécules complexes dont le comportement physique fait intervenir de nombreuses échelles de longueur qui vont de l'échelle atomique (quelques Å) jusqu'à l'échelle macroscopique (mm) en passant par diverses échelles intermédiaires comme le rayon de giration (quelques nm) et/ou la taille de micelles en solution (quelques μm). Modéliser chacune de ces échelles fait intervenir des méthodes radicalement différentes, depuis celles de la chimie quantique pour les briques de base du polymère jusqu'à celles de la physique statistique, basées sur des modèles discrets ou continus effectifs [145, 163, 172, 26], pour décrire le comportement à grande échelle en passant par des méthodes comme la dynamique moléculaire qui ne retiennent que certains des éléments essentiels pour modéliser le polymère aux échelles intermédiaires [54]. Un des problèmes majeur est d'assurer que l'ensemble des méthodes utilisées constitue une simulation multi-échelle cohérente du matériau considéré. On peut alors être contraint d'explorer des zones grises où les méthodes exactes ne sont plus utilisables mais où l'on ne dispose pas de schémas d'approximations bien contrôlés qui permettent de modéliser efficacement le système. C'est ici que la simulation quantique prend tout son intérêt : elle fournit un outil qui permet soit d'obtenir des prédictions dans ce régime ou, ce qui est plus intéressant encore, d'éprouver de manière contrôlée des modèles effectifs que l'on pense pertinents à ces échelles intermédiaires dites mésoscopiques.

Dans le même esprit, la physique moléculaire et la chimie théorique sont confrontées au même problème : même si on dispose d'un arsenal de méthodes théoriques et numériques pour calculer les spectres des atomes et des molécules, le temps de calcul sur ordinateur classique croît exponentiellement avec la taille des molécules. L'enjeu est pourtant extrêmement important. Calculer ne serait-ce que l'énergie du fondamental d'une molécule pour une position donnée des noyaux permet de faire des prédictions pour les taux de réactions chimiques en fonction de la température [279].

Un dénominateur commun à tous ces problèmes de simulation quantique est le calcul du spectre du hamiltonien de systèmes physiques ou, si on se limite à un objectif plus modeste, la caractérisation de son état fondamental et, encore plus *a minima*, le calcul de son énergie. Les hamiltoniens physiques sont locaux, c'est-à-dire que l'on peut toujours les réécrire comme un hamiltonien agissant sur un espace des états $\mathcal{H}^{\otimes N}$ où $\mathcal{H}$ est de dimension finie¹⁴. La localité implique que H est somme de contributions qui ne couplent qu'un

14. Ceci revient à discrétiser le problème soit en considérant un modèle sur réseau, soit en tronquant l'espace des états par exemple en ne considérant que certaines orbitales en chimie.

nombre fini de sites. On appelle hamiltonien p -local, un hamiltonien de la forme $H = \sum_{\alpha} H_{\alpha}$ où chaque H_{α} agit sur p composantes du produit tensoriel $\mathcal{H}^{\otimes N}$. On pourrait espérer que cette contrainte très forte limite la complexité du problème considéré pour un hamiltonien local couplant p voisins en fonction de la taille N du système.

✂ La classe des problèmes quantiques difficiles

Les informaticiens attaquent le problème de l'énergie du fondamental en cherchant des algorithmes qui permettent d'encadrer l'énergie du fondamental dans une fenêtre d'énergie dont la largeur décroît polynomialement en N . Bien entendu, si on se donne un vecteur d'état $|\psi\rangle$, le calcul de l'énergie moyenne $\langle\psi|H|\psi\rangle$ est facile et donc on s'attend à ce que la vérification que $|\psi\rangle$ soit dans la fenêtre visée soit facile. On dispose donc d'un certificat quantique efficace pour ce problème.

Cela ressemble à un problème NP, à ceci près que le certificat est quantique et que la vérification du certificat est dans BQP et non dans P. Cela nous amène à introduire la classe QMA¹⁵, définie dans [234, 231]. Pour généraliser la classe NP au cas quantique, on va donc considérer des problèmes de décision X tels qu'il existe un problème $Y \in \text{BQP}$ qui est le vérifieur quantique ainsi qu'un polynôme p qui borne la taille du certificat quantique. Toute instance x de X de taille n vérifie alors :

- si x est dans X , il existe un état $|\chi\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$ tel que $(x, |\chi\rangle)$ soit accepté par Y avec une probabilité plus grande que $2/3$;
- si x n'est pas dans X , alors pour tout état $|\chi\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, $(x, |\chi\rangle)$ sera accepté par X avec une probabilité inférieure à $1/3$.

Dans le cas du problème de l'encadrement de l'énergie du fondamental, l'instance x est une description compacte du hamiltonien H , et le vecteur $|\chi\rangle$ est le vecteur sur lequel on calcule l'énergie $\langle\chi|H|\chi\rangle$.

Bien entendu $\text{BQP} \subseteq \text{QMA}$. On peut se demander si on peut borner supérieurement la classe QMA. Watrous et Kitaev [232] ont montré que $\text{QMA} \subseteq \text{PP}$. Ceci qui permet finalement d'avoir la hiérarchie de classes

$$\text{BPP} \subseteq \text{BQP} \subseteq \text{QMA} \subseteq \text{PP}. \quad (8.85)$$

Enfin, une dernière relation nous donne la relation entre classes de problèmes difficiles classiques et quantiques :

$$\text{NP} \subseteq \text{QMA} \quad (8.86)$$

qui exprime la hiérarchie des puissances de calculs entre le quantique et le classique. Une liste des problèmes QMA-complets a été compilée [61].

15. QMA pour *Quantum Merlin Arthur* par référence à une classe de protocoles interactifs impliquant Merlin, être omniscient, capable de sortir la solution à un problème difficile de son chapeau et Arthur, pauvre mortel ne disposant que de capacités de calcul ordinaires.

Pour en revenir à la simulation quantique, le résultat général [311] est que le problème de l'encadrement de l'énergie du fondamental d'un hamiltonien couplant $p = 2$ voisins est génériquement QMA-complet. Concrètement, cela signifie que le problème général de l'encadrement de l'énergie du fondamental d'un hamiltonien local est un problème difficile aussi bien classiquement que quantiquement. Cependant, le point important est qu'il est possible de trouver des schémas d'approximations et des algorithmes pour lesquels le quantique permet un gain significatif.

✂ Algorithmes efficaces pour la chimie quantique

L'étape préliminaire indispensable consiste à transformer le problème de chimie quantique que l'on veut étudier en un problème de qubits. Or tout problème de chimie quantique peut se réécrire comme un problème de fermions itinérants sur un réseau abstrait¹⁶ avec des termes d'interaction à deux corps associés à l'interaction coulombienne entre électrons. Des techniques anciennes [222] ou plus récentes [68] permettent alors de réécrire ce problème comme un problème de qubits couplés, le point important étant de trouver comment minimiser le nombre de qubits nécessaires [328, 350]. L'opérateur hamiltonien s'écrit alors comme un polynôme en les matrices de Pauli agissant sur les différents qubits. Cette étape de réécriture est souvent appelée pré-compilation classique et c'est un préliminaire à la phase de calcul quantique à proprement parler. Pour celle-ci, une première idée consiste à utiliser l'algorithme d'estimation de phase de la section 8.2.3 pour estimer l'énergie de l'état fondamental [22]. Cependant cela nécessite un processeur quantique performant avec un faible taux d'erreurs. Une autre voie plus robuste a donc été explorée.

Cette dernière consiste à utiliser un coprocesseur quantique pour appliquer une méthode variationnelle. Il s'agit d'un algorithme hybride dans lequel on utilise un processeur quantique pour préparer un état parmi une famille d'états variationnels et calculer l'énergie moyenne de cet état. Un ordinateur classique est alors chargé de piloter le processeur quantique pour trouver le minimum de l'énergie au sein de cette famille variationnelle [299, 267]. Même si cet algorithme permet d'éviter d'avoir à calculer l'évolution cohérente, le problème de la préparation efficace des états de la famille variationnelle subsiste. Mais fort heureusement, une méthode standard utilisée en chimie quantique pour préparer des familles variationnelles états quantiques de systèmes de fermions (*unitary coupled cluster ansatz*) peut être mise en œuvre efficacement sur une machine quantique [384] alors qu'elle a une complexité exponentielle sur un ordinateur classique [299]. L'ensemble de ces étapes forme ce qu'on appelle un résolveur variationnel quantique (*variational quantum eigensolver*) dont le principe de fonctionnement est résumé sur la figure 8.18.

16. Les sites du réseau sont les orbitales atomiques et les amplitudes tunnel sont obtenues en calculant leur recouvrement et sont fonction de la géométrie de la molécule.

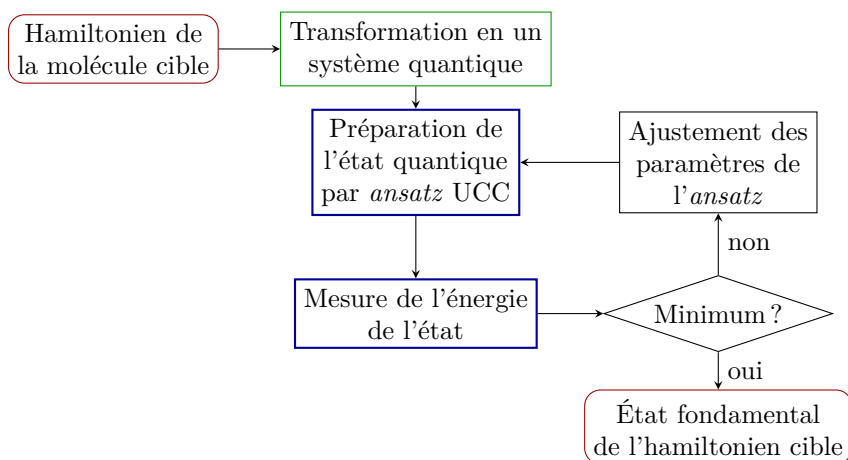


FIGURE 8.18 – Principe du résolveur variationnel quantique. Les étapes quantiques sont les deux boîtes bleues. Les étapes classiques sont les boîtes noires. La phase de précompilation est à gauche.

Les deux algorithmes que nous venons de décrire ont permis le calcul des énergies de molécules simples (H_2 et H_2^+) avec une précision suffisante pour permettre d'obtenir le bon ordre de grandeur des taux de réactions [290]. Plus récemment, le spectre des premiers états excités de la molécule H_2 a été calculé au moyen d'une adaptation de l'algorithme variationnel quantique [106]. Enfin, l'énergie du fondamental de molécules plus complexes comme BeH_2 et de nanoaimants a été calculée au moyen du résolveur variationnel quantique [226]. Cet algorithme a également été appliqué en physique nucléaire [141], ce qui constitue un retour intéressant étant donné que la paramétrisation des états à N corps par l'ansatz *unitary coupled cluster* fut au départ introduite dans ce domaine [102, 103] avant de devenir populaire en chimie. Ces travaux ouvrent la voie à la simulation quantique pour des molécules plus complexes sur lesquelles les méthodes classiques n'aboutiraient pas à des résultats aussi précis même avec les moyens de calcul les plus puissants.

8.5.2 ⚡ Le calcul adiabatique

⚡ Principe général

Le théorème adiabatique [252] nous dit que tout dispositif quantique dont le hamiltonien $H(t)$ dépend du temps de manière contrôlable par l'opérateur peut être utilisé pour préparer l'état fondamental des hamiltoniens continûment reliés à $H_0 = H(t=0)$, sous réserve de faire évoluer lentement les paramètres du système et de disposer d'une préparation du fondamental $|\Psi_0\rangle$ de H_0 . On peut alors imaginer utiliser une évolution adiabatique comme modèle de

calcul simplifié dans lequel on ne s'autorise à effectuer des mesures que lors de la préparation de $|\psi_0\rangle$, c'est-à-dire sur l'état initial et sur l'état final. C'est ce que l'on appelle le modèle du calcul quantique adiabatique [152] dont le développement est passé en revue dans [12].

De manière remarquable, il existe une équivalence de complexité entre ce modèle de calcul et le modèle de la machine de Turing quantique [11] lorsque l'on considère des hamiltoniens locaux. Pour être exploitable, cette méthode nécessite de savoir encoder le problème que l'on veut étudier dans un hamiltonien H_1 qui soit connecté continûment à un hamiltonien H_0 dont on sait préparer efficacement le fondamental. Cette question n'a pas de réponse générale simple. Si on la connaissait, on aurait déjà cerné l'étendue de la classe BQP. Inspiré par la théorie de la complexité, nous pouvons reformuler la question en nous demandant quels sont les hamiltoniens les plus simples qui décrivent les problèmes QMA les plus intéressants, à savoir les problèmes QMA-complets. De manière remarquable, point n'est besoin d'aller chercher des hamiltoniens très complexes : le hamiltonien du modèle d'Ising avec plus proches voisins et champ magnétique transverse devient QMA-complet dès lors qu'on y ajoute un terme de couplage XX, c'est-à-dire de la forme $\sigma_{x,i}\sigma_{x,j}$ entre plus proches voisins [52]. C'est un cas particulier du théorème de dichotomie quantique [113] qui donne la complexité de tous les hamiltoniens couplant les spins au plus par deux. Les hamiltoniens découplables, c'est-à-dire ramenables par transformation unitaire à une somme de hamiltoniens à un spin, sont dans la classe BQP. Les hamiltoniens qui ne peuvent pas être réduits à des hamiltoniens à un site mais que l'on peut localement diagonaliser¹⁷ correspondent à des problèmes NP-complets¹⁸. Les hamiltoniens qui ne font apparaître qu'un couplage ZZ et des couplages à un champ extérieur dans une autre direction appartiennent tous à la même classe intermédiaire qui est celle du modèle d'Ising avec champ extérieur transverse. Enfin, tous les autres sont QMA-complets.

Cela explique que finalement, pour réaliser un ordinateur quantique universel par calcul adiabatique, il suffit de réaliser un seul des hamiltoniens QMA-complets.

✂ Les verres d'Ising

Un cas particulier important consiste à trouver le fondamental d'un hamiltonien 2-local de type Ising qui ne contient que des couplages de type ZZ et des champs locaux selon Z . Cela s'avère être un modèle classique en remplaçant la trace sur l'espace de dimension 2^n par une somme sur les 2^n valeurs $s_i = \pm 1$ de chaque spin. Les couplages J_{ij} entre deux sites $i \neq j$ pouvant prendre des signes différents, le modèle n'est génériquement ni ferromagnétique, ni anti-ferromagnétique. C'est ce qu'on appelle un verre de spins d'Ising qui

17. C'est-à-dire qui ne comprennent que des opérateurs $\sigma_{z,i}$.

18. Cela inclut les hamiltoniens classiques présentant de la frustration, c'est-à-dire des constantes de couplages qui n'ont pas toutes le même signe.

présente un paysage en énergie extrêmement complexe en la configuration $\sigma = (s_i)_i$. De manière générique, trouver l'énergie du fondamental d'un verre de spins d'Ising en champ nul sur un réseau de dimension supérieure ou égale à trois, ou en champ extérieur non nul mais sur un réseau planaire, est un problème NP-dur¹⁹ [29].

Cette machine permet donc de résoudre l'ensemble des problèmes NP, ce qui inclut les problèmes k SAT. En particulier, la correspondance entre problèmes k SAT et hamiltoniens d'Ising est particulièrement simple. Introduisons n variables booléennes $x_i \in \{0, 1\}$ auxquelles on associe les spins classiques $s_i = 2x_i - 1$. À partir d'un problème $\mathfrak{p}$ de satisfiabilité caractérisé par m clauses, on définit la matrice des clauses $C^{(\mathfrak{p})}$ par $C_{ij} = 1$ si la variable x_i apparaît dans la clause j , -1 si c'est la variable $\neg x_i$ et 0 sinon. Le hamiltonien d'Ising correspondant au problème de satisfiabilité considéré est alors

$$H^{(\mathfrak{p})} = \sum_j \frac{1}{2^k} \prod_{i=1}^n \left(1 - C_{ij}^{(\mathfrak{p})} x_j \right), \quad (8.87)$$

dont la valeur, pour une configuration des variables booléennes donnée, est le nombre de causes violées. Le problème $\mathfrak{p}$ de décision k SAT consistant à trouver s'il est possible de satisfaire ou non toutes les clauses, il est donc équivalent à trouver si le fondamental de ce modèle est d'énergie nulle. Le hamiltonien $H^{(\mathfrak{p})}$ couple plus de deux spins entre eux mais il est possible de le ramener à un hamiltonien 2-local en introduisant des spins auxiliaires.

Même si répondre à la question de la nullité de l'énergie du fondamental d'un tel hamiltonien d'Ising est un problème difficile, la physique statistique des systèmes désordonnés permet d'obtenir non pas des résultats sur les pires cas rencontrés (ce qui fait que le problème est NP-complet) mais sur les cas typiques. On peut par exemple évaluer quelle est la fraction des problèmes k SAT qui vérifient toutes les clauses au sein de l'ensemble de toutes les instances possibles et observer comment elle évolue en fonction de la fraction m/n du nombre de clauses par rapport au nombre de spins. Cette très belle physique, qui illustre d'une manière intéressante les relations entre physique (pas forcément quantique), théorie de l'information et complexité est exposée dans l'ouvrage de Mézard et Montanari [270].

⚡ Plus rapide ou pas ?

La question de la rapidité d'un calcul adiabatique se décline en deux questions distinctes. Tout d'abord, il y a la complexité en temps de la réduction du problème considéré à un problème de recherche d'état fondamental. Ensuite, il y a la question de la rapidité avec laquelle on peut effectuer un chemin adiabatique pour trouver le fondamental recherché. En effet, pour être sûr d'atteindre ce fondamental, l'évolution doit être assez lente et cette condition

19. Ici la taille de la donnée est le nombre de sites du graphe considéré.

se formule en termes du *gap*, qui est l'écart d'énergie entre le fondamental et le premier état excité du hamiltonien instantané.

La question de la rapidité de la phase d'évolution du hamiltonien utilisé dans un calcul adiabatique est absolument centrale. Si on trouve un algorithme adiabatique pour un problème NP-complet pour lequel le hamiltonien associé possède un *gap* qui décroît assez lentement avec la taille de l'instance considérée, alors on peut espérer que la recherche du fondamental par évolution adiabatique donne un gain considérable en termes de complexité, voire fasse tomber ce dernier dans une classe quantique polynomiale comme BQP. C'est ce point qui a entraîné un intérêt extrêmement fort lors de la sortie de l'article fondateur du calcul adiabatique [152]. De manière plus précise, il faudrait que le *gap* diminue au plus en n^{-1} où n est la taille de l'instance considérée. Disons-le tout de suite, cet espoir ne s'est pas encore matérialisé mais il est intéressant de comprendre pourquoi.

Tout d'abord, on sait qu'il est assez facile d'avoir un *gap* qui décroît exponentiellement rapidement en la taille n pour certaines instances du problème. Mais cela ne veut pas dire que, pour des instances aléatoires du problème considéré, le *gap* décroisse aussi rapidement. Il est donc important de comprendre quelle peut être la dépendance générique du *gap* en la taille n du système dont on considère une famille à un paramètre de hamiltoniens $\lambda H_n(\lambda)$. C'est un problème compliqué pour les raisons que nous allons maintenant exposer.

Tout d'abord, un résultat de von Neumann et Wigner [286] assure que pour une famille à un paramètre d'opérateurs hermitiens $H_n(\lambda) \neq 0$, on n'observe en général aucune valeur de λ pour lequel le spectre de $H(\lambda)$ est dégénéré. Les exceptions proviennent de l'existence d'une symétrie dans $H_n(\lambda)$ pour une des valeurs de λ . Un mécanisme d'apparition de petits *gaps* bien connu est décrit dans la section 4.1.5 : il s'agit des anti-croisements de niveaux qui apparaissent quand on peut écrire $H(\lambda)$ comme la somme d'un hamiltonien non perturbé présentant des croisements de niveaux auquel on a rajouté une perturbation qui permet à ces niveaux de s'éviter. C'est un mécanisme standard en physique atomique et moléculaire.

Dans le cas du fondamental d'un hamiltonien $H_n(\lambda)$ d'un système complexe, un mécanisme similaire peut aussi être à l'œuvre. Loin des points de croisement évité, le spectre est quasiment celui du hamiltonien non perturbé et c'est la perturbation qui empêche l'apparition des croisements et donc des dégénérescences du fondamental. Le point important est que l'échelle d'énergie $\hbar\Omega$ qui gouverne cet anti-croisement peut être obtenue en prenant l'élément de matrice de l'opérateur intervenant dans la perturbation entre les états propres du hamiltonien non perturbé au voisinage de l'anti-croisement. Plus le recouvrement entre ces deux états sera faible, plus l'énergie $\hbar\Omega$ sera faible. L'idée avancée dans [13] est que, lorsque l'on considère les instances de taille n d'un problème NP-complet, on a affaire à un hamiltonien aléatoire²⁰ qui présente alors un effet connu sous le nom de localisation forte d'Anderson [15, 244].

20. L'aléatoire provient du fait que l'on considère l'ensemble des instances de taille n .

Cet effet a été initialement discuté pour les états d'une particule comme un électron dans un potentiel suffisamment désordonné, c'est-à-dire présentant d'importantes variations aux petites distances²¹. Dans ce cas, Anderson a prédit que la plupart des fonctions propres du hamiltonien ne sont plus des ondes planes ou des ondes de Bloch comme dans un cristal parfait mais des fonctions d'onde localisées qui décroissent exponentiellement sur une distance appelée longueur de localisation. On imagine que cet effet se produit aussi pour des hamiltoniens plus généraux, la localisation des fonctions d'onde ayant lieu dans l'espace paramétrant les différentes configurations, espace dont la taille croît exponentiellement avec n . Ainsi, les éléments de matrice de l'opérateur de perturbation entre deux telles fonctions d'onde différentes sont exponentiellement petits en n . Dans l'article [13], ce raisonnement est développé sur un exemple explicite : les auteurs partent d'un problème NP-complet pour lequel ils discutent une famille à un paramètre de hamiltoniens qui présentent en effet un gap décroissant plus rapidement qu'une exponentielle en la taille de l'instance. Mais, la donnée du problème NP-complet ne détermine pas à elle seule un seul algorithme adiabatique, c'est-à-dire une unique manière d'associer à chaque instance du problème une famille à un paramètre de hamiltoniens $H_N(\lambda)$. Choi [94] a montré que, dans le cas du problème considéré par Altshuler *et al.*, il était possible de trouver une famille de hamiltoniens à un paramètre ne possédant pas la pathologie de localisation forte. Cet argument fut généralisé pour montrer que pour un problème NP-dur bien spécifique, il était possible de trouver de manière efficace des hamiltoniens initiaux et finals tels que l'on échappait à ces croisements évités perturbatifs [136, 135]. En revanche, cela ne signifie pas que ces familles de hamiltoniens ne nécessitent pas un temps d'évolution adiabatique exponentiel en la taille du problème, tout simplement à cause de *gaps* exponentiellement petits d'une autre origine que celle discutée ici.

Ces considérations et les revues récentes [12, 250] montrent que le sujet est loin d'être clos. Néanmoins, il paraît peu probable qu'il soit possible de trouver un problème NP-complet qui bénéficie d'une accélération quantique exponentielle.

21. Dans le cas d'un métal, cela veut dire sur des distances comparables à la longueur d'onde de Fermi des électrons de conduction.

Chapitre 9

Dynamique des systèmes quantiques ouverts

Les règles de la théorie quantique exposées au chapitre 1 sont valides pour des systèmes quantiques fermés. Or, il s'avère que pour comprendre plus finement la physique des expériences modernes mais aussi pour attaquer des problèmes conceptuels majeurs de la théorie quantique, il est nécessaire de comprendre comment décrire des systèmes quantiques ouverts. La prise en compte de l'interaction entre le système et son environnement et l'intrication qui en résulte est en effet riche de conséquences. La conséquence majeure est sans aucun doute l'émergence de comportements « classiques » au sein du monde quantique.

Dans un premier temps, nous allons expliquer comment décrire la dynamique d'un système ouvert en supposant que l'on ne tient pas compte de l'information qui fuit dans l'environnement. Nous verrons que cela conduit à la disparition des cohérences quantiques, un phénomène que l'on appelle la décohérence. Pour aller ensuite plus loin, nous analyserons la dynamique d'un système ouvert en présence de mesures effectuées dans son environnement. Dans le cas où système et environnement sont initialement décorrélés, nous verrons que l'on peut récupérer une notion d'état quantique pur associé à une réalisation unique : c'est la notion de trajectoire quantique que nous aborderons par une approche d'intégrale de chemin en toute généralité.

L'objectif de ce chapitre est avant tout de présenter d'une façon la plus déductive possible les outils pour comprendre les systèmes ouverts. L'analyse des expériences ayant permis d'étudier la décohérence en profondeur est donnée dans le chapitre suivant afin d'insister ici sur la démarche générale. Les notions développées dans ce qui suit sont parmi les plus importantes de la physique quantique moderne, au même titre que l'intrication.

9.1 Les interférences quantiques en pratique

Le cœur du comportement quantique provient du phénomène d'interférence que les expériences d'interférométrie, comme les franges d'Young ou l'interférométrie Mach-Zehnder, illustrent de manière spectaculaire. Cependant, observer les interférences quantiques est une tâche délicate. En effet, la moindre acquisition d'information par un détecteur, ou plus généralement par un environnement, sur le chemin suivi par un système conduit à une atténuation drastique de la visibilité des franges d'interférences : c'est le phénomène de *décohérence*. Regardons comment cela se produit dans un modèle simple. Le but n'est pas pour le moment de présenter la théorie qui permet une comparaison précise avec les expériences mais de montrer les idées générales permettant l'analyse de la dynamique des systèmes quantiques ouverts.

Considérons une expérience de trous d'Young et supposons que l'interaction entre les particules injectées dans l'interféromètre et ce qui joue le rôle de détecteur se produit uniquement lorsque les particules passent au voisinage d'un des deux trous. Initialement, le détecteur est dans un état $|D_0\rangle$. Si la particule suit un chemin qui l'a conduit à passer par le trou 1 (respectivement 2), alors le détecteur bascule dans l'état $|D_1\rangle$ (respectivement $|D_2\rangle$). Considérons alors une base orthonormée $|J\rangle$ de l'espace des états du détecteur. Pour calculer la probabilité de trouver le détecteur dans l'état $|J\rangle$ à l'instant t_f , nous devons considérer le système complet et calculer l'amplitude de transition entre l'état $|s\rangle \otimes |D_0\rangle$ (correspondant à la particule émise par la source et le détecteur préparé dans l'état $|D_0\rangle$) au temps t_i et l'état $|x\rangle \otimes |J\rangle$ (correspondant à la particule détectée en position x et le détecteur passé dans l'état $|J\rangle$) au temps t_f . Les différents états $|J\rangle$ étant orthogonaux, la probabilité que la particule émise par la source soit détectée à la position x est égale à la somme des carrés de ces amplitudes sur les états $|J\rangle$. Si on considère les chemins γ_α qui passent par la fente α , l'amplitude est le produit de l'amplitude due à la dynamique propre de la particule le long de ce chemin multipliée par l'amplitude de détecter l'état $|J\rangle$ dans l'état $|D_\alpha\rangle$. Cela nous donne donc

$$p(S \mapsto x) = \sum_J \left| \sum_{\gamma_\alpha} \mathcal{A}[\gamma_\alpha] \langle J | D_\alpha \rangle \right|^2. \quad (9.1)$$

Nous pouvons alors développer le module au carré sous la forme d'une double somme sur les chemins :

$$p(S \mapsto x) = \sum_{\alpha_+, \alpha_-} \sum_{\gamma_{\alpha_+}, \gamma_{\alpha_-}} \mathcal{A}[\gamma_{\alpha_+}] \mathcal{A}[\gamma_{\alpha_-}]^* \langle D_{\alpha_-} | D_{\alpha_+} \rangle, \quad (9.2)$$

où nous avons utilisé la relation de fermeture $\sum_J \langle D_{\alpha_-} | J \rangle \langle J | D_{\alpha_+} \rangle = \langle D_{\alpha_-} | D_{\alpha_+} \rangle$ pour calculer explicitement la somme sur les états $|J\rangle$. Nous retrouvons là une forme tout à fait similaire à celle que nous obtiendrions pour une particule isolée à l'exception du facteur $\langle D_{\alpha_-} | D_{\alpha_+} \rangle$ qui est potentiellement

non trivial lorsque $\alpha_+ \neq \alpha_-$. Ce facteur, qui est de module inférieur ou égal à l'unité par l'inégalité de Cauchy-Schwarz, atténue les interférences entre les trajectoires passant par des fentes différentes.

Un détecteur qui n'est pas capable de distinguer entre les deux fentes sera envoyé sur deux états physiques identiques, ce qui signifie que $|D_1\rangle$ et $|D_2\rangle$ sont proportionnels. C'est le cas si le détecteur est un gros objet classique dont l'état n'est pas altéré par le passage de la particule. Les états étant normés, cela entraîne que $\langle D_1|D_2\rangle$ est de module unité. Les franges d'interférences sont donc au plus décalées mais leur contraste est maintenu. Cette situation correspond en optique à la propagation au sein d'un matériau transparent : la phase optique est corrigée du fait de l'indice de réfraction du milieu et l'insertion d'une lame transparente dans un interféromètre ne fait que décaler les franges d'interférences. À l'opposé, un détecteur parfaitement capable de distinguer entre les deux fentes donnera lieu à des états orthogonaux après le passage de la particule. Le facteur $\langle D_2|D_1\rangle$ sera alors nul et les termes d'interférences entre trajectoires passant par des fentes différentes sont totalement supprimés.

Le contraste des franges d'interférences est donc fonction des états $|D_\alpha\rangle$, c'est-à-dire des degrés de liberté couplés à ceux de la particule dont on étudie la dynamique. Ces états encodent l'empreinte laissée par la particule le long de son chemin. C'est exactement l'idée qui permet de décrire en toute généralité la dynamique des systèmes quantiques ouverts.

Du côté expérimental, le chemin fut long pour mettre en évidence ces effets de décohérence et mettre en œuvre les expériences de pensées simples élaborées par les fondateurs de la théorie. Des interférences atomiques furent observées au Japon au début des années 1990 grâce au développement des techniques de refroidissement radiatif [334]. L'idée est qu'à très basse température, il est possible que la longueur d'onde de de Broglie $\lambda(T) = h/\sqrt{2\pi mk_B T}$ devienne assez grande pour qu'un ensemble d'atomes piégés par des lasers forme une source thermique cohérente d'ondes de matière capable d'émettre plusieurs milliers de particules¹. Néanmoins, cette expérience remarquable ne permet pas encore d'étudier les effets de décohérence entre les atomes issus de la source et les atomes du gaz résiduel. En effet, la moindre collision dévie suffisamment un atome issu de la source pour qu'il n'arrive plus sur le détecteur. Pour mener des études interférométriques quantitatives, il était donc nécessaire d'employer des particules suffisamment massives par rapport aux atomes du gaz résiduel afin que les collisions n'entraînent qu'un changement d'impulsion le plus faible possible.

Ainsi, au début des années 2000, une expérience d'interférométrie fut réalisée par le groupe d'Anton Zeilinger à Vienne avec des molécules de fullerène [206]. La pièce maîtresse du dispositif expérimental est un interféromètre de Talbot-Lau pour molécules complexes [71] (figure 9.1).

Dans cette expérience, l'interféromètre est placé dans une enceinte à vide dont la pression résiduelle peut être contrôlée. Les données expérimentales pré-

1. À ne pas confondre avec des condensats de Bose-Einstein.

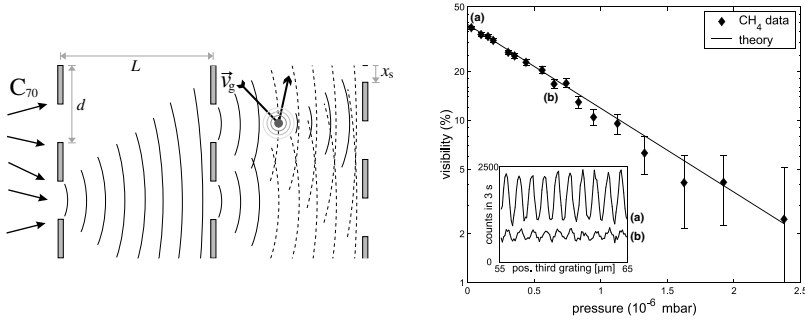


FIGURE 9.1 – À gauche : schéma de principe d'un interféromètre de Talbot-Lau pour des molécules. La première grille diffracte les ondes de matière de telle sorte qu'une onde cohérente atteigne plusieurs fentes de l'écran du milieu. En décalant la troisième grille, on peut ainsi mesurer l'accumulation des atomes après celle-ci en fonction de x_s et visualiser les franges d'interférences. À droite : on observe une décroissance exponentielle du contraste des franges d'interférences en fonction de la pression résiduelle du gaz dans l'enceinte à vide contenant l'interféromètre. La courbe pleine donne la prédiction théorique (sans paramètre ajustable). Reproduit avec l'autorisation de [206], © 2019 by the American Physical Society.

sentées dans la figure 9.1 montrent que le contraste des franges d'interférences dépend de la pression résiduelle du gaz, en l'occurrence des atomes d'argon : il décroît linéairement avec la pression du gaz et une pression très faible suffit à les faire presque totalement disparaître. Il s'agit ici d'un effet de décohérence quantique lié à la présence des molécules de gaz dans l'enceinte. L'ensemble du gaz résiduel agit comme un appareil de mesure qui est sensible à la trajectoire des molécules de fullerène : le passage d'une molécule se traduit par des collisions avec un certain nombre d'atomes d'argon qui enregistrent ainsi une information relative à la trajectoire de la molécule, exactement comme un appareil de mesure.

9.2 Empreintes quantiques et décohérence

Nous allons maintenant développer cette discussion dans un contexte plus général, c'est-à-dire sans supposer que l'interaction entre le système, dont nous étudions les interférences, et l'environnement soit aussi simple. La notion clé sera de nouveau celle d'empreinte laissée par le système dans les degrés de liberté environnementaux.

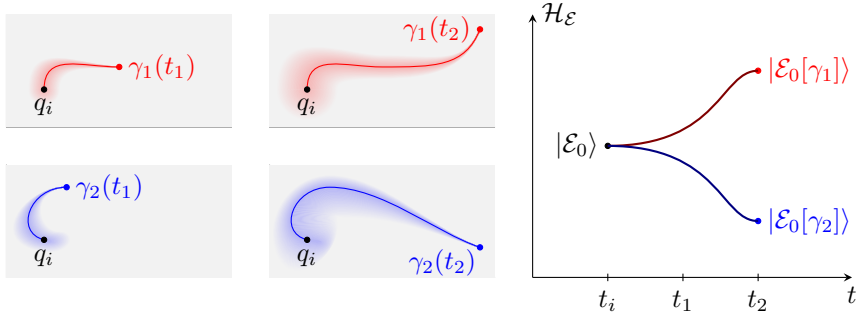


FIGURE 9.2 – Emprunte laissée dans l’environnement par une trajectoire donnée. Pour deux trajectoires différentes, l’état de l’environnement évolue vers deux états différents.

9.2.1 Empreintes et intrication

Considérons un système quantique $\mathcal{S}$ qui est plongé au sein d’un environnement $\mathcal{E}$, que nous allons aussi décrire comme un système quantique à part entière. L’espace des états de ce système composé n’est autre que le produit tensoriel de l’espace des états de chacun des deux sous-systèmes : $\mathcal{H} = \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{E}}$. Lorsque le système $\mathcal{S}$ suit une trajectoire γ bien définie dans son espace de configuration classique, il est vu par son environnement comme un forçage extérieur classique. En conséquence, on peut associer à chaque trajectoire γ de $\mathcal{S}$ un opérateur évolution $U[\gamma]$ qui décrit la dynamique de $\mathcal{E}$ en présence de ce forçage. Imaginons que l’environnement ait été initialement préparé dans un état $|\mathcal{E}_0\rangle$ et que le système $\mathcal{S}$ suive une trajectoire γ . Alors l’état final du système composé est $|\gamma(t_f)\rangle \otimes U[\gamma]|\mathcal{E}_0\rangle$ où $\gamma(t_f)$ désigne la configuration de $\mathcal{S}$ à l’instant final t_f . La figure 9.2 représente cette idée sur deux trajectoires γ_1 et γ_2 . Bien entendu, en termes d’intégrale de chemin, on doit sommer sur tous les chemins γ pour le système considéré. Cela nous montre que, en partant d’une configuration ou d’une position fixée $q_i = \gamma(t_i)$ pour $\mathcal{S}$ et d’un état $|\mathcal{E}_0\rangle$ pour l’environnement, l’état final du système composé est donné par la somme sur les chemins suivante :

$$|\Psi(t_f)\rangle = \sum_{\gamma/\gamma(t_i)=q_i} \mathcal{A}[\gamma] |\gamma(t_f)\rangle \otimes U[\gamma]|\mathcal{E}_0\rangle, \quad (9.3)$$

où $\mathcal{A}[\gamma]$ est l’amplitude de probabilité associée au chemin γ pour le système $\mathcal{S}$ isolé.

L’expression précédente, valable en toute généralité, met en évidence le rôle joué par l’empreinte laissée par le système dans son environnement, à savoir l’état $|\mathcal{E}_0[\gamma]\rangle = U[\gamma]|\mathcal{E}_0\rangle$. Nous pouvons généraliser encore un peu cette expression au cas où l’état initial de $\mathcal{S}$ n’est pas localisé sur une configuration ou une position précise mais est décrit par un état générique $|\varphi_0\rangle$. Partant

d'un état factorisé $|\varphi_0\rangle \otimes |\mathcal{E}_0\rangle$ au temps initial t_i , le système composé évolue donc vers un état *a priori* intriqué de la forme :

$$|\Psi(t_f)\rangle = \sum_{\gamma} \langle \gamma(t_i) | \varphi_0 \rangle \mathcal{A}[\gamma] |\gamma(t_f)\rangle \otimes |\mathcal{E}_0[\gamma]\rangle. \quad (9.4)$$

Cette expression formelle contient toute la physique de la décohérence et nous donnera les clés d'une notion très importante qui est celle de trajectoire quantique. Elle fait le lien entre toutes les approches de la décohérence actuellement utilisées : approches fonctionnelles basées sur le travail de Feynman et Vernon d'une part, approches opératorielles (super-opérateurs de Kraus) et approches par les trajectoires quantiques utilisées en optique quantique.

9.2.2 La fonctionnelle d'influence de Feynman-Vernon

Définition et signification physique

Intéressons-nous au système seul en traçant sur les degrés de liberté environnementaux. L'état du système est alors décrit par un opérateur densité réduit $\rho_S(t)$ obtenu en prenant la trace partielle sur $\mathcal{H}_E$ du projecteur $|\Psi(t)\rangle\langle\Psi(t)|$. Le résultat est l'opérateur densité réduit décrivant le système $\mathcal{S}$ au temps t_f . Il ne correspond pas en général à un état pur mais à un mélange statistique du fait de l'oubli de l'information imprimée dans l'environnement.

En utilisant l'équation (9.4), nous obtenons donc l'opérateur densité réduit sous la forme d'une double intégrale de chemins :

$$\begin{aligned} \rho_S(t_f) = & \sum_{(\gamma_+, \gamma_-)} \langle \gamma_+(t_i) | \rho_S(t_i) | \gamma_-(t_i) \rangle \times \mathcal{A}[\gamma_+] \mathcal{A}^*[\gamma_-] \\ & \times \langle \mathcal{E}_0[\gamma_-] | \mathcal{E}_0[\gamma_+] \rangle |\gamma_+(t_f)\rangle \langle \gamma_-(t_f)|. \end{aligned} \quad (9.5)$$

L'environnement se manifeste alors par l'apparition du produit scalaire $\mathcal{F}[\gamma_+, \gamma_-] = \langle \mathcal{E}_0[\gamma_-] | \mathcal{E}_0[\gamma_+] \rangle$ que l'on appelle la *fonctionnelle d'influence de Feynman-Vernon* [157]. Ici, nous l'avons introduite en supposant que l'environnement était initialement dans un état pur mais il est facile de la généraliser au cas d'un environnement initialement dans un état mélange : c'est la somme pondérée des fonctionnelles correspondant aux différents états purs du mélange.

La donnée de la fonctionnelle d'influence et des amplitudes de probabilité $\mathcal{A}[\gamma]$ pour le système $\mathcal{S}$ isolé permet de calculer comment l'opérateur densité du petit système évolue au cours du temps en décrivant comment la cohérence initiale est propagée au temps t_f au moyen d'une double intégrale de chemin donnée par l'équation (9.5). Cette méthode repose sur la seule hypothèse qu'initialement le système et son environnement sont décorrélés, appelée *approximation de Born*. La fonctionnelle de Feynman-Vernon, dont le module est au plus égal à l'unité, apparaît ainsi comme un facteur de contraste dans l'interférence quantique entre les amplitudes de probabilité associées aux chemins γ_+ et γ_- . Nous retrouvons ainsi une discussion qui généralise celle

présentée dans la section 9.1. Nous voyons que les empreintes associées à chaque chemin jouent exactement le rôle des états $|D_{1,2}\rangle$ dans la discussion simplifiée.

Regardons comment, dans le cas général, elle conduit à la notion de temps de cohérence. Pour deux chemins γ_+ et γ_- , il ne peut y avoir un effet d'interférence que si la fonctionnelle de Feynman-Vernon $\mathcal{F}[\gamma_+, \gamma_-]$ ne décroît pas significativement pour cette paire. Considérons alors une expérience de durée $t_f - t_i$. En l'absence de dissipation, les effets d'interférences quantiques proviendraient de l'interférence entre des paires de chemins qui diffèrent au plus d'une certaine distance $d(t_f - t_i)$ ². Cette distance dépend d'une part de la dynamique propre du système et d'autre part de l'état de départ $|\varphi_0\rangle$ de $\mathcal{S}$. En optique, on parlerait de la cohérence de la source pour caractériser cette dépendance en l'état initial. En présence de dissipation, on peut atteindre une durée T_0 telle que la distance $d(T_0)$ soit supérieure à la distance maximale sur laquelle la fonctionnelle de Feynman-Vernon ne décroît pas. Dans ce cas, les effets d'interférences quantiques cessent d'être visibles au temps $t_f = t_i + T_0$: T_0 est alors appelé le temps de cohérence quantique.

À ce stade, la discussion ci-dessus est encore bien désincarnée. On comprend bien que, de manière générique, la fonctionnelle d'influence puisse revêtir des formes très variées et que, par conséquent, décrire de manière générale l'évolution du système $\mathcal{S}$ soit un problème difficile. Cela étant, nous allons continuer à élaborer une vision globale en abordant le même problème d'un point de vue de l'environnement en se focalisant sur l'empreinte laissée par le système dans son environnement.

9.3 Trajectoires quantiques

9.3.1 Motivations

Dans la section précédente, nous avons considéré uniquement le système sans analyser l'information stockée dans l'environnement. Nous avons vu que la présence de l'environnement entraîne la destruction des interférences quantiques entre les trajectoires du système, ce qui entraîne que l'état de ce dernier n'est plus un état pur mais un mélange statistique d'état. Comme nous le verrons plus tard, ce phénomène de décohérence entraîne que les effets quantiques sont inobservables pour des systèmes complexes. La théorie de la décohérence constitue donc une indication en faveur de l'émergence du comportement classique au sein d'un monde quantique. Elle n'en constitue cependant pas encore une explication complète. En effet, la description en termes d'opérateur densité vise à décrire un ensemble statistique constitué par un grand nombre de réalisations du système, toutes préparées de manière identiques. Mais cela ne permet pas encore de comprendre pleinement le

2. Cette distance est à comprendre dans l'espace abstrait des chemins.

processus de mesure quantique et en particulier le problème de la sélection d'un résultat bien déterminé.

Dans cette section, nous allons plus particulièrement nous intéresser à l'information accessible dans l'environnement lui-même, au fur et à mesure que le système s'intrique avec ce dernier. Les idées introduites au chapitre 5 nous montreront comment émerge l'opérateur densité issu du processus de décohérence discuté précédemment à partir de la statistique des résultats de mesures effectuées dans l'environnement. Cela n'est toujours pas la réponse aux questions fondamentales ci-dessus mais nous verrons comment cela peut nous mettre sur la piste d'une solution. Au-delà des motivations très fondamentales discutées précédemment, c'est une question qui devient très concrète. En effet, avec les récents progrès dans les techniques expérimentales, il est maintenant possible d'effectuer des mesures faibles au sens de la section 5.4 du chapitre 5 sur des systèmes quantiques. Les expériences d'électrodynamique en cavité menées avec des atomes de Rydberg ou avec des nanocircuits supraconducteurs (qubits supra) illustrent particulièrement bien ces développements de la physique quantique mésoscopique.

9.3.2 Les états relatifs

Pour comprendre l'évolution de l'état du système conditionnée à des mesures dans l'environnement, reprenons la discussion de l'état intriqué du système et de son environnement en termes d'états relatifs comme au chapitre 5.

Supposons que nous disposions d'une base orthonormée de l'espace des états de l'environnement associée à la mesure de certaines quantités. Par exemple, dans le cas de l'expérience de Zeilinger présentée au début de ce chapitre, cela pourrait être la base des positions de chaque atome du gaz d'argon. Nous allons noter $|[X]\rangle$ cette base. La notation $[X]$ indexe les résultats de la mesure et est là pour rappeler que génériquement, son spectre n'a aucune raison d'être discret. Nous pouvons alors décomposer l'empreinte laissée dans l'environnement sur cette base et obtenir ainsi une nouvelle expression pour l'état intriqué de $(\mathcal{S})$ et $(\mathcal{E})$:

$$|\Psi(t_f)\rangle = \sum_X \sum_{\gamma} \langle \gamma(t_i) | \varphi_0 \rangle \mathcal{A}[\gamma] \langle [X] | \mathcal{E}_0[\gamma] \rangle |\gamma(t_f)\rangle \otimes |[X]\rangle. \quad (9.6)$$

Cette expression peut alors se réécrire sous la forme suivante :

$$|\Psi(t_f)\rangle = \sum_X |\psi(t_f, [X])\rangle \otimes |[X]\rangle, \quad (9.7)$$

où l'état non normé $|\psi(t_f, [X])\rangle$ est défini par l'intégrale de chemins suivante :

$$|\psi(t_f, [X])\rangle = \sum_{\gamma} \langle \gamma(t_i) | \varphi_0 \rangle \mathcal{A}[\gamma] \langle [X] | \mathcal{E}_0[\gamma] \rangle |\gamma(t_f)\rangle. \quad (9.8)$$

Pour se ramener à des états normalisés, introduisons la quantité $p[t_f, [X]] = \langle \psi(t_f, [X]) | \psi(t_f, [X]) \rangle$ qui n'est autre que la norme au carré de ce vecteur et notons $|\psi(t_f, [X])\rangle$ le vecteur normé tel que

$$|\psi(t_f, [X])\rangle = \sqrt{p[t_f, [X]]} |\psi(t_f|[X])\rangle, \quad (9.9)$$

l'équation ci-dessus devient alors :

$$|\Psi(t_f)\rangle = \sum_X \sqrt{p[t_f, [X]]} |\psi(t_f|[X])\rangle \otimes |[X]\rangle, \quad (9.10)$$

dans laquelle tous les vecteurs qui interviennent sont normalisés. Comme nous l'avons vu au chapitre 5, l'état $|\psi(t_f|[X])\rangle$ est l'état du système relatif au résultat de mesure $[X]$ dans l'environnement. La quantité $p[t_f, [X]]$ est alors la probabilité d'obtenir ce résultat de mesure³. Ce que nous venons de décrire n'est autre qu'une mesure généralisée effectuée sur le système au fil du temps. Au fur et à mesure que le temps t_f croît, le signal mesuré $[X]$ change : si c'est un signal dépendant du temps, on l'acquiert sur une durée plus longue. Si c'est un signal en fréquence, il faut inclure plus de fréquences pour décrire l'information relative à un intervalle de temps plus grand. Dans tous les cas, on peut suivre l'évolution du signal $[X]$ et donc de l'état relatif $|\psi(t_f|[X])\rangle$, qui décrit alors une *trajectoire quantique*⁴ dans l'espace des états $\mathcal{H}_S$.

Remarquons un point crucial : alors même que l'évolution de l'état du système complet ($\mathcal{S} + \mathcal{E}$) incluant les appareils de mesure qui sont dans ($\mathcal{E}$) est déterministe (et même unitaire), l'évolution de l'état relatif $|\psi(t_f|[X])\rangle$ est stochastique. En effet, pour chaque temps t_f le résultat de mesure lui-même est aléatoire (sa probabilité étant donnée par le carré de la norme de $|\psi(t_f, [X])\rangle$) et donc l'état relatif évolue avec une composante stochastique. C'est comme cela que l'on voit les fameux sauts quantiques. Nous verrons dans le prochain chapitre comment cela se décline sur des exemples concrets.

Le formalisme que nous avons développé ici donne une formule d'intégrale de chemins pour l'état $|\psi(t_f, [X])\rangle$ qui est l'état relatif multiplié par la racine carrée de la probabilité d'obtenir le résultat $[X]$ pour la mesure dans ($\mathcal{E}$). Cet état, qui n'est pas normé, possède une interprétation simple. Pour cela, revenons à la notion de probabilité et, plus précisément, aux probabilités conditionnelles dont nous avons parlé dans le chapitre 2. Une probabilité conditionnelle $p(A|B)$ est reliée à une probabilité jointe par la formule :

$$p(A, B) = p(A|B) p(B). \quad (9.11)$$

Compte tenu de la relation entre probabilités et amplitudes en mécanique quantique, la relation (9.9) entre l'état relatif $|\psi(t_f, [X])\rangle$ et l'état $|\psi(t_f|[X])\rangle$

3. Notons qu'il s'agit d'une probabilité conditionnelle : elle est conditionnée à la préparation initiale de l'ensemble système et environnement mais pour simplifier la notation, nous avons omis la condition initiale.

4. On parlera de trajectoire quantique pour le vecteur d'état par opposition aux trajectoires classiques qui apparaissent dans les intégrales de chemin.

apparaît comme l'analogue quantique de la relation (9.11). Il est donc naturel d'appeler état joint l'état non normé $|\psi(t_f, [X])\rangle$.

Le point remarquable dans cette approche à la Feynman est que c'est l'état joint qui possède une expression simple en termes d'intégrale de chemins. Ainsi l'expression (9.8) est exactement la formule d'intégrale de chemins pour un état pur mais avec un poids de Feynman effectif de la forme :

$$\mathcal{A}[\gamma, [X]] = \mathcal{A}[\gamma] \times \langle [X] | \mathcal{E}_0[\gamma] \rangle. \quad (9.12)$$

Cette expression est parfaitement intuitive : elle nous dit que la dynamique en présence de mesures effectuées sur l'environnement s'obtient en corrigeant la dynamique intrinsèque du système ($\mathcal{S}$) par le recouvrement entre l'empreinte qu'il laisse dans son environnement $|\mathcal{E}_0[\gamma]\rangle$ et l'état $|[X]\rangle$ représentant le résultat de la mesure auquel on conditionne. Cette relation exprime que mesurer l'environnement introduit un filtre dans l'intégrale de chemin. Ce filtre, qui n'est autre que l'amplitude $\langle [X] | \mathcal{E}_0[\gamma] \rangle$, comporte non seulement une phase que l'on peut voir comme une correction de la dynamique propre induite par la mesure (une renormalisation de la dynamique propre) mais minore aussi le poids associé à certaines trajectoires dès que $|\langle [X] | \mathcal{E}_0[\gamma] \rangle| < 1$. Nous pouvons encore voir cela comme un analogue de la relation de Bayes (9.11) mais au niveau des amplitudes élémentaires : l'amplitude de probabilité de voir le système suivre la trajectoire γ et d'observer le signal $[X]$ dans l'environnement est le produit de l'amplitude de voir le système suivre γ sous l'effet de sa dynamique propre multiplié par l'amplitude de voir le signal $[X]$ dans l'environnement sachant que le système a suivi γ .

Remarquons que l'équation (9.8) peut se traduire en disant qu'il existe un opérateur $M_{[t_i, t_f]}[X]$ qui donne comment l'état $|\varphi_0\rangle$ est transformé en l'état $|\psi(t_f, [X])\rangle$:

$$|\psi(t_f, [X])\rangle = M_{[t_i, t_f]}[X] |\varphi_0\rangle. \quad (9.13)$$

Cela illustre simplement le fait que nous sommes en train de décrire un processus de mesure généralisé (voir section 5.4 du chapitre 5). Les opérateurs $M_{[t_i, t_f]}[X]$ ne sont en général pas unitaires. En revanche, comme nous l'avons vu au chapitre 5, ils vérifient :

$$\sum_X M_{[t_i, t_f]}[X]^\dagger M_{[t_i, t_f]}[X] = \mathbb{1}. \quad (9.14)$$

Pour faire le lien avec la discussion de la section 9.2 qui donne un mélange statistique, remarquons que, lorsque l'on ignore les résultats de mesure, l'évolution du système est décrite par le mélange des états $|\psi(t_f, [X])\rangle$ avec la

probabilité $p[t_f, [X]]$, c'est-à-dire par l'opérateur densité réduit

$$\rho_S(t_f) = \sum_X p[t_f, [X]] |\psi(t_f|[X])\rangle\langle\psi_S(t_f|[X])| \quad (9.15a)$$

$$= \sum_X |\psi(t_f, [X])\rangle\langle\psi(t_f, [X])| \quad (9.15b)$$

$$= \sum_X M_{[t_i, t_f]}[X] \rho_S(t_i) M_{[t_i, t_f]}[X]^\dagger. \quad (9.15c)$$

Comme nous l'avons vu au chapitre 5, il s'agit d'une représentation de $\rho_S(t_f)$ qui découle du choix de mesure de $[X]$ dans l'environnement. Un autre choix donnerait une décomposition différente sous forme de mélange. De plus, même si les états $|[X]\rangle$ forment une famille orthonormée, il n'en est *a priori* rien pour les états $|\psi(t_f|[X])\rangle$. La décomposition ci-dessus n'est donc pas en général la décomposition de Schmidt de l'état intriqué $|\Psi(t_f)\rangle$. Cela serait le cas si et seulement si on effectuait à chaque instant des mesures fortes sur le système. Comme nous le verrons dans le prochain chapitre, certaines situations expérimentales correspondent à des mesures faibles au sens de la section 5.4.

Nous allons maintenant considérer un cas particulier assez générique qui est celui pour lequel l'environnement n'a pas de mémoire, ce qu'on appelle le cas markovien. Dans ce cas, nous allons trouver une équation d'évolution pour l'opérateur densité réduit $\rho_S(t)$ particulièrement importante, que l'on appelle équation de Lindblad, ainsi que l'équation pour les trajectoires quantiques associées. Nous reviendrons plus en détail sur la notion de markovianité dans la section 9.5.2.

9.4 Le cas markovien

Un environnement *markovien*, ou sans mémoire, est un environnement dont l'action sur le système ne dépend pas de l'histoire antérieure de ce dernier. Pour décrire un tel environnement, nous allons utiliser une modélisation phénoménologique qui consiste à discrétiser le temps en tranches de durées dt et à décrire l'environnement en termes de degrés de liberté associés à chacune des tranches. Sauf cas spécifique comme l'exemple de trains d'atomes (section 10.1), c'est une approximation dont la validité sera discutée dans la section 9.5.

9.4.1 Le modèle d'environnement markovien

Nous allons donc supposer que l'espace des états de notre environnement effectif est un produit tensoriel des espaces des états des degrés de liberté qui le composent comme $\mathcal{H}_{\mathcal{E}} = \bigotimes_i \mathcal{H}_{\mathcal{E}_i}$ que l'on peut ré-indexer en termes du temps t de sorte que l'espace des états des degrés de liberté environnementaux est le produit tensoriel des espaces des états associés aux degrés de liberté qui

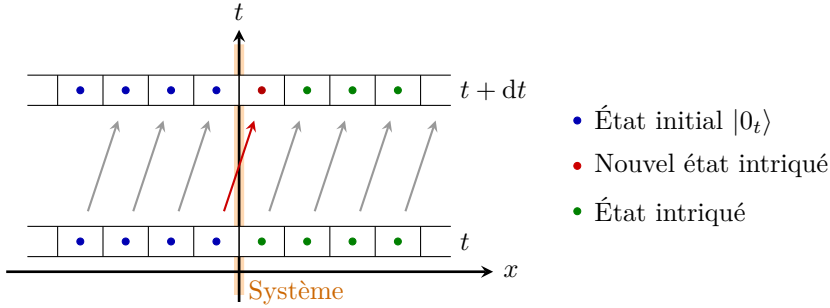


FIGURE 9.3 – Modélisation markovienne de l’environnement. L’interaction entre le système et l’environnement est sans mémoire.

intéragissent avec le système entre t et $t + dt$:

$$\mathcal{H}_{\mathcal{E}} = \bigotimes_t \mathcal{H}_t. \quad (9.16)$$

Une autre façon de penser le même problème est de considérer un environnement fixé qui est réinitialisé à chaque pas de temps. L’état de départ de l’environnement sera pris de la forme

$$|\mathcal{E}_0\rangle = \bigotimes_t |0_t\rangle, \quad (9.17)$$

où chaque état $|0_t\rangle$ représente l’état initial dans la tranche $[t, t+dt]$. Prendre un état $|\mathcal{E}_0\rangle$ invariant par translation dans le temps revient à dire que l’opérateur translation dans le temps d’une durée τ envoie l’état $|0_t\rangle$ sur l’état $|0_{t+\tau}\rangle$ pour tout t .

Lorsque le système suit une trajectoire γ et que nous la considérons sur l’intervalle $[t_i, t_f]$, l’interaction entre le système et son environnement étant locale en temps, cela entraîne que l’état de ce dernier devient alors

$$|\mathcal{E}[\gamma]\rangle = \left(\bigotimes_{t' < t_i} |0_{t'}\rangle \right) \otimes \left(\bigotimes_{t_i \leq t' \leq t_f} U[\gamma_{t'}] |0_{t'}\rangle \right) \otimes \left(\bigotimes_{t' > t_f} |0_{t'}\rangle \right), \quad (9.18)$$

où l’on voit que seuls les états associés aux tranches de temps dans l’intervalle $[t_i, t_f]$ sont altérés. Les opérateurs unitaires $U[\gamma_{t'}]$ correspondent à l’influence du système sur son environnement durant la tranche de temps $[t', t' + dt]$.

Le dernier ingrédient dont nous avons besoin est la définition des mesures effectuées dans l’environnement et donc, de manière équivalente, d’une base orthonormée de chaque espace $\mathcal{H}_t$. Là aussi, il est naturel de prendre la même base au fil du temps. Nous considérerons donc une base $|\mu\rangle$ orthonormée qui

est translatée de $\mathcal{H}_t$ à $\mathcal{H}_{t+\tau}$ au moyen de l'opérateur translation dans le temps. Les vecteurs $|\mu\rangle$ pour $\mu \neq 0$ sont orthogonaux à $|0\rangle$. Les données de mesure effectuées dans l'environnement sont donc la donnée pour chaque tranche temporelle $[t, t + dt]$ d'un état $|\mu_t\rangle$. On notera $[\mu]_t$ la collection des résultats de mesure entre t_i et t . Cette série de données de mesure au cours du temps est égale à la quantité que l'on notait $[X]$ auparavant : on retrouve bien l'idée d'un signal dépendant du temps qui est acquis de manière séquentielle.

Considérons alors la trajectoire quantique décrite par l'équation (9.8). Le facteur de filtrage $\langle [\mu]_t | \mathcal{E}[\gamma] \rangle$ dans l'intégrale de chemins donnant l'état $|\psi(t, [\mu]_t)\rangle$ est donné par

$$\langle [\mu]_t | \mathcal{E}[\gamma_t] \rangle = \prod_{t_i \leq t' \leq t} \langle \mu_{t'} | U[\gamma_{t'}] | 0_{t'} \rangle. \quad (9.19)$$

On voit donc apparaître un produit sur les différents intervalles temporels qui vérifie la même propriété de localité (1.6) que le poids de Feynman $\mathcal{A}[\gamma]$ pour un système isolé. En conséquence, nous en déduisons l'existence d'un opérateur $M_{[t_i, t]}[\mu]$ qui relie le vecteur $|\psi(t_i)\rangle$ au vecteur $|\psi(t, [\mu]_t)\rangle$: $|\psi(t, [\mu]_t)\rangle = M_{[t_i, t]}[\mu] |\psi(t_i)\rangle$. Néanmoins, comme le poids associé à chaque trajectoire n'est pas de module unité, nous ne pouvons pas dire que cet opérateur est unitaire ou, de manière équivalente, que l'état joint $|\psi(t, [\mu]_t)\rangle$ est normé. Par ailleurs, l'opérateur associé à l'intervalle temporel $[t_i, t]$ est le produit d'opérateurs associés à chaque intervalle $[t', t' + dt]$ pour $t' \in [t_i, t]$.

9.4.2 L'équation de Lindblad

Le caractère markovien suggère qu'il est possible de trouver une équation d'évolution pour l'opérateur densité réduit du système. C'est ce que nous allons faire maintenant en analysant plus précisément la forme des opérateurs $M_\mu(t) = M_{[t, t+dt]}[\mu]$ associés à chaque intervalle temporel.

Lorsque la largeur temporelle dt des intervalles temporels devient petite, le facteur de filtrage $\langle 0_t | U[\gamma_t] | 0_t \rangle$ tend vers 1. On s'attend donc, exactement comme dans l'intégrale de chemins pour les systèmes isolés (voir chapitre 1) à ce que l'opérateur $M_0(t)$ associé à l'intervalle $[t, t + dt]$ tende vers l'identité avec une correction linéaire en dt :

$$M_0(t) \simeq \mathbb{1} + L_0(t) dt + O((dt)^2). \quad (9.20)$$

Par ailleurs, nous savons que les opérateurs M_μ obéissent à la relation :

$$\sum_{\mu} M_{\mu}^{\dagger}(t) M_{\mu}(t) = \mathbb{1}. \quad (9.21)$$

En conséquence, cela montre que, dans la limite $dt \rightarrow 0$, les opérateurs $M_\mu(t)$ ont un comportement dominant en $\sqrt{dt} L_\mu(t)$, pour $\mu \neq 0$. En écrivant $L_0(t) = \frac{i}{\hbar} H(t) + \tilde{L}_0(t)$ où H et $\tilde{L}_0$ sont des opérateurs hermitiens, on peut

se servir de la relation (9.21) à l'ordre dt pour relier l'opérateur $L_0(t)$ aux opérateurs $L_\mu(t)$ pour $\mu \neq 0$:

$$L_0(t) = -\frac{1}{2} \sum_{\mu \neq 0} L_\mu^\dagger(t) L_\mu(t) - \frac{i}{\hbar} H(t). \quad (9.22)$$

Nous obtenons donc la forme suivante :

$$M_0(t) = \mathbb{1} - \frac{dt}{2} \sum_{\mu \neq 0} (L_\mu^\dagger L_\mu)(t) - \frac{i}{\hbar} H(t) dt, \quad (9.23a)$$

$$M_{\mu \neq 0}(t) = L_\mu(t) \sqrt{dt}. \quad (9.23b)$$

En utilisant ces expressions et en développant au premier ordre en dt l'équation d'évolution (9.15c), nous obtenons l'équation d'évolution pour l'opérateur densité réduit $\rho_S(t)$ sous la forme suivante, appelée *équation de Lindblad* :

$$\frac{d\rho_S(t)}{dt} = -\frac{i}{\hbar} [H(t), \rho_S(t)] \quad (9.24a)$$

$$- \frac{1}{2} \left\{ \sum_{\mu \neq 0} (L_\mu^\dagger L_\mu)(t), \rho_S(t) \right\} + \sum_{\mu \neq 0} L_\mu^\dagger(t) \rho_S(t) L_\mu(t), \quad (9.24b)$$

où $\{A, B\} = AB + BA$. La première ligne correspond à une équation pour l'opérateur densité d'un système isolé où $H(t)$ serait le hamiltonien dépendant du temps pour (S) . Cet opérateur peut contenir des effets environnementaux en plus de la dynamique de (S) isolé ; on dit que l'environnement renormalise la dynamique propre du système. Ici, c'est ce que l'on obtient lorsque l'interaction entre le système et l'environnement aboutit à ne jamais causer de transitions dans ce dernier : $|0_t\rangle \rightarrow |0_t\rangle$.

Les opérateurs L_μ avec $\mu \neq 0$, appelés *opérateurs de saut* ou *opérateurs de Lindblad*, sont associés à des transitions dans l'environnement : $|0_t\rangle \rightarrow |\mu_t\rangle$. Dans ce cas, l'évolution ne peut plus être décrite en termes hamiltoniens comme le montre la présence de la seconde ligne dans l'équation (9.24). Remarquons que les termes de la première et de la seconde ligne préservent la trace de l'opérateur $\rho_S(t)$.

9.4.3 L'équation de Schrödinger stochastique

Pour compléter la discussion, nous pouvons utiliser le caractère markovien pour écrire l'évolution des états $|\psi(t, [\mu]_t)\rangle$. L'équation stochastique ainsi obtenue est du premier ordre et, surtout, fait intervenir un bruit sans mémoire. Cela permet de la résoudre assez facilement sur ordinateur. On dispose alors d'une méthode alternative efficace pour calculer l'évolution du système ouvert. En effet, si on a affaire à un système dont l'espace des états est de dimension d , la résolution de l'équation de Lindblad porte sur une matrice $d \times d$ alors que

la résolution de l'équation stochastique que nous allons discuter porte sur un vecteur de $\mathbb{C}^d$. Lorsque le système est de grande dimension, la méthode des trajectoires stochastiques quantiques est plus efficace. Par ailleurs, comme nous l'avons souligné à la section 9.3, elle donne accès à la dynamique d'un système quantique unique et joue donc un rôle important, en particulier en ingénierie quantique (voir sections 10.1.4 et 10.2.5).

L'évolution de t à $t + dt$ étant donnée par les opérateurs $M_\mu(t)$, on a :

$$|\psi(t + dt, [\mu]_{t+dt})\rangle = M_\mu(t) |\psi(t, [\mu]_t)\rangle. \quad (9.25)$$

On peut alors décrire cette dynamique au travers d'une équation différentielle linéaire du premier ordre, appelée *équation de Schrödinger stochastique*, en introduisant une stochasticité qui décrit les différents résultats de mesure pour l'intervalle temporel $[t, t + dt]$:

$$d|\psi(t, [\mu]_t)\rangle = \left(-\frac{i}{\hbar} H(t) - \frac{1}{2} \sum_{\mu \neq 0} L_\mu^\dagger L_\mu \right) |\psi(t, [\mu]_t)\rangle dt \quad (9.26a)$$

$$+ \sum_{\mu} (\sqrt{dt} L_\mu - \mathbb{1}) |\psi(t, [\mu]_t)\rangle dN_\mu(t). \quad (9.26b)$$

Le membre de droite de la première ligne représente la partie déterministe de l'évolution du vecteur d'état, dans laquelle on identifie une partie unitaire donnée par le Hamiltonien effectif $H(t)$ et une partie non unitaire donnée par $-\sum_{\mu \neq 0} L_\mu^\dagger L_\mu / 2$. Elle représente l'effet de l'acquisition d'information sur le système en l'absence de transitions dans l'environnement. La seconde ligne contient la partie stochastique qui décrit les sauts quantiques associés aux transitions observées dans l'environnement. La quantité $dN_\mu(t)$ est une variable aléatoire classique valant 0 ou 1. Le cas où elle vaut 0 correspond à l'évolution en l'absence de saut quantique alors que, si $dN_\mu(t) = 1$, un saut quantique s'est produit et $|\psi(t + dt, [\mu]_{t+dt})\rangle = \sqrt{dt} L_\mu |\psi(t, [\mu]_t)\rangle$ et $\mu(t + dt) = \mu$. Pour reproduire la statistique des sauts quantiques, la valeur moyenne et la fluctuation de $dN_\mu(t)$ sont données par :

$$\mathbb{E}(dN_\mu(t)) = p_\mu[t, t + dt] dt, \quad (9.27a)$$

$$\mathbb{E}(dN_\mu(t) dN_\nu(t')) = p_\mu[t, t + dt] \delta_{\mu, \nu} \delta(t - t') dt, \quad (9.27b)$$

avec $p_\mu[t, t + dt]$ la probabilité d'effectuer le saut μ pendant la durée dt autour du temps t . En pratique, lorsque l'on souhaite calculer numériquement les trajectoires quantiques, on travaille souvent avec des états normés qui sont les états relatifs $|\psi(t|[\mu]_t)\rangle$. En effet, la norme de l'état solution de (9.26) évolue au cours du temps et on peut rapidement se retrouver avec des problèmes numériques. Le calcul de l'état relatif $|\psi(t|[\mu]_t)\rangle$ est facile à réaliser au moyen de l'algorithme suivant : on part de l'état $|\psi(t|[\mu]_t)\rangle$ au temps t , et on calcule les probabilités de saut sur l'intervalle $[t, t + dt]$ par la formule $p_\mu[t, t + dt] =$

$\langle L_\mu^\dagger L_\mu \rangle_{|\psi(t)|[\mu]_t} dt$ pour $\mu \neq 0$ et la probabilité qu'il ne se produise par de saut $p_0[t, t + dt] = 1 - \sum_{\mu \neq 0} p_\mu[t, t + dt]$. On tire un nombre aléatoire r entre 0 et 1. Si $r < p_0[t, t + dt]$, il n'y a pas de saut. On calcule alors :

$$|\psi(t + dt)|[\mu]_{t+dt}\rangle = \frac{M_0(t)|\psi(t)|[\mu]_t\rangle}{\sqrt{p_0[t, t + dt]}}. \quad (9.28)$$

Si r est plus grand que $\sum_{\nu=1}^{\mu} p_\nu[t, t + dt]$ mais strictement plus petit que $\sum_{\nu=1}^{\mu+1} p_\nu[t, t + dt]$, on dira que l'on a un saut de type $\mu \neq 0$. On calcule alors :

$$|\psi(t + dt)|[\mu]_{t+dt}\rangle = \frac{L_\mu|\psi(t)|[\mu]_t\rangle}{\sqrt{\langle L_\mu^\dagger L_\mu \rangle_{|\psi(t)|[\mu]_t}}}. \quad (9.29)$$

Puis on recommence le procédé avec ce nouvel état. La difficulté de cette procédure est que l'on peut se retrouver à explorer un grand nombre de trajectoires qui ont un poids statistique de plus en plus faible dans l'opérateur densité réduit $\rho_S(t)$. Il est donc recommandé de conserver la trace de la probabilité individuelle de chaque trajectoire que l'on met à jour d'une itération à la suivante par la formule :

$$p([\mu]_{t+dt}) = p_\mu[t, t + dt] p([\mu]_t), \quad (9.30)$$

où $\mu(t)$ correspond à ce qui est mesuré dans l'environnement pour l'intervalle $[t, t + dt]$. Parfois, il pourra être utile de tronquer les trajectoires si cette probabilité devient trop petite.

9.5 Super-opérateurs et markoviannité

Avant de quitter ce chapitre pour discuter d'exemples concrets, nous allons revenir sur quelques-uns des concepts présentés ici en les abordant sous un angle plus fréquemment employé dans la littérature. On consultera en particulier avec profit le livre de Haroche et Raimond [194] ainsi que l'ouvrage de Breuer et Petruccione [70] qui, lui, développe ces questions de manière plus mathématique.

9.5.1 Le théorème de Kraus

Dans les sections précédentes, nous avons pris le point de vue du physicien, qui consiste à discuter la dynamique du système en présence de son environnement en partant de l'hypothèse que celui-ci imprime une empreinte $|\mathcal{E}[\gamma]\rangle$ dans son environnement le long de chacun des chemins qu'il suit, sachant qu'au départ l'environnement est dans un état pur donné. Dans ce cas, nous avons obtenu une représentation de la dynamique quantique du système sous la forme d'une mesure généralisée (équation (9.15c)) :

$$\rho_S(t_f) = \sum_X M[X] \rho_S(t_i) M[X]^\dagger, \quad (9.31)$$

où les opérateurs $M[X]$ correspondent aux différents résultats de mesures effectuées dans $\mathcal{E}$. Mais on peut se poser le problème sous une forme plus abstraite en se demandant tout d'abord quelles sont les contraintes sur les transformations d'un opérateur densité décrivant l'état d'un système quantique. Puis dans un second temps, on peut se demander quelles sont les transformations les plus générales qui vérifient ces contraintes.

Les super-opérateurs complètement positifs

La réponse à la première question est la suivante : pour une condition initiale factorisée entre un système et son environnement, l'évolution de l'opérateur densité décrivant l'évolution d'un système quantique ouvert est décrite par une transformation $\mathcal{L}_{[t_i, t_f]}$, appelée *super-opérateur*, qui vérifie les conditions suivantes :

1. $\mathcal{L}_{[t_i, t_f]}$ transforme un mélange statistique d'opérateurs densité $\rho = \sum_i p_i \rho_i$ en le mélange des images $\sum_i p_i \mathcal{L}_{[t_i, t_f]}(\rho_i)$;
2. $\mathcal{L}_{[t_i, t_f]}$ envoie les opérateurs hermitiens sur les opérateurs hermitiens ;
3. $\mathcal{L}_{[t_i, t_f]}$ préserve la trace ;
4. considérons un système auxiliaire dont l'espace des états est $\mathcal{H}_{\text{aux}}$. Alors, si $\rho_{\text{tot}}(t_i)$ désigne un opérateur densité sur $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\text{aux}}$, $\text{tr}_{\mathcal{H}_{\text{aux}}} ((\mathbb{1} \otimes \mathcal{L}_{[t_i, t_f]})(\rho_{\text{tot}}(t_i)))$ est un opérateur hermitien dont le spectre est positif.

Seule la dernière propriété, appelée *positivité complète*, n'est pas très intuitive car les trois premières sont ce qu'on attend pour la transformation qui envoie $\rho_{\mathcal{S}}(t_i)$ sur $\rho_{\mathcal{S}}(t_f)$. La positivité complète est une condition plus forte que d'imposer la positivité de l'opérateur $\mathcal{L}_{[t_i, t_f]}(\rho_{\mathcal{S}}(t_i))$. Elle signifie que $\mathcal{L}_{[t_i, t_f]}$ décrit bien l'évolution de l'état de $\mathcal{S}$ y compris en présence d'un système auxiliaire avec lequel il a été intriqué dans le passé mais avec lequel il n'interagit pas durant l'intervalle temporel $[t_i, t_f]$. En revanche, un point crucial est qu'une telle écriture n'existe qu'à cause de l'hypothèse sur la condition initiale factorisée entre le système et l'environnement avec lequel il interagit sur l'intervalle $[t_i, t_f]$. En effet, comme nous le verrons dans la section suivante 9.5.2, si on part d'une situation corrélée entre $\mathcal{S}$ et $\mathcal{E}$, il n'est plus possible de décrire $\rho_{\mathcal{S}}(t_f)$ uniquement à partir de la donnée de $\rho_{\mathcal{S}}(t_i)$ et de $\rho_{\mathcal{E}}(t_i)$ car on oublie alors l'information contenue dans les corrélations quantiques entre système et environnement à t_i .

Le théorème de Kraus

La forme générale des super-opérateurs complètement positifs est donnée par le *théorème de Kraus* suivant :

Théorème 9.1. *Un super-opérateur complètement positif $\mathcal{L}$ agissant sur les matrices densité d'un système dont l'espace des états est de dimension N est*

de la forme

$$\mathcal{L}(\rho) = \sum_{a=1}^{N^2} M_a \rho M_a^\dagger, \quad (9.32)$$

où les opérateurs M_a vérifient⁵

$$\sum_{a=1}^{N^2} M_a^\dagger M_a = \mathbb{1}. \quad (9.33)$$

On trouvera une très élégante démonstration dans le livre de Haroche et Raimond [194] ainsi qu'une démonstration élémentaire en dimension finie dans [51]. Sans détailler cette démonstration, plusieurs remarques permettent de comprendre l'importance et la signification de ce résultat.

Remarquons tout d'abord que le théorème de Kraus est plus fort que la représentation générale (9.31) : en effet, il nous apprend que nous avons besoin d'au plus N^2 générateurs M_a alors que dans la représentation précédente, le nombre de générateurs $M[X]$ pouvait être bien supérieur. On a ainsi la taille de la représentation minimale du processus d'évolution de l'état d'un système ouvert. Dans le cas où les M_a sont de rang 1, ils sont tous de la forme $M_{(i,j)} = |F_j\rangle\langle I_i|$ où $(|F_j\rangle)$ et $(|I_i\rangle)$ sont deux bases orthonormées de $\mathcal{H}_S$. En conséquence, ils correspondent à des transitions pour le système. L'idée que l'on puisse caractériser l'évolution du système au travers d'un tel jeu de transitions paraît physiquement sensée. En termes plus informationnels, le théorème de Kraus nous dit que le processus physique qui implique l'intrication entre notre système et un environnement dont l'espace des états est éventuellement de dimension infinie peut être simulé par l'intrication avec un environnement fictif dont la dimension est au plus N . Cet environnement fictif possède la dimension minimale pour enregistrer l'information sur toutes les transitions qui peuvent se produire au niveau du système S durant l'intervalle $[t_i, t_f]$.

Les multiples déploiements des trajectoires quantiques

Cette discussion suggère que la représentation d'un super-opérateur complètement positif donné sous la forme d'une mesure généralisée n'est pas unique.

En effet, en définissant de nouveaux opérateurs $N_b = \sum_a U_{b,a} M_a$ avec une matrice unitaire U , nous pouvons vérifier aisément que grâce à l'unitarité de U les équations (9.32) et (9.33) sont vérifiées en remplaçant les M_a par les N_b , qui peuvent être bien plus nombreux que les M_a non nuls.

Nous retrouvons là les idées discutées dans la section 5.3.2 du chapitre 5 ainsi que dans la section 9.3 : pour une condition initiale $\rho_S(t_i) = |\psi_S(t_i)\rangle\langle\psi_S(t_i)|$, l'opérateur densité à l'instant final est décrit comme un mélange d'états correspondants aux différentes trajectoires quantiques. Chaque jeu de trajectoire

5. Cette condition assure que la trace est préservée.

correspond à un choix de mesures effectuées dans l'environnement et donc à une mesure généralisée décrite par des générateurs de Kraus. Si on change la mesure de l'environnement, on change les générateurs de Kraus. Choisir une mesure dans l'environnement pour décrire le super-opérateur $\mathcal{L}_{[t_i, t_f]}$ revient à choisir ce qu'on appelle un déploiement de l'évolution de $\mathcal{S}$ sur un jeu de trajectoires quantiques. Nous verrons dans le prochain chapitre comment émergent différents déploiements des trajectoires quantiques dans des dispositifs physiques concrets.

Remarquons que ces considérations sont valables indépendamment de l'hypothèse de markoviannité. Même si le cas markovien revêt une grande importance pratique, l'exploration des systèmes mésoscopiques rend de plus en plus nécessaires de discuter également les situations non markoviennes.

9.5.2 Cas markoviens et non markoviens

Jusqu'à présent, nous avons donné une définition de la markoviannité basée sur l'absence de mémoire de l'environnement. En réalité, la notion de markoviannité nous vient du monde classique et son adaptation au monde quantique est subtile. Diverses définitions, qui ne sont pas forcément équivalentes, ont donc été proposées pour classer une dynamique comme « markovienne » ou « non markovienne ». Afin de comprendre ces subtilités et de justifier la définition que nous avons donnée, nous allons revenir sur cette notion pour les processus stochastiques classiques avant d'aborder le cas quantique. Le lecteur pourra se référer aux revues [314, 69] pour une approche plus exhaustive.

Processus markoviens classiques

Un processus stochastique peut être vu comme l'évolution d'une variable aléatoire au cours du temps. Si l'on s'intéresse à n tranches de temps, indexées par $t_1, \dots, t_n$, le processus est entièrement caractérisé par la probabilité jointe

$$p(x_n, t_n; \dots; x_2, t_2; x_1, t_1) \quad (9.34)$$

qui est celle d'obtenir les résultats $(x_1, \dots, x_n)$ aux instants $(t_1, \dots, t_n)$. On peut alors adopter plusieurs approches afin de décrire la dynamique d'un tel processus.

Une première approche consiste à travailler directement sur les distributions de probabilités. C'est l'approche de type « équation pilote » que nous avons vue plus tôt dans le cas quantique, où l'on travaille sur la matrice densité qui contient l'ensemble des réalisations. Dans ce cas, la description de la dynamique s'effectue à l'aide de la probabilité d'obtenir le résultat x_{n+1} à l'instant t_{n+1} à partir des résultats aux instants précédents. L'application du théorème de Bayes permet alors, en composant les probabilités conditionnelles ainsi que la probabilité au temps initial, d'obtenir la probabilité pour l'ensemble des événements (9.34). Une seconde approche consiste à travailler à l'échelle de la

réalisation unique et à suivre la « trajectoire » de la variable aléatoire au cours du temps. C'est l'approche que nous avons développée précédemment dans le cas quantique avec l'équation de Schrödinger stochastique. La dynamique à l'échelle de la trajectoire unique est aléatoire et la probabilité de sauter en x_{n+1} au temps t_{n+1} est conditionnée au reste des valeurs prises le long de la trajectoire. Dans ce cas-là, pour retrouver les probabilités, il faut échantillonner sur un grand ensemble de réalisations. Dans les deux cas, on peut décrire l'intégralité du processus sous la forme d'un arbre, dans lequel figurent les probabilités conditionnelles des différents événements (figures 9.4 et 9.5).

Dans le cas général, il n'y a bien sûr pas de limites sur les corrélations que l'on peut établir à différents temps. Cela rend la modélisation des processus stochastiques compliquée, car on a besoin d'autant plus d'information sur la dynamique que l'on avance dans le temps. Cependant, il est raisonnable d'imaginer un temps de « mémoire » limité, voire nul. C'est le cas notamment des processus markoviens, pour lesquels les probabilités à l'instant t_n conditionnées au reste de l'histoire suivie par la variable aléatoire, ne dépend en fait que de la valeur prise par la variable aléatoire à l'instant précédent t_{n-1} . Cela se traduit sur les probabilités conditionnelles par :

$$p(x_n, t_n | x_{n-1}, t_{n-1}; \dots; x_1, t_1) = p(x_n, t_n | x_{n-1}, t_{n-1}). \quad (9.35)$$

Dans ce cas, la quantité $p(x_n, t_n | x_{n-1}, t_{n-1})$ donne l'intégralité de la dynamique et peut être vue comme une matrice de transition entre les instants t_{n-1} et t_n . Cela implique notamment que la matrice de transition donne accès non seulement à la probabilité de faire une observation à un temps donné, mais également aux corrélations entre différents instants.

Revenons au cas d'un processus stochastique général. On peut s'interroger par exemple sur le cas d'un observateur qui n'aurait pas accès aux corrélations entre différents instants, mais seulement aux distributions de probabilité à un temps donné. En physique classique, où l'on peut effectuer des mesures sans altérer les probabilités futures du système mesuré, cela peut sembler un choix arbitraire. Mais nous verrons que si l'on cherche à définir la notion de dynamique markovienne en physique quantique, il est impossible de garder l'ensemble des corrélations simplement car mesurer le système modifie sa dynamique et donc les corrélations à différents temps.

On peut alors se demander comment on différencie un processus non markovien d'un processus markovien, à partir des probabilités à un temps. Naturellement, la description de la dynamique markovienne est locale en temps, au moyen de la matrice de transition $T_{t_n \rightarrow t_{n+1}}$, dont les éléments de matrice sont donnés par les probabilités conditionnelles entre un instant et le suivant.

Cette matrice satisfait les trois propriétés suivantes :

$$\sum_{x_{n+1}} T(x_n, t_n \rightarrow t_{n+1}, x_{n+1}) = 1, \quad (9.36a)$$

$$T_{t_n \rightarrow t_{n+1}} \times T_{t_{n+1} \rightarrow t_{n+2}} = T_{t_n \rightarrow t_{n+2}}, \quad (9.36b)$$

$$T(x_n, t_n \rightarrow x_{n+1}, t_{n+1}) \geq 0. \quad (9.36c)$$

La propriété (9.36a) assure que les probabilités restent normalisées. La propriété (9.36b), quant à elle, assure que l'on peut composer les évolutions locales, de proche en proche, de façon à avoir l'évolution de la probabilité à un temps. Enfin, la condition (9.36c) permet d'interpréter les éléments de la matrice de transition comme des probabilités conditionnelles. Cela assure notamment que les probabilités restent positives, même si cette propriété n'est pas nécessaire pour cela.

Une condition suffisante pour avoir une dynamique locale en temps est que la matrice donnée par $p(x_n, t_n | x_1, t_1)$ soit inversible. Cela signifie que l'on peut parcourir la dynamique à rebrousse temps, et ainsi avoir accès aux probabilités initiales en partant des probabilités finales. Cela n'est évidemment pas toujours le cas : un processus d'effacement par exemple, sera irréversible. Néanmoins, si l'on tire les probabilités au hasard, on pourra presque toujours effectuer ce processus. Dans ce cas-là, on peut définir une matrice de transition entre t_n et t_{n+1} en utilisant les probabilités conditionnelles, et en la définissant comme la composition d'une évolution à rebrousse temps, depuis t_n vers t_1 et une évolution normale de t_1 à t_{n+1} :

$$T(x_n, t_n \rightarrow x_{n+1}, t_{n+1}) = \sum_{x_1} p(x_{n+1}, t_{n+1} | x_1, t_1) P^{-1}(x_n, t_n | x_1, t_1). \quad (9.37)$$

On a alors une façon locale en temps d'écrire la dynamique. Cela implique notamment que les propriétés (9.36a) et (9.36b) sont valides. Cependant, la propriété (9.36c) ne l'est pas toujours : il est possible d'avoir des taux de transition négatifs. Cela vient notamment du fait que l'on s'attend à avoir, dans la plupart des cas, des taux de transitions négatifs pour la dynamique renversée. Cela est illustré par l'exemple dans la figure 9.4, dans lequel la variable aléatoire peut prendre les valeurs 0 ou 1, sur les temps t_1, t_2 et t_3 . Ici, le processus est non markovien, car les probabilités conditionnelles en t_3 dépendent à la fois de l'état en t_1 et en t_2 . Dans ce cas, il est néanmoins possible de calculer une matrice de transition. Cependant, celle-ci ne reproduit pas la propriété (9.36c), car on a $T(1, t_2 \rightarrow 0, t_3) = -1/2 < 0$. Cela empêche donc d'interpréter les taux de transitions comme des probabilités conditionnelles. Ce cas illustre également une différence importante entre le point de vue donné par l'équation pilote et celui donné par les trajectoires stochastiques. En effet, s'il est possible de donner une expression de la probabilité à un temps de façon locale, les trajectoires, elles, évoluent de façon non locale en temps. Notamment, ce dernier formalisme semble plus intuitif pour discuter la notion de divisibilité ou de markovianité.

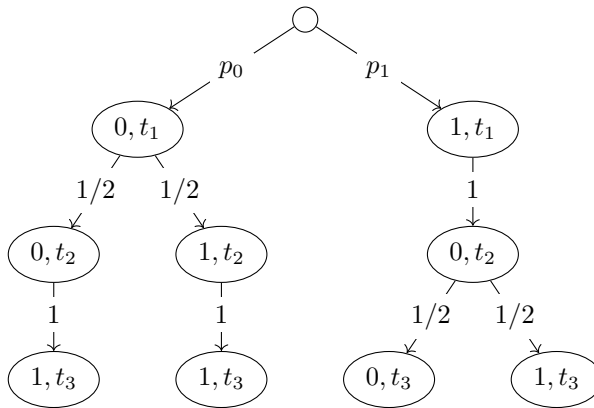


FIGURE 9.4 – Exemple de processus non markovien dans lequel les probabilités à un temps peuvent être décrites par une matrice de transition.

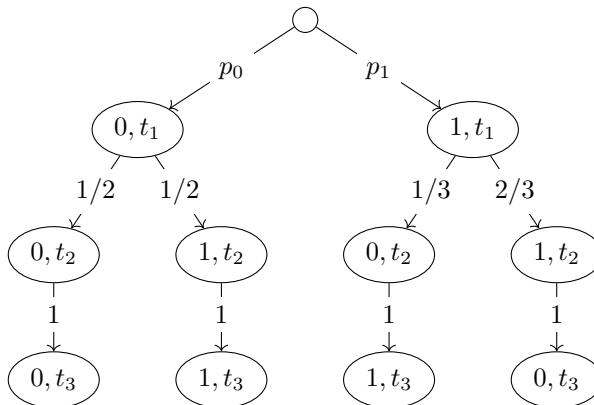


FIGURE 9.5 – Exemple d'un processus non markovien divisible. Dans ce cas-ci, même si le processus est non markovien, il existe un processus markovien qui redonne les mêmes probabilités à un temps. Un observateur qui n'aurait pas accès aux corrélations entre plusieurs instants ne pourrait donc pas distinguer celui-ci d'un processus markovien.

En revanche, si la propriété (9.36c) est satisfaite, on peut interpréter les éléments de la matrice de transition comme des probabilités conditionnelles. Dans ce cas, on parle de processus divisible. La différence entre un processus divisible et un processus markovien est parfaitement invisible aux yeux d'un observateur qui n'aurait accès qu'à des probabilités à un temps. En revanche, il est tout à fait possible d'avoir un processus non markovien qui est divisible, comme c'est le cas sur l'exemple montré sur la figure 9.5.

Ainsi, la bonne notion pour un observateur n'ayant accès qu'aux probabilités à un temps n'est pas celle de markoviannité mais celle de divisibilité. Bien sûr, si le processus n'est pas divisible, l'observateur en déduira qu'il n'est pas markovien. Mais s'il l'est, il lui est parfaitement impossible de distinguer un processus markovien d'un processus non markovien.

Markoviannité en physique quantique

La notion de markoviannité en classique s'applique aux probabilités d'obtenir un événement particulier. Une transcription directe de cette notion en physique quantique implique donc de spécifier les mesures effectuées dans l'environnement. La difficulté est que, contrairement au cas classique, la mesure influence drastiquement la dynamique. En fait, selon ce que l'on mesure, on peut même altérer le caractère markovien. Prenons l'exemple de deux qubits, que l'on prépare dans les états factorisés $|0, 0\rangle$ ou $|1, 0\rangle$ en t_1 . L'interaction est telle que les états des deux spins sont échangés en t_2 , et qu'ils sont ré-échangés en t_3 . Un observateur observant le premier spin dans la base $(|0\rangle, |1\rangle)$ pour ces trois temps verrait soit l'histoire $|0\rangle \rightarrow |0\rangle \rightarrow |0\rangle$, soit $|1\rangle \rightarrow |0\rangle \rightarrow |1\rangle$, ce qui est clairement une signature de non-markoviannité, peu surprenante étant donné la dynamique. En revanche, s'il entreprend de mesurer le spin dans la base $(|+\rangle, |-\rangle)$ en t_3 , alors on a affaire à un processus markovien.

La définition d'un processus markovien en physique quantique est donc nécessairement différente de celle de la physique classique. Dans la section 9.4, nous avons adopté une approche basée sur les trajectoires, dans laquelle on peut découper l'environnement en tranches indépendantes. Cette approche a l'avantage de coller au plus proche de la notion intuitive de markoviannité. L'inconvénient principal de cette définition est qu'elle nécessite de modéliser entièrement la dynamique entre le système et l'environnement ce qui est *a priori* un problème très complexe. Cependant, dans le cas où le couplage entre le système et son environnement est assez faible, nous verrons comment formuler un critère de validité de l'approche markovienne effective présentée précédemment.

Une autre voie consiste à donner une définition qui se base sur la dynamique réduite du système seule. Même en posant cette restriction, plusieurs approches s'opposent et ne mènent pas nécessairement aux mêmes résultats. Nous allons donner un aperçu de quelques-unes des définitions et faire le lien avec l'approche précédente (section 9.4). Pour cela, nous allons devoir transposer la notion de divisibilité au cas quantique. Dans le cas classique, cette notion repose sur les propriétés de l'opérateur décrivant l'évolution des probabilités entre deux instants arbitraires. Notamment, celui-ci doit pouvoir s'interpréter comme un arrangement de probabilités conditionnelles. En physique quantique, l'objet avec lequel nous allons travailler est le super-opérateur qui décrit l'évolution de la matrice densité.

9.5.3 Conditions de validité de l'approche markovienne

Commençons par examiner comment formaliser l'idée d'un temps de mémoire court et donc quelles sont les conditions qui font que l'approche présentée dans la section 9.4 est valable dans une situation physiquement réaliste.

Le point de départ est une modélisation du système S et de son environnement E avec la donnée du hamiltonien décrivant la dynamique. On doit également spécifier une condition initiale. Deux choix naturels sont alors possibles :

- On peut supposer que les deux ont évolué jusqu'à se thermaliser en un état d'équilibre pour SE en interaction. Cette condition est évidemment particulièrement adaptée au cas d'un environnement fortement couplé au système. À l'instant $t = 0$, on initialise alors le système au moyen d'un opérateur agissant uniquement sur lui. Dans cette description, le système et l'environnement sont corrélés du fait de leur interaction avant l'initialisation.
- On peut partir d'une condition dite factorisée (ou de Born) dans laquelle seul l'environnement est thermalisé et le système initialisé dans un état initial $\rho_S(0)$. L'opérateur densité complet initial est alors $\rho_{SE}(0) = \bar{\rho}_E \otimes \rho_S(0)$ où $\bar{\rho}_E$ représente l'état d'équilibre de l'environnement. Cette description suppose qu'il n'y a pas de corrélations entre le système et l'environnement et qu'elles n'apparaissent qu'après initialisation du système. Cette hypothèse de factorisation n'est donc pertinente que dans les situations où le couplage entre système et environnement est assez faible pour qu'on puisse négliger les corrélations avant initialisation.

L'approche présentée dans les sections précédentes suppose une condition initiale factorisée. À partir de là, l'opérateur densité $\rho_{SE}(t)$ pourra être décomposé de manière à faire apparaître les corrélations classiques et quantiques entre système et environnement comme

$$\rho_{SE}(t) = \rho_S(t) \otimes (\bar{\rho}_E + \delta\rho_E(t)) + \delta\rho_{SE}(t), \quad (9.38)$$

où $\delta\rho_E(t)$ décrit l'écart à l'équilibre de l'environnement et $\delta\rho_{SE}(t)$ son intrication avec le système. L'idée est de considérer des situations physiques où $\delta\rho_E(t)$ et $\delta\rho_{SE}(t)$ ont des temps caractéristiques de relaxation τ_c qui sont courts devant le temps caractéristique d'observation τ . Cela revient à dire que l'environnement ne s'écarte de son état d'équilibre qu'entre t et $t + \tau_c$, c'est-à-dire sur une durée courte devant le temps τ . Dans ce cas, tout se passe comme si l'environnement était réinitialisé à son état d'équilibre sur chaque pas de temps Δt qui est petit devant le temps d'observation τ mais grand devant le temps de mémoire τ_c . C'est dans ce régime que l'approximation markovienne, parfois aussi appelée *approximation de Born-Markov*, reste valable.

Lors d'une étape d'interaction, le couplage entre le système et l'environnement induit une phase $V\tau_c/\hbar$ où on a noté V l'ordre de grandeur de cette interaction. Ce point est essentiel et non trivial. En effet, on s'attendrait naïvement à ce que la phase induite dépende de t comme $Vt/\hbar$ et non de τ_c . Ceci

ne se produit pas ici grâce à la grande complexité de l'environnement composé d'un très grand nombre de modes. Cet effet, nommé rétrécissement par le mouvement, peut se montrer plus rigoureusement en partant des équations de la dynamique exacte.

Pendant t/τ_c étapes, la phase totale accumulée va donc diffuser aléatoirement puis s'étaler de $(\Delta\phi(t))^2 = t/T_r$ avec $T_r = \hbar^2/V^2\tau_c$. Ce temps est celui nécessaire pour que l'environnement entraîne une évolution significative de l'état du système. La condition de Markov, ou de mémoire courte, signifie que le temps de mémoire est très court devant l'échelle de temps des phénomènes de relaxation et de décohérence de (S) induits par l'environnement, soit $\tau_c \ll T_r$. La condition $V\tau_c/\hbar \ll 1$ à laquelle cela conduit signifie que le couplage entre S et E est assez faible pour n'induire qu'une petite phase pendant le temps de mémoire de l'environnement. Ceci nous fournit un critère explicite pour considérer que l'approche de la section 9.4 est bien valable :

$$\tau_c \ll \tau \ll T_r \text{ ainsi que } \rho_{SE}(0) = \rho_S(0) \otimes \bar{\rho}_E. \quad (9.39)$$

Ce critère est celui utilisé en pratique pour tester si une approche markovienne est adaptée. Dans de très nombreux systèmes où est présente une nette séparation des échelles, il sera satisfait. Cependant, il existe deux familles de situations physiques où cette approximation ne sera pas justifiée. Il s'agit tout d'abord des systèmes assez fortement couplés pour que les corrélations avec le système invalident nos hypothèses sur l'état de l'environnement. C'est parfois le cas en physique de la matière condensée, notamment dans les problèmes d'impuretés quantiques comme le problème Kondo d'une impureté magnétique couplée au spin des électrons d'une bande de conduction dans un métal. De nombreuses techniques furent développées pour traiter ces problèmes [376]. Sortent également du cadre markovien les environnements présentant une large distribution de temps caractéristiques. Là encore, cette situation se rencontre souvent en physique de la matière condensée, la plupart des systèmes présentant un bruit avec une distribution spectrale en $1/\omega$ qui provient de défauts structuraux aux énergies et couplages distribués sur une large plage [143].

9.5.4 Vers une équation de type Lindblad

Au-delà de ce critère portant sur la modélisation du système et sur l'idée de temps de mémoire court, il est intéressant de rechercher des formes générales de dynamiques du système pouvant correspondre à une dynamique markovienne en généralisant au cas quantique la notion de divisibilité.

Nous allons donc nous intéresser aux dynamiques pour lesquelles l'évolution de l'opérateur densité se décrit de manière locale en temps. Pour un temps discret, cela revient à dire que l'on peut calculer $\rho(t_{n+1})$ en fonction de $\rho(t_n)$. Dans le cas d'un temps continu, cela veut dire qu'il existe une équation différentielle pour ρ .

Dans le cas général, l'opérateur densité décrivant le système au cours du temps $\rho(t)$ peut être obtenu en appliquant le super-opérateur $\mathcal{L}_{[t_i, t]}$ sur l'état initial $\rho(t_i)$. Afin de décrire une évolution physique, le super-opérateur $\mathcal{L}_{[t_i, t]}$ doit être linéaire, préserver l'hermiticité, la trace et être complètement positif. En revanche, comme pour les processus stochastiques classiques, il n'est *a priori* pas possible de chercher une évolution partant d'un temps $t > t_i$. Il n'est même pas dit que l'opérateur $\mathcal{L}_{[t, t']}$ existe pour $t \neq t_i$. Si l'on imagine par exemple un cas extrême dans lequel la totalité de l'information sur le système à l'instant t_i est déplacée dans l'environnement à l'instant t , avant de réapparaître à l'instant t' , on ne peut rien prédire sur l'évolution du système à partir de l'instant t .

Écrire une équation locale en temps pour la matrice densité nécessite l'existence d'un opérateur $\mathcal{L}_{[t, t']}$. Tout comme dans le cas classique, une condition suffisante pour cela est que l'inverse de l'opérateur $\mathcal{L}_{[t_i, t']}$ existe. Dans ce cas, on peut écrire le super-opérateur qui donne la matrice densité en t' en fonction de la matrice densité en t comme

$$\mathcal{L}_{[t, t']} = \mathcal{L}_{[t_i, t']} (\mathcal{L}_{[t_i, t]})^{-1}. \quad (9.40)$$

Ce n'est pas parce que le super-opérateur $\mathcal{L}_{[t, t']}$ existe qu'il décrit une évolution physique partant d'états factorisés. Celui-ci est linéaire, préserve l'hermiticité ainsi que la trace mais il n'est pas nécessairement complètement positif, ni même positif. En d'autres termes, on peut se servir de $\mathcal{L}_{[t_i, t]}^{-1}$ pour calculer $\rho(t_i)$ à partir de la donnée d'un $\rho(t)$, mais ce n'est que le résultat d'un calcul, la dynamique renversée n'existant pas en tant que telle. Le point remarquable est que si l'on est capable d'écrire une équation locale pour la matrice densité, celle-ci possède alors une forme analogue à celle de Lindblad [92, 333] :

$$\frac{d\rho_S(t)}{dt} = -\frac{i}{\hbar} [H(t), \rho_S(t)] \quad (9.41a)$$

$$+ \sum_i \gamma_i(t) \left(A_i^\dagger(t) \rho_S(t) A_i(t) - \frac{1}{2} \left\{ (A_i^\dagger A_i)(t), \rho_S(t) \right\} \right). \quad (9.41b)$$

Notons que si on a $\gamma_i(t) \geq 0$, on retrouve l'équation de Lindblad, en posant $L_i(t) = \sqrt{\gamma_i(t)} A_i(t)$. Néanmoins, *a priori*, il n'y a pas une telle contrainte sur les $\gamma_i(t)$. Une question se pose de savoir à quelle condition cette forme décrit une dynamique physique, de sorte que la dynamique intégrée redonne bien un super-opérateur complètement positif préservant la trace. Dans le cas général, la question est toujours ouverte. Cependant, si $\mathcal{L}_{[t, t']}$ possède des propriétés supplémentaires, telles que la positivité ou la positivité complète, il est possible de donner des conditions sur les γ_i et les A_i qui assurent que l'équation provient d'une dynamique physique. Enfin, quel serait l'analogue quantique de la notion de divisibilité, et comment cela se traduit-il sur ces quantités ?

✱ Cas où $\mathcal{L}_{t,t'}$ est complètement positif

Le cas le plus simple est sans doute celui dans lequel le super-opérateur $\mathcal{L}_{t,t'}$ est complètement positif. On parlera alors de dynamique CP-divisible. La dynamique peut alors se découper par tranche de temps. On est dans ce cas-là si et seulement si les $\gamma_i(t)$ sont tous positifs. On retrouve alors l'équation de Lindblad. Si l'on se contente d'observer la dynamique du système réduit, on a donc quelque chose d'équivalent à la notion de markovianité que nous avons développée plus tôt, qui redonnait également l'équation de Lindblad.

Étant donné que l'évolution entre deux tranches de temps est donnée par un opérateur complètement positif, il est toujours possible de voir l'évolution entre t et $t + dt$ comme une collision entre une tranche de l'environnement et le système. Si l'on considère uniquement la tranche $[t, t + dt]$, on peut tout à fait imaginer le système dans un état pur à l'instant initial. La dynamique a tendance à transformer cet état pur en un mélange statistique décrit par un opérateur densité en $t + dt$. Néanmoins, l'opérateur étant complètement positif, on peut écrire sa décomposition de Kraus, et interpréter l'ensemble des opérateurs de Kraus comme une mesure généralisée. On retrouve la modélisation markovienne initialement présentée où l'on décrit mathématiquement l'évolution entre t et $t + dt$ comme une évolution unitaire qui transforme un état factorisé entre le système et une tranche d'environnement fictif vers un état intriqué. Les différentes tranches arrivent toujours vierges et sont indépendantes. Dans le cas où à l'instant t on a un état mélange, il suffit de le décomposer sur les états purs. Cette décomposition mathématique n'est rien d'autre qu'un modèle où à chaque pas de temps, on a une collision, sans mémoire, entre le système et une tranche vierge d'environnement. Cela justifie notamment la définition que l'on a donnée précédemment d'une dynamique markovienne.

À partir de ce modèle collisionnel, il est également possible de remonter aux trajectoires stochastiques. Un dépliement⁶ des trajectoires s'obtient en choisissant une mesure complète sur chaque tranche indépendante de l'environnement fictif. On a alors, pour chaque enregistrement opéré dans l'environnement, un état pur pour le système, qui n'est rien d'autre que la trajectoire stochastique. Comme les tranches de l'environnement sont indépendantes, on peut déduire l'évolution au pas de temps suivant uniquement avec la donnée de l'état à l'instant présent. Tout comme pour le cas classique, il semble impossible de donner un tel algorithme, sans mémoire, au niveau des trajectoires, lorsque la dynamique n'est pas CP-divisible. Cela est un argument fort pour identifier la markovianité à la CP-divisibilité.

Notons cependant que la notion de CP-divisibilité n'est pas strictement équivalente à celle que nous avons développée précédemment. S'il est possible de trouver une dynamique collisionnelle pour le couple système-environnement, cela n'est qu'une décomposition mathématique. Rien ne dit qu'en réalité la

6. *Unraveling* dans la littérature.

dynamique se décompose comme telle. Il est en effet possible de trouver des dynamiques complètement positives alors que l'état initial contient de l'intrication [77, 73]. On peut alors imaginer des cas dans lesquels la dynamique unitaire de l'ensemble génère à des temps différents des états qui sont différents de ceux issus d'un modèle collisionnel, mais dont on peut trouver une dynamique effective, collisionnelle, qui reproduit la même dynamique réduite pour le système. Si l'on n'a accès qu'à l'évolution de la matrice densité, il est donc impossible de faire la distinction entre les deux cas.

✱ D'autres définitions de la markoviannité

Il est également possible d'utiliser des notions informationnelles pour caractériser la markoviannité. Nous avons vu dans la section 6.5.2 qu'il était possible de caractériser la discernabilité de deux états ρ_1 et ρ_2 émis respectivement avec la probabilité p_1 et p_2 par la probabilité de se tromper sur l'état estimé. Nous avons vu avec l'équation (6.39) que celle-ci était donnée par la quantité $(1 - \|\Delta\|_1)/2$, où $\Delta = p_1\rho_1 - p_2\rho_2$ est la matrice de Helstrom. Un critère possible pour la markoviannité peut donc être l'augmentation de l'indiscernabilité au cours du temps pour n'importe quel couple d'états. Cette propriété est motivée par le fait que l'on souhaite que l'information contenue dans le système fuie dans l'environnement sans jamais pouvoir revenir. Cela ressemble notamment au cas que nous avons vu dans la section 2.3.5, dans laquelle une dynamique stochastique markovienne contracte la divergence de Kullback-Leibler. Ici, augmenter l'indiscernabilité de tous les couples d'états revient donc à contracter la matrice de Helstrom.

On peut montrer qu'un super-opérateur $\mathcal{L}$ est positif si et seulement si il contracte la matrice de Helstrom :

$$\|\mathcal{L}[\Delta]\|_1 \leq \|\Delta\|. \quad (9.42)$$

On a donc une dynamique contractante si et seulement si elle est P-divisible. C'est notamment l'approche suivie dans [69] pour définir la notion de markoviannité. Étant donné qu'une dynamique P-divisible est également CP-divisible, c'est une notion moins contraignante que celle définie précédemment.

Il est également possible de caractériser la P-divisibilité au niveau de l'équation de Lindblad. Pour cela, les $\gamma_i(t)$ et les $A_i(t)$ doivent satisfaire la condition suivante :

$$\sum_i \gamma_i(t) |\langle n | A_i(t) | m \rangle|^2 \geq 0 \quad (9.43)$$

pour n'importe quel couple d'états orthogonaux ($|m\rangle, |n\rangle$). Certains $\gamma_i(t)$ peuvent donc devenir négatifs, rendant impossible l'application d'un algorithme de type « trajectoire ».

Notons enfin que, si la dynamique est P-divisible mais pas CP-divisible, il est possible d'observer un retour de la discernabilité si l'on intrique le système physique avec un système auxiliaire qui n'évolue pas. Cela provient du fait

que dans ce cas-là, la dynamique de ces deux systèmes n'est plus positive, et qu'il existe donc des états pour lesquels la probabilité d'erreur ne fait pas qu'augmenter.

Chapitre 10

Vers une ingénierie quantique

Les développements spectaculaires des techniques expérimentales ont permis de faire de nombreuses expériences sur des systèmes quantiques simples. Sachant maintenant les préparer et les manipuler de manière contrôlée, les physiciens ont pu réaliser les expériences de pensées élaborées par les fondateurs de la théorie quantique. Ces systèmes permettent ainsi de tester de nombreux aspects des fondements de la théorie quantique comme la physique de l'intrication, de la mesure, la dynamique des systèmes quantiques ouverts et la décohérence. Ils ont également ouvert tout un champ de recherche appelé ingénierie quantique ou, plus généralement lorsque l'on inclut la dimension applicative, technologies quantiques.

Nous allons aborder deux choses dans ce chapitre. La première concerne la réalisation en laboratoire de qubits et d'oscillateurs harmoniques pouvant être manipulés de manière contrôlée (chapitre 4) dans le contexte de l'électrodynamique en cavité et des qubits supraconducteurs. Le second objectif va être, à travers la physique de ces systèmes simples, d'illustrer l'ensemble des concepts développés lors de l'étude des systèmes quantiques ouverts au chapitre 9. Ceci permettra de poser des jalons utiles pour éclairer les fondements de la théorie quantique.

10.1 Atomes en cavité

Les expériences d'*électrodynamique en cavité* menées au sein du groupe de Serge Haroche à l'École normale supérieure puis au Collège de France à Paris utilisent des atomes soigneusement préparés et les font traverser une cavité supraconductrice à très haut facteur de qualité. Durant leur passage, ils interagissent avec un mode du champ électromagnétique piégé dans la cavité. Du fait du moment dipolaire électrique élevé des atomes de Rydberg et du très haut facteur de qualité de la cavité, le couplage atome-champ domine totalement le couplage entre le mode de la cavité et l'environnement. Cette expérience permet d'étudier, sous une forme quasiment idéale, le système matière-lumière le plus simple possible : un mode quantifié du champ et

un système atomique à deux niveaux. Avec celles sur les ions piégés, ces expériences ont permis d'illustrer de manière spectaculaire la caractérisation, la manipulation et le contrôle de l'état de systèmes quantiques uniques.

Dans ces expériences, les atomes constituent une sonde de l'état du champ tout à fait inédite. En effet, en ajustant les fréquences atomiques par effet Stark de manière à ce qu'elles ne coïncident pas avec la fréquence propre du mode dans la cavité, les atomes ne peuvent pas absorber ou émettre de photon. Néanmoins, ils emportent une information sur l'état du champ dans la cavité sous la forme d'une différence de phase mesurable par interférométrie.

Ceci permet en retour d'étudier la physique des systèmes quantiques ouverts développée au chapitre 9. Nous montrerons en effet comment il est possible de révéler les trajectoires quantiques suivies par l'état du champ sous l'influence de l'environnement formé par les atomes traversant la cavité, illustrant sur un exemple à temps discret les idées de la section 9.4 sur la dynamique markovienne d'un système quantique ouvert. Enfin, nous verrons comment cette technique permet d'étudier la dynamique du mode quantifié du champ sous l'action de l'autre source de décohérence que sont les fuites de photons hors de la cavité.

10.1.1 Dispositif expérimental et modélisation

Le dispositif expérimental est représenté sur la figure 10.1 : il permet de faire interagir les deux systèmes quantiques simples que nous avons discutés dans le chapitre 4, à savoir un système à deux niveaux et un oscillateur harmonique quantique.

Côté atomique, le système à deux niveaux est caractérisé par l'écart d'énergie $\hbar\omega_{eg}$ entre l'état excité $|e\rangle$ et le fondamental $|g\rangle$ de la transition considérée. Côté cavité, c'est un mode propre du champ de pulsation ω_0 . Enfin, le couplage entre la cavité et le mode du champ est caractérisé par une énergie $\hbar g$. Une discussion des ordres de grandeurs montre que $g \ll \omega_0, \omega_{eg}$; ces dernières étant dans le domaine des micro-ondes ($\omega_0/2\pi \simeq \omega_{eg}/2\pi = 51,1$ GHz) alors que $g/2\pi \simeq 50$ kHz. Nous renvoyons le lecteur au livre de S. Haroche et J.-M. Raimond [194] pour plus de détails sur le dispositif lui-même¹.

L'analyse complète nécessite de prendre en compte les effets dissipatifs associés d'une part à la fuite de photons de la cavité et d'autre part à l'émission spontanée atomique qui est susceptible de les faire transiter vers des niveaux hors du doublet ($|e\rangle, |g\rangle$). Le temps d'émission spontanée de l'atome est de l'ordre de 22 ms, ce qui est bien supérieur au temps de vol des atomes dans la cavité, qui est lui de l'ordre de 100 μ s². Le temps que met un photon à fuir de la cavité est dépendant de la fabrication de la cavité : alors que les premières avaient un temps de fuite de l'ordre de 160 μ s, les cavités de nouvelle

1. Dans [194], notre g correspond à la pulsation du Rabi du vide Ω_0 introduite plus loin. Dans la littérature, $g = \Omega_0/2$ est aussi souvent utilisé.

2. La cavité fait 5 cm de long et le processus de sélection de vitesse permet de sélectionner des vitesses allant de 140 m s⁻¹ à 600 m s⁻¹.

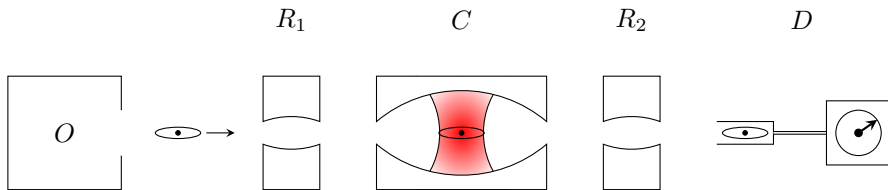


FIGURE 10.1 – Schéma du dispositif utilisé dans l'expérience réalisée dans le groupe de S. Haroche. La cavité supraconductrice C , qui possède un haut facteur de qualité, contient un état du champ électromagnétique préparé à l'aide d'une source classique, non représentée sur le schéma. La cavité est réglable au moyen d'un dispositif électro-mécanique, ce qui permet de réaliser aussi bien des expériences avec une interaction atome-champ résonante que non résonante. Elle est traversée par des atomes excités dits de Rydberg, émis par une source atomique O associée à un dispositif de préparation et de sélection de vitesse. Un détecteur à ionisation D est placé en aval du dispositif. Enfin, deux autres cavités de très basse surtension R_1 et R_2 entourant la cavité principale sont spécifiquement conçues pour pouvoir appliquer des impulsions $\pi/2$ à l'état atomique. Il s'agit donc d'une réalisation expérimentale d'un interféromètre de Ramsey (section 4.1.5).

génération ont atteint un temps de fuite record d'environ 130 ms [242] soit un gain de trois ordres de grandeurs.

Le modèle de Jaynes-Cummings

La dynamique du système formé par un atome et le champ électromagnétique quantifié peut être obtenue en faisant un modèle microscopique à partir de l'électrodynamique en cavité, valide dans le cadre de l'approximation dipolaire électrique et de l'approximation de l'onde tournante³. Le résultat de cette analyse est donné par le *hamiltonien de Jaynes-Cummings* [114] :

$$H = H_{\text{at}} + H_{\text{cav}} + H_{\text{int}} , \quad (10.1a)$$

$$H_{\text{at}} = \frac{\hbar\omega_{eg}}{2} \sigma_z , \quad (10.1b)$$

$$H_{\text{cav}} = \hbar\omega_0 a^\dagger a , \quad (10.1c)$$

$$H_{\text{int}} = \frac{\hbar g}{2} (\sigma_+ a + \sigma_- a^\dagger) , \quad (10.1d)$$

où l'on a introduit des matrices de Pauli agissant sur les degrés de libertés atomiques : $\sigma_+ = |e\rangle\langle g|$, $\sigma_- = |g\rangle\langle e|$, $\sigma_z = |e\rangle\langle e| - |g\rangle\langle g|$ ainsi que les opérateurs de création et de destruction a et $a^\dagger$ pour le mode quantifié du champ

3. Cette dernière approximation, qui revient à ne retenir que les termes dit séculaires, suppose que l'échelle caractéristique de l'interaction g est bien inférieure aux pulsations ω_{eg} et ω_0 de l'atome et de la cavité.

électromagnétique dans la cavité (section 4.2). La constante de couplage g provient du couplage dipolaire électrique entre le moment dipolaire atomique et le champ électrique. Dans le montage expérimental, elle dépend de la position de l'atome au sein de la cavité, ce qui signifie qu'elle va varier au fil de la traversée de la cavité par l'atome. Cependant, pour comprendre la physique sous-jacente à ce modèle, nous allons la supposer constante.

Le hamiltonien se diagonalise par blocs en remarquant que le nombre d'excitations total $a^\dagger a + \sigma_+ \sigma_-$ est une quantité conservée. Ainsi, les sous-espaces dynamiquement stables sont d'une part celui engendré par $|g, 0\rangle$ et d'autre part les sous-espaces de dimension deux engendrés par les états $(|e, n\rangle, |g, n+1\rangle)$. La matrice représentant H dans ce sous-espace est celle d'un système à deux niveaux effectif, atome-champ :

$$H^{(n)} = \hbar \left(n + \frac{1}{2} \right) \omega_0 \mathbb{1} + \frac{1}{2} (\delta \sigma_z + g \sqrt{n+1} \sigma_x), \quad (10.2)$$

où $\delta = \omega_{eg} - \omega_0$ désigne le désaccord entre la transition atomique et la cavité⁴. On s'attend donc à observer des oscillations pour chacun de ces systèmes à deux niveaux effectifs, qui se traduiront par des oscillations de Rabi des populations atomiques analogues à celles discutées au chapitre 1. Les pulsations $\Omega_n(\delta)$ des oscillations sont données par la différence des valeurs propres $\omega_n^{(\pm)}$ des $H^{(n)}$, avec :

$$\omega_n^{(\pm)} = \omega_0 \left(n + \frac{1}{2} \right) \pm \frac{1}{2} \sqrt{\delta^2 + g^2(n+1)} \quad (10.3)$$

et donc $\Omega_n(\delta) = \sqrt{\delta^2 + g^2(n+1)}$. Pour le premier doublet ($n=0$, états $|e, 0\rangle$ et $|g, 1\rangle$), à résonance ($\omega_{eg} = \omega_0$), on voit ainsi des oscillations à la pulsation caractéristique g que l'on appelle *la pulsation de Rabi du vide*⁵.

À g fixé, au sein du doublet n , on doit distinguer entre deux régimes bien différents. Dans le *régime résonant*, défini par l'égalité $\omega_{eg} = \omega_0$ (soit $\delta = 0$), les états propres sont les combinaisons linéaires symétriques et antisymétriques

$$\frac{|g, n+1\rangle \pm |e, n\rangle}{\sqrt{2}}, \quad (10.4)$$

qui sont séparées en énergie par $\Omega_n(0) = \hbar g \sqrt{n+1}$.

À l'inverse, on parle de *régime dispersif* lorsque le désaccord δ entre l'atome et la cavité est grand par rapport à $g \sqrt{n+1}$. Les états propres sont alors très proches des états $|e, n\rangle$ et $|g, n+1\rangle$. Au second ordre, dans le régime dispersif, la théorie des perturbations permet d'écrire un nouveau hamiltonien effectif :

$$H_{\text{eff}}^{(n)} \simeq \hbar(\omega_0 + \chi \sigma_z) n + \frac{\hbar(\omega_{eg} + \chi)}{2} \sigma_z, \quad (10.5)$$

4. Il faut bien faire attention que le système à deux niveaux pour la dynamique donnée par l'équation (10.1) et celui donné par l'équation (10.2) ne sont pas les mêmes. Le premier se réfère à $(|e\rangle, |g\rangle)$, l'atome lui-même, tandis que le second se réfère au système composite atome-champ $(|e, n\rangle, |g, n+1\rangle)$.

5. C'est pourquoi g est parfois noté Ω_0 .

où $\chi = g^2/4\delta$. Cette description montre une renormalisation de la fréquence de transition atomique ω_{eg} en $\omega_{eg} + \chi$ en l'absence de photons ($n = 0$) : c'est le *déplacement de Lamb*. De plus, les niveaux atomiques sont décalés en fonction du nombre de photons présents dans la cavité : c'est ce qu'on appelle l'*effet Stark dynamique* (ou déplacement lumineux). Une interprétation équivalente consiste à dire que la pulsation du mode propre dans la cavité ω_0 est modifiée par l'état atomique en $\omega_0 \pm \chi$. C'est un effet d'indice de réfraction, puisque, hors résonance, les atomes sont transparents au rayonnement, ne pouvant ni émettre ou ni absorber de photon. Ils se comportent donc comme des diélectriques dont l'indice de réfraction dépend de l'état atomique. Ce sont ces effets qui seront mis à profit pour caractériser l'état du champ dans la cavité au moyen du passage d'un train d'atomes. À l'ordre de la théorie de perturbation considéré, ces différents effets sont linéaires mais il faut garder à l'esprit que, dans l'expérience, on peut observer des déviations non linéaires provenant des ordres supérieurs dans le traitement perturbatif.

10.1.2 Les trajectoires quantiques

Préparer le système dans le régime dispersif a l'intérêt de pouvoir voir l'atome comme un petit appareil de mesure généralisée du champ. Ceci peut ensuite être exploité en envoyant un train d'atomes dans la cavité pour reconstruire une information bien plus complète, à savoir la distribution de probabilité du nombre de photons dans l'état quantique du mode en réalisant, dans la limite des grands nombre d'atomes, une mesure projective de l'opérateur nombre de photons. Ainsi, ces montages permettent d'illustrer de nombreux concepts : mesure projective, mesure généralisée, état relatif ou encore trajectoire quantique.

La collision atome-champ non résonante

Nous pouvons décrire le passage d'un atome à travers la cavité en régime dispersif comme une collision. En partant d'un état atomique $|a\rangle$ avec $a = e$ ou g et d'un état du champ quelconque, l'état atome-champ évolue, selon l'équation (10.5), comme

$$|\psi(t)\rangle = U(t) \left(\sum_{n=0}^{+\infty} \alpha_n |n\rangle |a\rangle \right) = e^{i\varphi_a(t)} \sum_{n=0}^{+\infty} \alpha_n e^{in\delta_a(t)} |n\rangle |a\rangle \quad (10.6)$$

avec $\varphi_a(t) = \mp(\omega_{eg} + \chi)t/2$ et $\delta_a(t) = -(\omega_{eg} \pm \chi)t$. Les différents états de Fock $|n\rangle$ impriment donc une information sur n dans une phase qui dépend de l'état atomique. Or, rien de tel qu'un interféromètre pour récupérer une information de phase.

Pour cela, on prépare au moyen d'une radiofréquence résonante classique durant une durée adaptée (impulsion $\pi/2$, section 4.1.5) un état atomique de la forme $(|g\rangle + |e\rangle)/\sqrt{2}$ de sorte que, si on a exactement n photons dans la

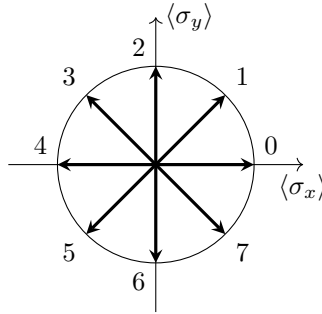


FIGURE 10.2 – Représentation des positions des vecteurs de Bloch associés aux états atomiques $|+\pi n/q\rangle$ intriqués aux différents états de Fock $|n\rangle$ dans la cavité, avec $n = 0 \dots 2q - 1$ et $q = 4$.

cavité, l'état atome-champ en sortie de la cavité est de la forme, en factorisant une phase globale,

$$|\psi(t)\rangle = \frac{|g\rangle + e^{i\varphi_n(t)} |e\rangle}{\sqrt{2}} \otimes |n\rangle \quad (10.7)$$

avec $\varphi_n(t) = \varphi_e(t) - \varphi_g(t) + n(\delta_e(t) - \delta_g(t))$. Sur la sphère de Bloch, l'état en entrée de la cavité, qui correspond au vecteur $\mathbf{e}_x$ de la sphère de Bloch, tourne dans le plan équatorial de l'angle φ_n . Pour révéler cette phase, et donc le nombre de photons n , nous allons effectuer une impulsion inverse de la précédente réalisant ainsi un interféromètre de Ramsey (section 4.1.5) avec une différence de phase $\varphi_e - \varphi_g$ entre les deux chemins.

Effet d'une seule mesure

En jouant sur la phase $\varphi_n(t)$ de l'équation (10.7), nous pouvons préparer l'atome dans une superposition d'états générique. On suppose ici que l'état de l'atome en sortie de la cavité est de la forme :

$$|+\pi n/q\rangle = \frac{|e\rangle + e^{i\pi n/q} |g\rangle}{\sqrt{2}}, \quad (10.8)$$

où n est le nombre de photons dans la cavité et q est un paramètre choisi par l'expérimentateur. Nous allons voir que cela permet de déterminer de façon unique le nombre de photons de la cavité entre 0 et $2q - 1$.

En général, pour un état initial du mode du champ dans la cavité $|\psi_i\rangle = \sum_n \alpha_n |n\rangle$, l'état du système atome-champ en sortie de la cavité est donné par

$$|\psi_i\rangle \otimes |+_0\rangle \rightarrow \sum_{n=0}^{+\infty} \alpha_n e^{-i\omega_0 n T} e^{-i\pi n/2q} |n\rangle \otimes |+\pi n/q\rangle, \quad (10.9)$$

où T est le temps de vol de l'atome à travers la cavité. Les états $|+\pi n/q\rangle$ sont donc les états de l'atome relatifs aux différentes valeurs du nombre de photons dans la cavité. Ils sont dans le plan équatorial de la sphère de Bloch comme indiqué sur la figure 10.2 : ce sont en quelque sorte la position de l'aiguille de l'appareil qui va nous servir à mesurer le nombre de photons. En mesurant l'état atomique dans la base $(|+\varphi\rangle, |-\varphi\rangle)$, correspondant aux vecteurs de Bloch situés dans le plan équatorial et d'azimuts respectifs φ et $\varphi + \pi$, les oscillations des amplitudes suivantes

$$|+\varphi_n\rangle = \cos\left(\frac{\varphi - \varphi_n}{2}\right) |+\varphi\rangle + \sin\left(\frac{\varphi - \varphi_n}{2}\right) |-\varphi\rangle \quad (10.10a)$$

$$|-\varphi_n\rangle = \cos\left(\frac{\varphi - \varphi_n}{2}\right) |-\varphi\rangle + \sin\left(\frac{\varphi - \varphi_n}{2}\right) |+\varphi\rangle \quad (10.10b)$$

en fonction de φ permettent de remonter à la phase φ_n .

États relatifs du champ

Nous pouvons maintenant analyser l'effet sur l'état du champ électromagnétique d'une mesure effectuée sur l'atome à l'instant t , première étape dans l'étude des trajectoires quantiques de l'état du champ en présence de mesures effectuées sur les atomes traversant la cavité.

Pour cela, nous allons écrire l'état atome-champ d'une manière adaptée pour faire apparaître les états du champ relatifs aux états atomiques $|\pm\varphi\rangle$. Si $|\psi(t)\rangle$ désigne l'état du champ juste avant la mesure qui dure un temps T :

$$U(T) |\psi(t)\rangle \otimes |+_0\rangle = \sqrt{p_+} |\psi(t+T|+\varphi)\rangle \otimes |+\varphi\rangle + \sqrt{p_-} |\psi(t+T|-\varphi)\rangle \otimes |-\varphi\rangle . \quad (10.11)$$

Les états relatifs $|\psi(t+T|\pm\varphi)\rangle$ du mode de la cavité juste après la mesure sont alors donnés par

$$|\psi(t+T|+\varphi)\rangle = \frac{e^{-i\varphi/2}}{\sqrt{p_+}} \sum_n \alpha_n e^{-i\omega_0 n T} \cos\left(\frac{1}{2}\left(\varphi - \frac{\pi n}{q}\right)\right) |n\rangle , \quad (10.12a)$$

$$|\psi(t+T|-\varphi)\rangle = \frac{e^{-i\varphi/2}}{\sqrt{p_-}} \sum_n \alpha_n e^{-i\omega_0 n T} \sin\left(\frac{1}{2}\left(\varphi - \frac{\pi n}{q}\right)\right) |n\rangle . \quad (10.12b)$$

Les probabilités $p_{\pm}$ qui apparaissent dans ces formules sont celles d'obtention des deux résultats possibles dans la base $|\pm\varphi\rangle$:

$$p_+ = \sum_n |\alpha_n|^2 \cos^2\left(\frac{1}{2}\left(\varphi - \frac{\pi n}{q}\right)\right) , \quad (10.13a)$$

$$p_- = \sum_n |\alpha_n|^2 \sin^2\left(\frac{1}{2}\left(\varphi - \frac{\pi n}{q}\right)\right) . \quad (10.13b)$$

Les opérateurs $M_+(\varphi)$ et $M_-(\varphi)$, qui correspondent aux résultats de mesure associés à la détection de l'atome dans les états $|\pm\varphi\rangle$ (section 5.4) :

$$M_+(\varphi) = \cos\left(\frac{1}{2}\left(\varphi - \frac{\pi}{q}N\right)\right), \quad (10.14a)$$

$$M_-(\varphi) = \sin\left(\frac{1}{2}\left(\varphi - \frac{\pi}{q}N\right)\right), \quad (10.14b)$$

où N est l'opérateur nombre de photons dans la cavité. Ces opérateurs vérifient $M_+(\varphi)^\dagger M_+(\varphi) + M_-(\varphi)^\dagger M_-(\varphi) = \mathbb{1}$. Ils décrivent donc une mesure généralisée (effectuée par l'atome) du champ dans la cavité.

Nous allons maintenant décrire ce qui se passe quand on envoie un train de N atomes à travers la cavité, pour comprendre l'extraction d'information et l'évolution de l'état du champ comme une trajectoire en fonction de l'historique des résultats de mesures.

Trajectoires quantiques associées à un train d'atomes

Pour comprendre l'effet d'un train d'atomes, il faut voir celui-ci comme un environnement sans mémoire pour le mode du champ électromagnétique piégé dans la cavité. En effet, le numéro de l'atome au sein du train va jouer le rôle du temps et l'espace des états $\mathcal{H}_l \sim \mathbb{C}^2$ du l -ème atome va correspondre à l'espace $\mathcal{H}_t$ associé à un intervalle de temps dans la section 9.4.

Le passage de N atomes à travers la cavité crée un état intriqué entre le mode du champ dans la cavité et les N atomes. Une trajectoire quantique est alors la donnée des états relatifs associés aux résultats de mesures successives effectuées du premier au dernier atome. On obtient alors 2^N trajectoires quantiques différentes. Les équations (10.12) nous disent que chaque mesure modifie les amplitudes de l'état du champ sur les différents états à $|k\rangle$ photons pour $k = 0$ à $k = 2q - 1$: chacune de ces amplitudes est multipliée par un facteur qui dépend du résultat de la mesure et qui diminue le poids associé à certaines valeurs de k .

La figure 10.3 représente schématiquement l'effet de la détection de quelques atomes sur le champ dans la cavité. Si φ est choisi de la forme $\pi k/q$, chaque détection d'un atome correspond à l'application d'un opérateur qui vaut 0 en une valeur de N unique (modulo $2q$) : $N = k$ pour M_- et $N = k + q$ pour M_+ . Après quelques mesures, on va ainsi sélectionner des états du champ compatibles avec les résultats obtenus. Notons que dans le cas $q = 1$, on effectue ainsi une mesure projective de la parité du nombre de photons en un seul coup. En effet, la détection de l'atome permet de discerner parfaitement entre les états atomiques relatifs à la mesure de la parité du nombre de photons dans la cavité. Cela fut exploité pour mettre en évidence les sauts quantiques dans la cavité peuplée par un faible nombre de photons [178].

Cette discussion qualitative suggère deux choses : en premier lieu, elle montre que l'on peut utiliser ce processus stochastique d'origine quantique

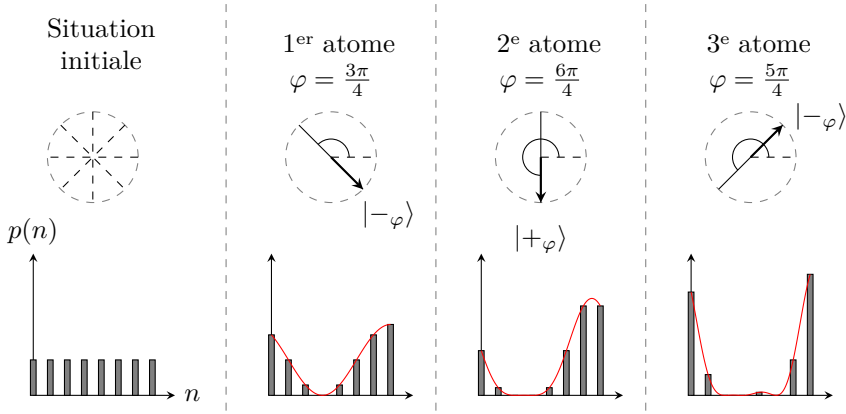


FIGURE 10.3 – Effet des mesures sur le champ dans la cavité. Pour chaque passage d'un atome, on choisit une base de mesure définie par l'angle φ , parmi 8 bases possibles, et on détecte l'atome dans un des deux états $|\pm\varphi\rangle$. Les amplitudes pour l'état du champ dans la cavité relatif au résultat sont alors les amplitudes initiales, multipliées par un cosinus ou un sinus, ce qui correspond à l'application de l'opérateur $M_{\pm}(\varphi)$. Ici, on a représenté l'évolution des probabilités $p(n)$ qui sont égales aux modules des amplitudes au carré, en partant d'un champ dans l'état connu $\sum_{i=0}^7 |i\rangle/\sqrt{8}$. On voit qu'à chaque mesure dans une nouvelle base on va éliminer la composante du champ qui correspond de manière exacte à l'état orthogonal au résultat mesuré pour l'atome.

pour affiner notre connaissance de la distribution de probabilité du nombre de photons présents dans la cavité le long de la trajectoire considérée. En second lieu, elle suggère qu'à la limite d'un grand nombre d'atomes, le processus projette l'état du champ sur un état à nombre de photons donné et donc prépare un état de Fock. Le processus limite est en fait une mesure projective du nombre de photons dans le mode électromagnétique.

Commençons par discuter comment extraire l'information sur la distribution de probabilité du nombre de photons associée à l'état du champ dans la cavité. Pour cela, on doit travailler avec des probabilités bayésiennes qui reflètent notre connaissance de l'état du champ dans la cavité (section 2.4). Supposons que, au bout de k passages, on ait inféré une distribution de probabilité $p_k(n|R_k)$ pour le nombre de photons n où $R_k = \{(\epsilon_j, \varphi_j)_{j=1\dots k}\}$ est l'ensemble des résultats de mesures ϵ_j pour le choix de phase φ_j pour les k atomes déjà passés. Par la règle de Bayes, cette distribution est actualisée en utilisant le résultat de la $(k+1)$ -ème mesure :

$$p_{k+1}(n|R_{k+1}) = \frac{p(\epsilon_{k+1}, \varphi_{k+1}|n) p_k(n|R_k)}{\sum_{m=0}^{2q} p_m(n|R_m) p(\epsilon_{m+1}, \varphi_{m+1}|n)}. \quad (10.15)$$

Le germe du processus, c'est-à-dire la distribution de probabilité initiale, est une distribution de probabilité qui reflète notre connaissance *a priori* de la distribution du nombre de photons présents dans la cavité. Si on ne suppose rien sur le processus de préparation de l'état du champ dans la cavité, le plus simple est de prendre une distribution plate $p_0(n) = 1/2q$. Comme expliqué dans la section 2.4.3, le gain d'information dans le processus bayésien est positif en moyenne.

Le point non trivial est que, pour une trajectoire donnée, ce processus finit toujours par aboutir à une distribution de probabilité qui est concentrée sur une valeur donnée du nombre de photons. Cette valeur n'est pas la même d'une histoire à l'autre : c'est une variable aléatoire. Comme la mesure n'entraîne pas d'émission ou d'absorption de photons, sa loi de probabilité n'est autre que la distribution du nombre de photons issue de la préparation du champ dans la cavité avant que ne débutent les mesures. Dans la limite d'un nombre infini de mesures, on peut donc démontrer que l'état du champ dans la cavité tend vers l'état de Fock $|n\rangle$ avec une probabilité $p_\infty(n) = p_i(n)$, où $p_i(n)$ est la probabilité d'avoir n photons dans l'état initial à l'issue du processus de préparation. Autrement dit, un train suffisamment long d'atomes réalise une mesure projective du nombre de photons dans la cavité. La démonstration de ce résultat sur laquelle nous reviendrons dans le prochain paragraphe est faite de manière détaillée dans [34].

Pour finir cette discussion, intéressons-nous à l'état du champ dans la cavité après la traversée des N atomes. Celui-ci est décrit au moyen d'un opérateur densité car on doit additionner les deux possibilités associées aux deux résultats $|\pm_\varphi\rangle$. L'opérateur densité donnant l'état de la cavité à l'instant $t + T$ est alors

$$\rho(t + T) = M_+(\varphi)\rho(t)M_+^\dagger(\varphi) + M_-(\varphi)\rho(t)M_-^\dagger(\varphi). \quad (10.16)$$

L'évolution des éléments de matrice dans la base de Fock $\rho_{nn'} = \langle n'|\rho|n\rangle$ est alors donnée par

$$\rho_{nn'}(t + T) = \rho_{nn'}(t)e^{i\omega_0(n-n')T} \cos\left(\frac{\pi}{2q}(n - n')\right), \quad (10.17)$$

où le facteur de phase reflète l'évolution libre du mode d'oscillateur et le facteur en cosinus provient de l'interaction avec les atomes qui ont traversé la cavité. Cette équation est l'analogue en temps discret de l'équation de Linblad (9.24) qui régit l'évolution de l'opérateur densité réduit décrivant l'état du système sous l'action de sa dynamique propre et du couplage à son environnement. Ici, le passage de chaque atome se fait durant un intervalle de temps de durée T et on a donc un processus markovien discret.

Imaginons maintenant que l'on envoie N atomes à travers la cavité pendant la durée totale $T_{\text{exp}} = NT$, l'état final de la cavité est alors décrit par :

$$\rho_{nn'}(T_{\text{exp}}) = \rho_{nn'}(0)e^{i\omega_0(n-n')T_{\text{exp}}} \left[\cos\left(\frac{\pi}{2q}(n - n')\right) \right]^{T_{\text{exp}}/T}. \quad (10.18)$$

Les termes diagonaux qui donnent les probabilités d'occupation des divers états de la base de Fock ne sont pas affectés, ce qui est attendu vu que les atomes sont en régime dispersif et n'absorbent ni n'émettent aucun photon dans la cavité. En revanche, les termes non diagonaux sont exponentiellement atténués : on voit ici la décohérence dans la base de Fock provenant de l'empreinte quantique laissée par le mode de la cavité sur le train d'atomes.

Développements expérimentaux

Une question naturelle consiste à savoir combien d'atomes il faut envoyer à travers la cavité pour réaliser cette sélection d'états à nombre de photons fixé. En effet, dans l'optique de pouvoir mesurer la distribution $p_i(n)$ le plus efficacement possible, il va être nécessaire de reproduire un grand nombre de fois l'expérience jusqu'à l'étape sélectionnant une valeur de n donnée, et donc essayer d'atteindre cette étape le plus vite possible. Dans les premières expériences, les angles de mesures φ furent choisis suivant une séquence bien déterminée. Mais c'est loin d'être le plus efficace. Le protocole optimal a été proposé par Brune, Raimond et Haroche en 1992 [193]. L'idée [194] consiste à utiliser des choix de paramètres qui permettent d'accéder à la décomposition en binaire des nombres de photons susceptibles d'être présents dans la cavité. Si $\bar{n}_{\text{cav}}$ désigne le nombre maximum de photons que l'on souhaite sonder, il faudra $\log_2 \bar{n}_{\text{cav}}$ atomes pour que le processus d'inférence bayésienne converge. Pour sonder chaque chiffre de la décomposition du nombre de photons en binaire, on devra changer la phase de l'interféromètre mais aussi le désaccord atome-cavité ou bien le temps d'interaction atome-champ, ce qui revient à changer le paramètre q de l'équation (10.8). En termes de théorie d'information, on peut dire que ce processus permet, au mieux, un gain d'un bit d'information par atome ce qui est optimal. Brune *et al.* interprètent ce protocole comme la réalisation quantique d'un convertisseur analogique/numérique qui procède par étapes : on commence par tester si la tension mesurée est entre 0 et $V_{\text{max}}/2$ ou entre $V_{\text{max}}/2$ et V_{max} . Suivant le résultat, on décale la tension puis on l'amplifie alors d'un facteur deux et on recommence. Cela donne en N étapes une numérisation à N bits de la tension sur l'intervalle $[0, V_{\text{max}}/2]$.

Alors que la mesure de la parité du nombre de photons fut réalisée en 1999 [289] par une technique totalement différente, la mesure non destructive du nombre de photons jusqu'à 7 ($q = 4$) ne fut réalisée qu'en 2007 [189] car, pour y arriver, il faut disposer de cavités supraconductrices avec un temps de vie suffisamment long pour que la relaxation n'ait pratiquement pas d'action durant le temps de passage d'un train permettant la réduction sur un état de Fock [242].

Ces expériences montrent qu'il est possible de préparer un état de Fock à nombre de photons donné dans la cavité et que, sur un très grand nombre de réalisations de l'expérience, ce nombre est aléatoire avec une statistique qui reflète la préparation de l'état du champ dans la cavité [189]. La figure 10.4

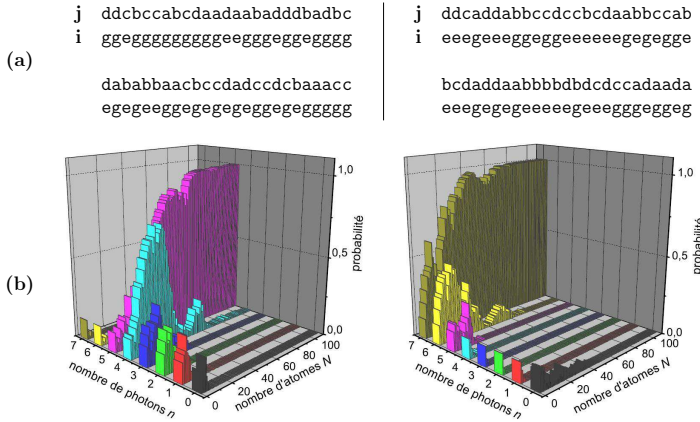


FIGURE 10.4 – Effondrement progressif de la distribution du nombre de photons $p_k^{[R_k]}$ vers une distribution δ_{n,n_∞} pour deux réalisations d’une expérience où l’on envoie un train d’atomes en régime dispersif à travers une cavité. (a) Résultats expérimentaux avec en première ligne la liste des angles et en dessous les résultats correspondants. (b) Histogrammes représentant les $p_k^{[R_k]}$ pour $k = 0, \dots, 110$. Graphe de gauche : $n_\infty = 5$. Graphe de droite : $n_\infty = 7$. Reproduit avec l’autorisation de *Nature/Springer/Palgrave* [189], © 2019.

montre la stabilisation de la distribution de probabilité $p_k^{[R_k]}$ sur une distribution p_∞ . Actuellement, le protocole optimal n’a pas encore été déployé. En revanche, un protocole adaptatif qui choisit la mesure à faire en fonction d’une estimation du gain d’information a été mis en œuvre [296, 295].

10.1.3 Tomographie d’un champ en cavité

Une application spectaculaire des concepts développés ici est la reconstruction de l’état quantique du champ électromagnétique au sein de la cavité au moyen des mesures effectuées avec les atomes non résonants traversant la cavité.

Préparation d’états non classiques

Mettre en œuvre un protocole de tomographie quantique n’est intéressant que si l’on souhaite reconstruire des états non classiques du système. Voyons comment préparer de tels états pour le champ.

Nous avons vu à la section 4.2 que les états du champ qui correspondent aux configurations classiques dans la théorie de Maxwell sont les états cohérents du champ électromagnétique. Ils sont caractérisés, pour un mode dans une

cavité, par une amplitude complexe α qui encode la valeur moyenne du champ dans ce mode. L'état correspondant est alors noté $|\alpha\rangle$.

L'interaction atome-champ produisant un déphasage par photon qui dépend de l'état atomique, il est facile de voir grâce à l'équation (4.53) que, si on envoie un atome préparé dans l'état $(|e\rangle + |g\rangle)/\sqrt{2}$ à travers la cavité où se trouve un état cohérent $|\alpha\rangle$, alors l'état du système atome-champ après le passage de l'atome est intriqué :

$$|\Psi_{\text{int}}\rangle = \frac{|e\rangle |\alpha e^{i\delta_e}\rangle + |g\rangle |\alpha e^{i\delta_g}\rangle}{\sqrt{2}}. \quad (10.19)$$

Pour préparer un état superposé du champ, on va agir sur l'atome de manière à effacer l'information sur l'état de la cavité. Cela est réalisé par l'application d'un opérateur σ_x sur l'état atomique, en utilisant une impulsion $\pi/2$ après la traversée de la cavité. L'état ainsi obtenu est

$$|\Psi_{\text{out}}\rangle = \frac{|e\rangle + |g\rangle}{2} \otimes |\alpha e^{i\delta_e}\rangle + \frac{|e\rangle - |g\rangle}{2} \otimes |\alpha e^{i\delta_g}\rangle. \quad (10.20)$$

La détection de l'atome dans un des deux états $|e\rangle$ ou $|g\rangle$ produit alors les états relatifs du champ dans la cavité

$$|\psi(\text{out}|e)\rangle = \frac{|\alpha e^{i\delta_e}\rangle + |\alpha e^{i\delta_g}\rangle}{\sqrt{2}}, \quad (10.21a)$$

$$|\psi(\text{out}|g)\rangle = \frac{|\alpha e^{i\delta_e}\rangle - |\alpha e^{i\delta_g}\rangle}{\sqrt{2}}. \quad (10.21b)$$

Ce processus produit donc un état quantique superposition de deux états cohérents de même nombre moyen de photons mais avec des phases différentes. Un tel état ne peut être généré au moyen d'une source de courant classique (section 4.2.2).

Tomographie de l'état du champ

Comment reconstituer l'état quantique du champ électromagnétique dans la cavité à partir de mesures effectuées sur les atomes ?

En pratique, on ne mesure pas directement les éléments de matrice de l'opérateur densité ρ du champ dans la base en nombre d'occupation ou en « position » (section 4.2). Les mesures donnent en fait accès à une fonction équivalente à l'opérateur densité que l'on appelle la *fonction de Wigner* qui dépend de la variable complexe $\alpha = X + iP$:

$$W_\rho(X + iP) = \int_{-\infty}^{+\infty} \left\langle X + \frac{Y}{2} \left| \rho \right| X - \frac{Y}{2} \right\rangle e^{iPY} dY. \quad (10.22)$$

Contrairement aux éléments de matrice non diagonaux de l'opérateur densité, qui sont génériquement complexes, la fonction de Wigner est réelle. Ses

marginales, c'est-à-dire l'intégration sur une des deux variables, donnent la distribution de probabilité pour la variable restante. Mais surtout, on peut y accéder directement par des mesures car la valeur de la fonction de Wigner en $\alpha = (X + iP)/\sqrt{2}$ est donnée par [262]

$$W_\rho(\alpha) = \frac{2}{\pi} \operatorname{tr} (D(\alpha) \rho D(-\alpha) (-1)^N), \quad (10.23)$$

où $D(\alpha)$ est l'opérateur déplacement défini par l'équation (4.54) et N est l'opérateur nombre. Il est alors possible de reconstruire la fonction de Wigner et donc l'opérateur densité ρ avec une mesure de la parité $(-1)^N$ du nombre de photons. En effet, partant de l'état du champ ρ , le branchement d'une source classique convenablement calibrée transforme celui-ci en $D(\alpha)\rho D(-\alpha)$. Une mesure de la parité du nombre de photons réalisée par la technique interférométrique décrite dans la section précédente permet alors de reconstruire W_ρ à partir des données expérimentales pour un grand nombre de répétitions de l'expérience. Remarquons qu'il est important d'avoir une cavité dont le taux de fuite est faible et un train d'atomes assez court pour que cette procédure fournisse un instantané sur une échelle de temps courte devant l'échelle de temps sur laquelle les photons s'échappent de la cavité.

☞ Séduisante en théorie, cette méthode ne fonctionne pas aussi simplement en pratique du fait d'une dépendance non linéaire en le nombre de photons de la différence de phase entre les deux états atomiques accumulée lors de la traversée de la cavité. On utilise donc une procédure de reconstruction de l'opérateur densité réduit qui s'appuie sur un ensemble plus large de données expérimentales. On remarque que les probabilités de détection atomiques p_e et p_g après traversée de l'interféromètre sont données par

$$\begin{aligned} p_e - p_g &= \operatorname{tr} (D(\alpha) \rho D(-\alpha) (M_+(\varphi)^2 - M_-(\varphi)^2)) \\ &= \operatorname{tr} \left(D(\alpha) \rho D(-\alpha) \cos \left(\varphi - \frac{\pi}{q} N \right) \right). \end{aligned} \quad (10.24)$$

Comme le nombre de photons ne varie pas en régime dispersif, le passage d'un train suffisamment compact d'atomes donne accès à la quantité $p_e - p_g$ en faisant la statistique des détections atomiques successives [125]. Ces techniques ont permis d'obtenir des images au cours du temps de l'état du champ initialement préparé dans une superposition de deux états cohérents de phases opposées [126].

10.1.4 Fuites de photons

Nous allons maintenant prendre en compte les imperfections de la cavité, en laissant la possibilité pour les photons de sortir de celle-ci. Cela correspond donc à l'ajout d'un environnement pour les photons avec lequel ils peuvent interagir. Là encore, notre approche consiste à modéliser le processus non pas microscopiquement mais de manière simplifiée tout en capturant l'essentiel de

la physique pour décrire les expériences. Cette section est ainsi une application directe, sur le cas particulier des photons dans une cavité, des méthodes génériques présentées dans le chapitre 9.

Modélisation et équation des trajectoires

Supposons que les photons qui s'échappent de la cavité sont détectés par un détecteur ayant deux états : l'état $|0\rangle$ (pas de détection) et l'état $|1\rangle$ (détection d'un photon). On suppose de plus que le détecteur peut détecter au plus un photon pendant la durée dt . Les opérateurs d'échelle D et $D^\dagger$ du détecteur sont définis par :

$$D|0\rangle = 0 \quad D|1\rangle = |0\rangle \quad D^\dagger|1\rangle = 0 \quad D^\dagger|0\rangle = |1\rangle. \quad (10.25)$$

Le détecteur absorbe les photons. On modélise donc le hamiltonien décrivant l'interaction champ-détecteur par un opérateur de la forme $aD^\dagger$, auquel on doit rajouter le conjugué hermitien :

$$H_{\text{int}} = \hbar g_d (aD^\dagger + a^\dagger D). \quad (10.26)$$

Le détecteur et la cavité interagissent pendant un temps dt , à la suite de quoi l'état du détecteur est remis à $|0\rangle$. En représentation d'interaction [252], nous obtenons :

$$U(t, t+dt) |\psi(t)\rangle \otimes |0\rangle = \left(\mathbb{1} - \frac{g_d^2 dt^2}{2} a^\dagger a \right) |\psi(t)\rangle \otimes |0\rangle - i g_d dt a |\psi(t)\rangle \otimes |1\rangle. \quad (10.27)$$

Calculons maintenant le taux de fuite de la cavité en fonction de la force du couplage g_d . La probabilité de ne pas avoir eu d'émission de photons entre 0 et t se calcule *via* l'histoire, notée $[0]$ dans laquelle le détecteur reste dans son état $|0\rangle$. En découpant le temps en tranches de durée dt , de sorte que $t = Ndt$, on a pour un état initial à n photons :

$$|\psi(t, [0])\rangle = \left(1 - \frac{ng_d^2 dt^2}{2} \right)^N |\psi(0)\rangle = \sqrt{p_0(n, t)} |\psi(0)\rangle, \quad (10.28)$$

où $p_0(n, t)$ est la probabilité de ne pas avoir d'émission de photons pendant la durée t . Elle est donnée par

$$p_0(n, t) = \lim_{N \rightarrow +\infty} \left(1 - \frac{ng_d^2 dt^2}{2} \right)^{2N} = e^{-n(g_d^2 dt)t}, \quad (10.29)$$

où l'on a utilisé $(dt)^2 = (t/N)dt$ pour prendre la limite $N \rightarrow +\infty$. Cette probabilité décroît exponentiellement au cours du temps avec un taux de décroissance proportionnel à n . On retrouve ici le facteur n d'émission stimulée [146] qui, comme expliqué par Feynman [156, Chapitre XX], provient des

interférences constructives liées à la statistique bosonique (section 1.2.5) dans le processus d'émission. Le taux de fuite de la cavité κ , qui est obtenu pour $n = 1$, est relié à l'échelle de temps dt et à la constante de couplage par

$$g_d^2 dt = \kappa. \quad (10.30)$$

L'équation (10.27) prend alors la forme suivante en fonction du paramètre physique κ :

$$U(t, t + dt) |\psi(t)\rangle \otimes |0\rangle = \left(\mathbb{1} - \frac{\kappa dt}{2} a^\dagger a \right) |\psi(t)\rangle \otimes |0\rangle - i\sqrt{\kappa dt} a |\psi(t)\rangle \otimes |1\rangle. \quad (10.31)$$

Elle décrit les trajectoires quantiques pour un mode du champ électromagnétique placé dans une cavité imparfaite dont le taux de fuite est κ . Dans cet exemple, l'espace $\mathcal{H}_t$ de la section 9.4 est l'espace des états $\mathbb{C}^2$ du détecteur entre t et $t + dt$. Les trajectoires stochastiques quantiques sont alors décrites par les règles suivantes pour l'état du mode de la cavité joint à l'enregistrement $[n]$ des états des détecteurs :

$$|\psi(t + dt), [0_{t+dt}, n]\rangle = \left(1 - \frac{\kappa dt}{2} a^\dagger a \right) |\psi(t)[n]\rangle, \quad (10.32a)$$

$$|\psi(t + dt), [1_{t+dt}, n]\rangle = -i\sqrt{\kappa dt} a |\psi(t)[n]\rangle. \quad (10.32b)$$

La comparaison avec l'équation (9.23) montre qu'il n'y a qu'un opérateur de saut non trivial $L = \sqrt{\kappa} a$. Celui-ci correspond à la destruction d'un photon dans la cavité qui fuit dans un environnement à température nulle. L'équation de Lindblad correspondant à ce problème (équations (9.24) et (9.26)) s'écrit alors

$$\frac{d\rho}{dt} = -\frac{\kappa}{2} \{a^\dagger a, \rho(t)\} + \kappa a \rho(t) a^\dagger. \quad (10.33)$$

Elle permet de calculer l'évolution temporelle de l'opérateur densité $\rho(t)$ décrivant l'état de la cavité seule. L'évolution temporelle du nombre moyen de photons dans le mode de la cavité $\langle N \rangle_{\rho(t)}$ est décrite par l'équation :

$$\frac{d\langle N \rangle_{\rho(t)}}{dt} = -\kappa \langle N \rangle_{\rho(t)} \quad (10.34)$$

qui prédit une diminution exponentielle du nombre de photons dans la cavité au cours du temps, avec un temps caractéristique $1/\kappa$, comme attendu pour une relaxation avec un taux de fuite κ . Nous allons maintenant utiliser l'équation (10.31) pour décrire les phénomènes de relaxation et de décohérence pour un mode du champ en cavité dans quelques exemples physiquement importants.

Relaxation de la cavité

Commençons par une cavité initialement peuplée avec un état de Fock à n photons, $|n\rangle$. Lorsqu'un photon est détecté en dehors de la cavité, cela signifie que celle-ci a transité vers un état comprenant un photon de moins : $|n\rangle \rightarrow |n-1\rangle$ ($n \geq 1$). Entre deux tels événements, appelés sauts quantiques, l'état de Fock de la cavité n'évolue pas, étant un état propre de la partie hermitienne de l'équation (10.33). Lorsque le vide est atteint, plus rien ne se passe. La relaxation de la cavité se fait donc par émissions successives de photons.

Il est intéressant de regarder la statistique des temps d'émission ou, de manière équivalente, du nombre de photons émis entre l'instant initial $t_i = 0$ et un instant $t \geq 0$. La probabilité que l'on n'émette aucun photon au bout du temps t décroît exponentiellement pour $n\kappa t \gg 1$. L'expression (10.29) pour la probabilité de ne pas avoir émis de photon durant l'intervalle $[0, t]$ permet alors de déduire la distribution de probabilité pour le temps de première émission d'un photon $\mathbf{p}(n, \tau)$. En effet, comme $p_0(n, t) = \int_t^\infty \mathbf{p}(n, \tau) d\tau$, la densité de probabilité pour la date de première émission est donnée par :

$$\mathbf{p}(n, \tau) = n\kappa e^{-\kappa n \tau} . \quad (10.35)$$

qui est une distribution correspondant à un processus poissonien où la probabilité d'émission par unité de temps est $n\kappa$. À partir de cette expression, et compte tenu du fait qu'après un saut quantique l'état du champ dans la cavité est également un état de Fock, il est possible de trouver une expression pour la distribution de probabilité $p(k, t|n, 0)$ d'avoir émis k photons sur l'intervalle temporel $[0, t]$ sachant qu'on en avait n initialement :

$$p(k, t|n, 0) = \frac{n!}{(n-k)! k!} (1 - e^{-\kappa t})^k e^{-(n-k)\kappa t} . \quad (10.36)$$

Cette statistique binomiale peut s'interpréter en disant que la cavité émet k photons parmi n et que la probabilité qu'un photon soit émis est $1 - e^{-\kappa t}$. Tout se passe donc comme si les n photons dans la cavité se comportaient comme des objets classiques devant passer au travers d'une barrière de transparence $1 - e^{-\kappa t}$. La moyenne sur l'ensemble des trajectoires redonne la décroissance exponentielle de l'énergie moyenne stockée dans la cavité avec la constante de temps κ conformément à l'équation générale (10.34). La figure 10.5 résume ces résultats pour un état de Fock initial $|n = 10\rangle$.

Il est intéressant de considérer le cas d'un état cohérent $|\alpha\rangle$. Comme $|\alpha\rangle$ est un état propre de l'opérateur a , l'action de l'opérateur de saut est donnée par

$$\frac{a|\alpha\rangle}{\sqrt{\langle N \rangle_{|\alpha\rangle}}} = e^{i \arg(\alpha)} \frac{|\alpha|}{\sqrt{|\alpha|^2}} |\alpha\rangle = e^{i \arg(\alpha)} |\alpha\rangle . \quad (10.37)$$

L'état est donc uniquement modifié par une phase. En conséquence, la fuite d'un photon hors de la cavité laisse un tel état invariant. Mais alors, d'où vient

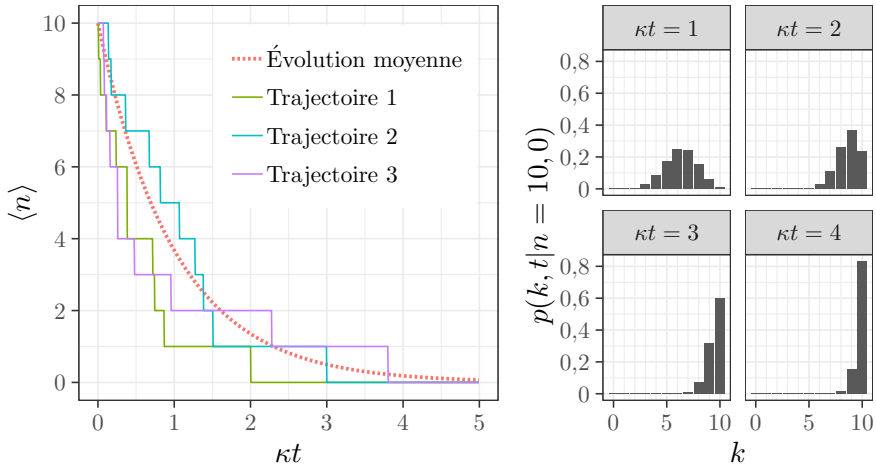


FIGURE 10.5 – Évolution d'un état de Fock avec $n = 10$ au cours du temps. Pour une trajectoire donnée, on a des sauts correspondant à la perte d'un photon, qui ont lieu à des instants aléatoires. Néanmoins, en moyenne sur toutes les trajectoires possibles, on retrouve simplement une évolution exponentiellement décroissante du nombre moyen de photons dans la cavité. Les histogrammes représentent la distribution de probabilité $p(k, t | n = 10, 0)$ d'avoir émis k photons sur l'intervalle $[0, t]$ sachant qu'il y avait 10 photons à l'instant initial. Comme attendu, plus on attend longtemps, et plus le nombre de photons émis augmente (tout en étant limité par la valeur $n = 10$).

l'énergie qui est emportée vers l'extérieur ? En premier lieu, soulignons qu'il n'y a aucun problème avec la conservation de l'énergie. En effet, comme l'état cohérent n'est pas un état propre de l'observable $\hbar\omega\hat{N}$ qui donne l'énergie du mode, parler de conservation de l'énergie n'a de sens qu'en valeur moyenne. Pour pouvoir en parler à l'échelle de la réalisation unique, il faudrait que le saut quantique se fasse entre deux états d'énergie bien définie, ce qui n'est pas le cas ici.

Dans le cas d'un état cohérent du champ, la décroissance de l'énergie moyenne décrite par l'équation (10.34) reflète l'acquisition d'information sur l'état entre deux sauts quantiques. En effet, en décomposant l'état $|\psi_{\text{nj}}(t|\alpha)\rangle$ obtenu à partir de l'état $|\alpha\rangle$ après une durée t sans saut quantique sur la base des états à nombre de photons fixé, nous voyons que :

$$|\psi_{\text{nj}}(t|\alpha)\rangle = |e^{-i\omega_0 t} e^{-\kappa t/2} \alpha\rangle. \quad (10.38)$$

C'est donc encore un état cohérent mais son paramètre est amorti exponentiellement, précisément de la manière qui correspond au mouvement d'un oscillateur harmonique amorti avec un taux de relaxation κ : $\alpha(t) = \alpha e^{-i\omega_0 t} e^{-\kappa t/2}$. Cette discussion illustre la nature bayésienne des états quantiques qui véhiculent

toute l'information dont on dispose sur le processus de préparation du système. Dans le cas présent, si on ne voit rien sortir de la cavité, cela nous apprend qu'il y a moins de photons présents que ce que nous pensions. C'est exactement ce qui est exprimé par l'équation (10.38).

La statistique de comptage du nombre de photons émis sur l'intervalle temporel $[0, t]$ s'effectue suivant les mêmes étapes que pour un état à nombre de photons fixés mais donne un résultat radicalement différent. On obtient une statistique poissonnienne dont la moyenne est donnée par la conservation de l'énergie moyenne :

$$p(k, t|\alpha, 0) = e^{-\bar{n}_\alpha(t)} \frac{(\bar{n}_\alpha(t))^k}{k!}, \quad (10.39)$$

où $\bar{n}_\alpha(t) = |\alpha|^2 - |\alpha(t)|^2 = |\alpha|^2(1 - e^{-\kappa t})$. Ce résultat est à comparer avec la statistique binomiale (10.36) qui correspond à une statistique de comptage pour des particules en nombre fixé n s'échappant d'une boîte. Ces résultats sont résumés sur la figure 10.6

On peut cependant remarquer que la loi de Poisson (10.39) au temps $t > 0$ s'obtient en prenant la convolution de la distribution en photon de l'état cohérent préparé au temps $t = 0$ par la loi binomiale (10.36) avec $p = 1 - e^{-\kappa t}$:

$$p(k, t|\alpha, 0) = \sum_{n \geq k} e^{-n} \frac{|\alpha|^{2n}}{n!} \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k}, \quad (10.40)$$

ce qui permet de comprendre le processus de relaxation d'un état cohérent en termes classiques sur les états à nombre de photons fixés. Cela n'est pas surprenant car, comme un miroir semi-réfléchissant ne change pas le nombre de photons, la détection d'un photon en aval de ce dernier peut toujours être décrite en termes de la détection d'un photon en entrée.

Pour une condition initiale $|\alpha\rangle$, nous obtenons à une phase près un état conditionné qui est le même pour toutes les trajectoires quantiques, c'est-à-dire l'état cohérent $|\alpha(t)\rangle$. Ceci montre que lorsque la cavité est initialement peuplée avec un état cohérent, elle ne s'intrique pas avec son environnement. Cette propriété remarquable des états cohérents exprime leur robustesse par rapport au couplage entre la cavité et son environnement. Ces états ont été appelés des *états pointeurs* (approchés) par Zurek et sont cruciaux pour comprendre comment émerge un comportement classique au sein d'un monde totalement quantique [394].

Décohérence d'une superposition d'états cohérents

Pour finir cette discussion de la dynamique de l'état d'un mode du champ en présence de fuite de photons, considérons comme état la superposition quantique de deux états cohérents.

Lors d'un saut quantique, un état cohérent est modulé par une phase qui est exactement la phase de son paramètre (équation (10.37)). Une superposition

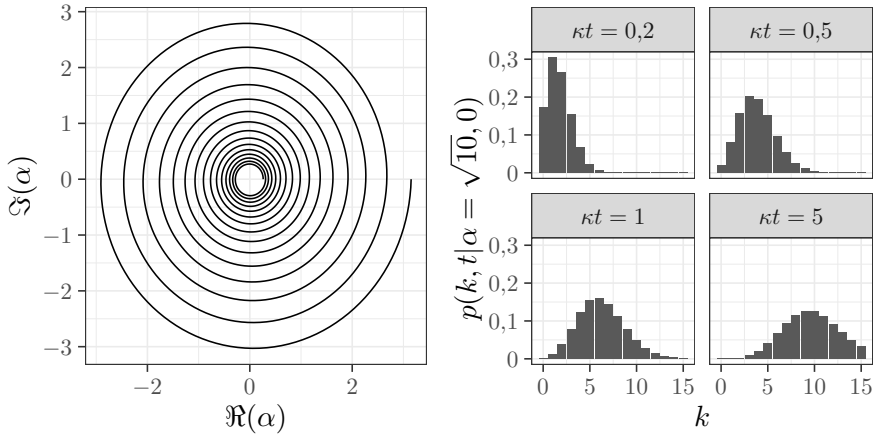


FIGURE 10.6 – Évolution d'un état cohérent avec $\alpha = \sqrt{10}$ au cours du temps. L'évolution de cet état est exactement identique à celle d'un oscillateur harmonique amorti classique. À droite, on représente la distribution de probabilité $p(k, t | \alpha = \sqrt{10}, 0)$ d'avoir émis k photons sur l'intervalle $[0, t]$ sachant qu'on était dans cet état cohérent à l'instant initial. Cette distribution tend au cours du temps vers la distribution de photons dans l'état cohérent initial.

de deux états cohérents qui ne se recouvrent pratiquement pas et dont les paramètres diffèrent par une phase se transforme donc en une superposition avec une phase relative entre les deux termes

$$\frac{e^{i\arg(\alpha)}|\alpha\rangle + e^{i(\theta+\arg(\alpha))}|e^{i\theta}\alpha\rangle}{\sqrt{2}}. \quad (10.41)$$

Comme la différence de phase entre les paramètres des deux composantes cohérentes de la superposition reste constante dans l'évolution entre deux sauts quantiques, nous voyons que le long d'une trajectoire avec p sauts, l'état quantique final est

$$\frac{|\alpha(t)\rangle + e^{ip\theta}|\alpha(t)e^{i\theta}\rangle}{\sqrt{2}}. \quad (10.42)$$

Cette fois, l'évolution en présence de fuites de photons a engendré un ensemble d'états qui diffèrent du fait de la phase relative $e^{ip\theta}$. C'est le signe que l'état du champ s'est intriqué avec les degrés de liberté extérieurs à la cavité. Lorsque l'on ignore l'information sur la fuite des photons, l'opérateur densité, obtenu en moyennant sur toutes les trajectoires, vaut :

$$\rho(t) = \frac{1}{2}(|\alpha(t)\rangle\langle\alpha(t)| + |e^{i\theta}\alpha(t)\rangle\langle e^{i\theta}\alpha(t)|) \quad (10.43a)$$

$$+ \frac{1}{2}(\mathcal{D}(t)|\alpha(t)\rangle\langle e^{i\theta}\alpha(t)| + \mathcal{D}(t)^*|e^{i\theta}\alpha(t)\rangle\langle\alpha(t)|). \quad (10.43b)$$

Le coefficient de décohérence $\mathcal{D}(t)$ est donné par

$$\mathcal{D}(t) = \overline{e^{iN[0,t]\theta}}. \quad (10.44)$$

où la moyenne porte sur la statistique du nombre de sauts quantiques $N[0, t]$ sur l'intervalle temporel $[0, t]$. Or, pour une superposition de la forme considérée ici, la probabilité d'avoir un saut quantique durant le temps dt est $dp = |\alpha|^2 \kappa [1 + \Re(\langle \alpha | e^{i\theta} \alpha \rangle)] dt$, qui n'est autre que $|\alpha|^2 \kappa dt$ car on a supposé que l'on pouvait considérer les états $|\alpha\rangle$ et $|\alpha e^{i\theta}\rangle$ comme orthogonaux. De fait, la moyenne dans l'équation (10.44) peut se calculer en utilisant la statistique de comptage pour les sauts quantiques dans le cas de l'état cohérent $|\alpha\rangle$. Cela nous donne :

$$\begin{aligned} \mathcal{D}(t) &= \overline{e^{iN[0,t]\theta}} = \sum_{k=0}^{\infty} p(k, t | \alpha, 0) e^{ik\theta} = e^{-\bar{n}_\alpha(t)} \sum_{k=0}^{\infty} \frac{(\bar{n}_\alpha(t) e^{i\theta})^k}{k!} \\ &= \exp(\bar{n}_\alpha(t) (e^{i\theta} - 1)). \end{aligned} \quad (10.45)$$

Le module du coefficient de décohérence décroît donc comme $e^{-d(t)}$ où

$$d(t) = 2 \bar{n}_\alpha(t) \sin^2 \theta / 2 = 2 |\alpha|^2 (1 - e^{-\kappa t}) \sin^2 \theta / 2 \quad (10.46)$$

en utilisant l'équation (10.34) pour la deuxième égalité. Cette décroissance est exponentielle aux temps courts avec un temps caractéristique donné par

$$\tau_\theta = \frac{\kappa^{-1}}{2 |\alpha|^2 \sin^2 \theta / 2}. \quad (10.47)$$

Le dénominateur n'est autre que le carré de la distance, mesurée dans le plan complexe, entre les paramètres des deux états cohérents. L'échelle de temps κ^{-1} au numérateur est l'échelle de temps de la relaxation. Cette expression explique pourquoi deux états cohérents assez séparés pour ne pas se recouvrir vont décohérer sur un temps court devant leur temps caractéristique de relaxation κ^{-1} . La figure 10.7 montre l'évolution d'une superposition d'états cohérents au cours du temps et le phénomène de décohérence qui s'en suit. Ce fait est très général : une superposition quantique d'états pointeurs perd sa cohérence en un temps court devant le temps d'évolution de chacun des membres de la superposition. Ici, l'échelle de temps de la décohérence est le temps d'émission du premier photon $\kappa^{-1} / |\alpha|^2$ qui suit la densité de probabilité donnée par $p(0, \tau) = |\alpha|^2 \kappa e^{-\kappa \tau} e^{-|\alpha|^2 (1 - e^{-\kappa \tau})}$.

10.1.5 Études expérimentales de la décohérence

L'étude de la relaxation de la cavité peuplée par un nombre de photons bien défini a été rendue possible par la mise au point de la méthode de mesure non destructive du nombre de photons exposée à la section 10.1.2. En effet,

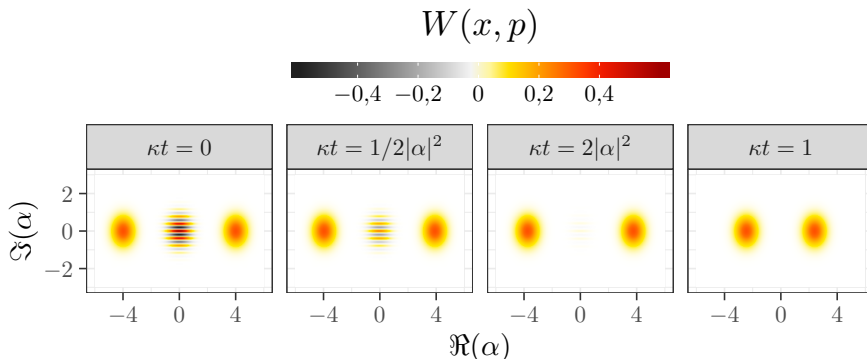


FIGURE 10.7 – Évolution d’une superposition d’états cohérents de nombre de photons moyen $|\alpha|^2 = 16$ (en opposition de phase $\theta = \pi$) en représentation de Wigner. Au cours du temps, les deux gaussiennes, correspondant à chacun des membres de la superposition initialement centrés en $\alpha = \pm 4$, se rapprochent de l’origine sur une échelle de temps $1/\kappa$. La figure d’interférence, au centre, indique la cohérence entre chacun des états cohérents. Celle-ci disparaît sur une échelle de temps $1/2\kappa|\alpha|^2$.

elle permet de préparer, de manière non déterministe, un état à nombre de photons fixé puis de mesurer celui-ci au fil du temps en envoyant des trains successifs d’atomes. Il fut alors possible d’observer les sauts quantiques de l’état du champ dans la cavité [178] puis, par une étude plus quantitative de l’évolution de l’état du champ dans la base de Fock [75], de confirmer la validité de la description par l’équation de Lindblad (10.33).

La décohérence d’une superposition quantique de deux états cohérents a été observée *via* le processus de tomographie décrit à la section 10.1.3. Une première étude quantitative du processus de décohérence avait été menée vingt ans auparavant [76] au moyen des corrélations entre deux atomes traversant la cavité en régime dispersif. Le passage du premier atome puis sa mesure créent un état superposé du champ dans la cavité. Après décohérence partielle, le passage du second atome, lui aussi en régime dispersif, permet de récupérer l’information sur la décohérence subie par l’état quantique du champ entre le passage des deux atomes. Cette expérience a confirmé quantitativement le modèle de décohérence exposé dans le précédent paragraphe.

L’observation par tomographie [126] a permis d’observer de manière complète la décohérence d’un état chat de Schrödinger du champ électromagnétique et donc d’illustrer de manière spectaculaire la dynamique en temps réel d’un système quantique ouvert. Elle montre comment les protocoles de mesure, tout comme les effets de décohérence, proviennent de l’intrication entre le système considéré et son environnement, qu’il s’agisse d’un appareil de mesure ou de degrés de liberté non observés.

Ce type de travaux a ouvert la voie au développement d'une véritable ingénierie quantique qui vise à préparer, manipuler et caractériser l'état quantique de systèmes à l'échelle de la réalisation unique. Un enjeu important étant de stabiliser certains états quantiques non classiques en vue des applications, les expériences récentes menées sur ces systèmes ont porté sur le contrôle quantique, c'est-à-dire la mise en œuvre de protocoles de rétro-action sur l'état du champ au moyen d'actuateurs classiques [320] ou quantiques [387].

De telles manipulations sont maintenant possibles sur des circuits électroniques qui, *a priori*, sont nettement plus complexes qu'un mode du champ électromagnétique et un atome. L'avantage de ces dispositifs réside dans leur procédé de fabrication et leur miniaturisation qui permet d'envisager le développement de circuits électroniques quantiques comportant un grand nombre de composants.

10.2 Électrodynamique quantique sur circuit

Les expériences d'électrodynamique sur circuit reposent sur des circuits électroniques supraconducteurs dont le comportement est dominé par les lois de la théorie quantique. Le développement de ces dispositifs constituent un progrès majeur des vingt dernières années : leur simplicité d'utilisation permet de manipuler finement l'état quantique du circuit. L'un de leurs atouts majeurs est de pouvoir simuler de nombreux modèles importants de la physique de la matière condensée. Ils ont donc ouvert un nouveau champ de possibilités, y compris pour développer de nouvelles expériences de physique fondamentale. Avant d'étudier la dynamique de ce type de système quantique ouvert, il faut comprendre comment les supraconducteurs ont permis de fabriquer ces dispositifs quantiques électroniques et quels types de qubits ont pu être construits à partir de là.

10.2.1 Circuits électriques quantiques

Dans le cadre de la théorie classique de Maxwell, les matériaux sont décrits au travers de fonctions de réponse qui donnent la densité de charge et de courant en fonction des champs appliqués. En pratique, calculer microscopiquement ces fonctions de réponse nécessite de comprendre les propriétés quantiques du matériau. Il existe cependant des formes phénoménologiques, comme la conductivité de Drude, étudiées bien avant l'invention de la théorie quantique [388, 341]. Cette théorie de la réponse locale des matériaux est pertinente pour une antenne radio ou pour de petits circuits opérés à des fréquences très élevées, c'est-à-dire lorsque la taille du dispositif est comparable à la longueur d'onde du rayonnement électromagnétique ou des ondes de densité de charge qui s'y propagent. En revanche, à plus basse fréquence, les approximations des régimes quasi stationnaires pour le champ électromagnétique et la description du circuit en éléments discrets (résistances, capacités,

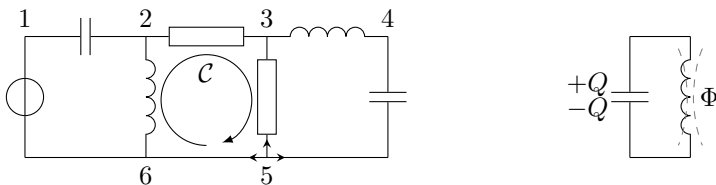


FIGURE 10.8 – À gauche : représentation d'un circuit électrique en termes d'éléments discrets. Les deux lois principales sont la loi des mailles et la loi des nœuds. À droite : circuit LC décrit en termes du flux Φ à travers l'inductance et de la charge Q stockée sur une armature du condensateur.

inductances) conviennent parfaitement. Cela revient à décrire le circuit en termes de coordonnées collectives qui fournissent une description effective du système électronique [385, 223]. Étonnamment, cette description classique du circuit reste pertinente dans le cas où il est assez isolé de son environnement pour que les lois de la théorie quantique s'y appliquent. La procédure de quantification canonique (section 1.3.3) fournit alors une description de la dynamique quantique du circuit en termes de ses variables collectives.

Pour cela, il nous faut préciser quelles sont les variables pertinentes pour décrire un circuit électrique (figure 10.8). Deux suffisent : le courant $I_{nm}(t)$ traversant une branche orientée $\alpha = [nm]$ ainsi que la tension $V_\alpha(t)$ que l'on peut écrire comme la différence du potentiel en chaque nœud $V_\alpha(t) = V_m(t) - V_n(t)$. Les lois élémentaires des circuits, dues à Kirchhoff, expriment, d'une part, qu'il n'y a pas d'accumulation de charge dans le circuit et, d'autre part, que la différence de tension le long d'un chemin fermé est nulle (pas de flux magnétique dépendant du temps à travers les boucles du circuit) : $\sum_\alpha I_{nm}(t) = 0$ et $\sum_\alpha V_{nm}(t) = 0$, où $\mathcal{C}$ est un chemin fermé sur le circuit électrique. Ces relations peuvent s'exprimer en termes de quantités qui sont des charges et des flux associés à chacune des branches orientées : $Q_\alpha(t) = \int_{-\infty}^t I_\alpha(t') dt'$ et $\Phi_\alpha(t) = \int_{-\infty}^t V_\alpha(t') dt'$.

Considérons maintenant le circuit LC de la figure 10.8. La variable de flux Φ à travers l'inductance décrit ce circuit car la tension V aux bornes de la capacité et de l'inductance n'est autre que $\dot{\Phi}$. Cela suggère de prendre comme lagrangien la différence de l'énergie électrostatique de la capacité et de l'énergie magnétique de l'inductance :

$$L(\dot{\Phi}, \Phi) = \frac{C\dot{\Phi}^2}{2} - \frac{\Phi^2}{2L}. \quad (10.48)$$

Les équations du mouvement sont celles d'un oscillateur harmonique de pulsation $\omega_0 = 1/\sqrt{LC}$. Le moment conjugué à Φ est $\Pi_\Phi = C\dot{\Phi}$, c'est-à-dire la charge Q stockée sur une armature du condensateur. On en déduit que le

hamiltonien décrivant ce circuit est un hamiltonien d'oscillateur harmonique ⁶

$$H = \frac{Q^2}{2C} + \frac{\Phi^2}{2L} \quad (10.49)$$

exprimé en termes des variables conjuguées Q et Φ . Par quantification canonique, les quantités conjuguées Q_α et Φ_α deviennent des opérateurs qui vérifient les relations de commutation canoniques

$$[Q_\alpha, \Phi_\beta] = -i\hbar \delta_{\alpha,\beta} \mathbb{1}, \quad (10.50)$$

où nous avons pris soin d'orienter chaque arête du circuit. Si on suppose que le courant est porté par des charges q , il est utile d'introduire le nombre de porteurs correspondants $N_\alpha = Q_\alpha/q$. Si on pose $\Phi_\alpha = \hbar\varphi_\alpha/q$ alors N_α et φ_α obéissent à la relation $[\varphi_\alpha, N_\beta] = i\delta_{\alpha,\beta} \mathbb{1}$.

Même si ces considérations sont pertinentes pour des conducteurs métalliques, elles prennent tout leur intérêt dans le cas des supraconducteurs. En effet, l'état supraconducteur se caractérise par une rigidité qui permet de décrire l'état du fluide électronique au moyen de variables collectives dont la dynamique est, dans les bonnes conditions, quantique.

10.2.2 Circuits supraconducteurs simples

Les supraconducteurs

La supraconductivité est un état possible pour le fluide électronique dans un conducteur. Sous l'effet de l'interaction entre électrons et vibrations du réseau cristallin, le fluide électronique entre dans une phase où son état fondamental est séparé de ses états excités par un écart d'énergie ou *gap* Δ [30, 256]. Pour comprendre comment l'état supraconducteur diffère d'un métal habituel, commençons par étudier les excitations d'un tel état ainsi que sa réponse à un champ extérieur.

Contrairement aux excitations élémentaires dans un métal, il faut une énergie minimale supérieure à celle du *gap* pour créer des excitations dans un supraconducteur. De plus, alors que celles-ci ont bien une charge $\pm e$ à très haute énergie (c'est-à-dire grande devant Δ), celles qui ont une énergie minimale sont en fait neutres. Elles ne peuvent alors être identifiées à des excitations de type électron ou trou qu'à haute énergie. L'existence du *gap* a des conséquences importantes : le supraconducteur est rigide au sens où il ne répond pas facilement à une perturbation extérieure. Il est isolant tant qu'on ne lui applique pas une différence de potentiel de l'ordre de Δ/e ou un champ électrique de fréquence $\gtrsim \Delta/\hbar$. Dans la plupart des cas, Δ/k_B est de l'ordre de quelques degrés Kelvin. Rappelons que 1 K correspond à une énergie $k_B T$ de l'ordre de 86,2 μeV qui est l'énergie d'un photon de fréquence 20,8 GHz.

6. Dans l'analogie entre ce circuit et un oscillateur, un flux correspond à une position, une capacité à une masse, une inductance à une constante de raideur, une tension à une vitesse et un courant électrique, qui est le quotient d'un flux par une inductance, à une force.

À basse température ($T \ll \Delta/k_B$) et à basse fréquence ($f \ll \Delta/h$), la réponse du supraconducteur à des excitations extérieures comme des champs électromagnétiques reflète la capacité de l'état fondamental du fluide électronique à se « déformer », en modulant son paramètre d'ordre dans l'espace et dans le temps. On peut montrer que le supraconducteur répond à un champ électromagnétique classique extérieur par un courant électrique moyen donné par l'équation de London [261] :

$$\mathbf{j}_s(\mathbf{r}, t) = -\frac{\pi\sigma_n\Delta}{\hbar} \left(\mathbf{A} + \frac{\hbar}{2e} \nabla\varphi \right) (\mathbf{r}, t), \quad (10.51)$$

où σ_n désigne la conductivité du métal qui décrit sa réponse linéaire en courant au champ électrique appliqué (loi d'Ohm) : $\mathbf{j}_n(\mathbf{r}, t) = \sigma_n \mathbf{E}(\mathbf{r}, t)$. L'équation (10.51) nous apprend de plus que la phase du paramètre d'ordre φ n'est pas un invariant de jauge. En effet, puisque le courant électrique doit l'être, le changement $\mathbf{A}$ en $\mathbf{A} + \nabla\chi$ doit entraîner un changement de φ en $\varphi - 2e\chi/\hbar$. C'est ce que l'on attendrait si le paramètre d'ordre supraconducteur $\Delta e^{i\varphi}$ représentait une fonction d'onde collective pour des quanta portant la charge $-2e$. En effet, effectuer la transformation de jauge définie par les équations (1.20) altère la fonction d'onde de tels quanta de charge $-2e$ en la multipliant par $e^{-2ei\chi(\mathbf{x}, t)/\hbar}$ lorsqu'on l'évalue en $(\mathbf{x}, t)$ (équation (1.21) pour $q = -2e$). Ces considérations nous suggèrent donc de considérer le paramètre d'ordre complexe du supraconducteur, décrivant l'état fondamental, comme correspondant à une fonction d'onde $\psi(\mathbf{r}, t)$ partagée par un grand nombre de quanta de charge $-2e$ condensés dans un même état quantique macroscopique. Ceci conduit à écrire phénoménologiquement

$$\psi(\mathbf{r}, t) = \sqrt{n(\mathbf{r}, t)} e^{i\varphi(\mathbf{r}, t)}, \quad (10.52)$$

où $n(\mathbf{r}, t)$ représente la densité de ces quanta appelés *paires de Cooper*.

Dans un bloc supraconducteur à la topologie triviale, les variations spatiales de la phase peuvent donc être effacées par une transformation de jauge. Mais deux phases uniformes différentes ne correspondent pas au même état du supraconducteur. En effet, si on les réalisait sur deux parties d'un même bloc supraconducteur *via* des conditions au bord appropriées et si elles correspondaient au même état physique du supraconducteur, il n'y aurait pas de courant entre les deux parties. Or, cela contredirait l'équation (10.51) qui nous dit que tout gradient de phase induit un courant électrique. Pour un bloc supraconducteur de topologie triviale, les différents états du fluide électronique en phase supraconductrice sont donc indexés par une phase uniforme⁷. Dans la suite, la phase supraconductrice sera supposée homogène.

7. Cela permet de comprendre pourquoi les supraconducteurs brisent une symétrie $U(1)$: chaque état d'un bloc supraconducteur est décrit par une phase alors même que la dynamique du matériau et du champ électromagnétique préserve l'invariance de jauge et donc, en particulier, est symétrique par changement de phase global.

Avant de passer aux circuits électriques, montrons comment plusieurs effets caractéristiques des supraconducteurs peuvent être retrouvés à partir des éléments que nous avons déjà discutés. Le premier, appelé *effet Meissner*, est l'expulsion des champs magnétiques des supraconducteurs. En effet, dans la jauge de Lorenz ($\text{div} \mathbf{A} + \partial V/c^2 \partial t = 0$), l'équation satisfaite par le potentiel vecteur est une équation de propagation d'onde

$$\frac{\partial^2 \mathbf{A}}{\partial^2 t} - c^2 \Delta \mathbf{A} = \frac{\mathbf{j}}{\varepsilon_0}. \quad (10.53)$$

Dans un supraconducteur homogène, l'équation (10.51) implique que le courant électrique est proportionnel au potentiel vecteur. Lorsque l'on impose un champ magnétique constant hors du supraconducteur, ce terme d'amortissement dans l'équation (10.53) correspond à une longueur de pénétration $\lambda_M = \sqrt{\hbar c^2 \varepsilon_0 / \pi \Delta \sigma_n}$ dans ce dernier. Cet écrantage du champ magnétique à l'intérieur d'un supraconducteur plus épais que λ_M est dû aux courants électriques au voisinage de la surface.

L'équation (10.51) permet aussi de comprendre pourquoi le supraconducteur est parfait, au sens où il n'oppose pas de résistance au passage d'un courant électrique. En effet, la combinaison

$$V(\mathbf{r}, t) - \frac{\hbar}{2e} \frac{\partial \varphi(\mathbf{r}, t)}{\partial t} \quad (10.54)$$

est invariante de jauge. En utilisant $\mathbf{E} = -\nabla V - \partial_t \mathbf{A}$ dans l'équation (10.51), nous voyons alors que

$$\frac{\partial \mathbf{j}_s}{\partial t} = \frac{\pi \sigma_n \Delta}{\hbar} \left(\mathbf{E} + \nabla \left[V - \frac{\hbar}{2e} \frac{\partial \varphi}{\partial t} \right] \right). \quad (10.55)$$

Dans un supraconducteur homogène placé dans un champ électrique extérieur uniforme et constant, on s'attend à ce que la fonction (10.54) ne dépende ni de la position ni du temps. On obtient alors $\partial_t \mathbf{j}_s(\mathbf{r}, t) \propto \mathbf{E}$. C'est exactement la manière dont répondrait un fluide de particules chargées placées dans un champ électrique en l'absence de tout frottement : l'accélération donnée par le champ électrique serait constante et le courant croîtrait linéairement avec le temps.

Placer un supraconducteur dans un circuit électrique

Une dernière conséquence de l'équation de London (10.51) est la quantification du flux magnétique passant à travers un anneau supraconducteur plus épais que λ_M . En effet, considérons un anneau supraconducteur dans un champ magnétique $\mathbf{B}$ uniforme et constant, et faisons un tour autour de l'anneau en passant dans le matériau, c'est-à-dire là où $\mathbf{j}_s = 0$. Le long de ce tour, la phase φ varie de $-(2e/\hbar) \times \Phi_{\mathbf{B}}$, où $\Phi_{\mathbf{B}}$ est le flux magnétique perçant l'anneau. Or le saut de phase dans un tour complet est nécessairement un

multiple entier de 2π afin que la fonction d'onde correspondante soit continue. On en déduit donc que $\Phi_{\mathbf{B}}$ est quantifiée par un multiple entier de $\Phi_0 = h/2e$, que l'on appelle le *quantum de flux* supraconducteur.

Considérons maintenant un circuit formé d'éléments discrets. Les bonnes variables sont alors le nombre de porteurs de charges N_α associé à une arête et la phase φ_α conjuguée. On peut dès lors facilement comprendre comment prendre en compte des éléments discrets supraconducteurs dans un tel circuit. Chaque nœud n du circuit est caractérisé par sa phase supraconductrice φ_n . Si on met un composant supraconducteur sur une arête connectant deux nœuds supraconducteurs n et m , alors les phases des nœuds et le flux Φ_{nm} seront reliés par

$$\Phi_{nm} = \frac{\hbar}{2e}(\varphi_m - \varphi_n). \quad (10.56)$$

Il est donc possible d'utiliser les phases φ_n comme coordonnées pour décrire la dynamique du circuit, sachant encore une fois qu'elles ne sont pas indépendantes entre elles et qu'il faudra donc éliminer les redondances. Comme dans un supraconducteur les phases correspondent à la phase d'une fonction d'onde macroscopique pour les porteurs de charge $q = -2e$, l'équation du mouvement qui découle de la procédure de quantification correspond à l'équation d'évolution pour cette fonction d'onde macroscopique. L'idée des circuits supraconducteurs consiste à exploiter pleinement ce résultat de manière à constituer un système quantique bien spécifique réalisant la dynamique quantique qui nous intéresse.

La jonction Josephson

Nous pouvons maintenant étudier la *jonction Josephson*, composant principal de l'électronique supraconductrice permettant de construire des atomes artificiels. L'architecture d'une jonction Josephson est simple : elle se construit à partir d'un matériau isolant très fin entre deux morceaux d'un supraconducteur de *gap* Δ , caractérisés par des phases différentes φ_1 et φ_2 comme indiqué sur la figure 10.9.

Sous réserve que l'on n'applique pas une différence de potentiel supérieure à Δ/e entre les deux, la seule possibilité de transfert de charge à travers la jonction est le passage d'un quanta de charge $-2e$, c'est-à-dire d'une paire de Cooper, d'un supraconducteur à l'autre. Dans un passage de 1 vers 2, la phase varie de $\varphi_2 - \varphi_1$. Or, comme $[\varphi_{21}, N_{21}] = i\mathbb{1}$, nous voyons que $e^{i\varphi_{21}} N_{12} e^{-i\varphi_{21}} = N_{12} + \mathbb{1}$. L'opérateur $e^{i\varphi_{21}}$ correspond donc au transfert d'une paire à travers l'arête orientée $\alpha = (12)$. Le hamiltonien décrivant la physique de la jonction doit donc contenir les contributions, hermitiennes conjuguées, associées aux processus de passage dans un sens et dans l'autre. Cela conduit à écrire une première contribution au hamiltonien du circuit sous la forme

$$\mathcal{A}_{12} e^{-i\varphi_{12}} + \mathcal{A}_{12}^* e^{i\varphi_{12}}, \quad (10.57)$$

où $\mathcal{A}_{12}$ est une amplitude, homogène à une énergie, qui décrit les propriétés de transmission quantique de la jonction pour les paires de Cooper. Cette amplitude a un module carré valant E_J , où E_J est appelée l'énergie Josephson de la jonction. Le signe de ces amplitudes va être choisi négatif de sorte que $\varphi_{12} = 0$ est bien un minimum d'énergie lorsque l'énergie Josephson croît. Physiquement, cette limite correspond à rendre la jonction Josephson de plus en plus transparente. Or, à la limite où elle n'existe plus, la différence de phase doit s'annuler car on serait dans un supraconducteur homogène. Il est donc logique de choisir $\mathcal{A}_{12} = \mathcal{A}_{12}^* = -\sqrt{E_J}$.

Par ailleurs, pour une jonction réaliste, on doit prendre en compte le fait que sa géométrie (deux conducteurs face à face) forme un condensateur de capacité C_J (figure 10.9). Dans la théorie des circuits, il convient donc d'ajouter la contribution de l'énergie électrostatique liée à cette capacité. Celle-ci ne dépend que de la différence des potentiels aux bornes $V_{12} = \Phi_1 - \Phi_2$, ce qui donne un terme cinétique de la forme suivante au lagrangien du circuit :

$$\frac{Q}{2C_J} = \frac{C_J \dot{\Phi}_{12}^2}{2} = \frac{\hbar^2 C_J}{8e^2} \dot{\varphi}_{21}^2. \quad (10.58)$$

Si on passe en formalisme hamiltonien et que l'on introduit maintenant N_{12} conjugué de φ_{12} , la contribution de la jonction au hamiltonien du circuit est donc

$$H_J = \frac{2e^2}{C_J} N_{12}^2 - E_J \cos \varphi_{12}, \quad (10.59)$$

où on reconnaît $E_C = (2e)^2/2C_J$ qui est l'énergie correspondant à la charge d'une paire de Cooper stockée dans la capacité C_J . On peut calculer le courant électrique $I_{12} = -2e\partial_t N_{12}$ qui traverse la jonction dans le sens $1 \rightarrow 2$ au moyen du hamiltonien (10.59). Cela donne

$$I_{12} = -2e \frac{E_J}{\hbar} \sin \varphi_{12}. \quad (10.60)$$

Le courant est donc une fonction non linéaire du flux magnétique associé à la jonction $\Phi_{12} = -\hbar\varphi_{12}/2e$. Cette non-linéarité de la jonction Josephson est un point clé dans la réalisation des qubits supraconducteurs. Pour les petites valeurs du flux, la relation devient linéaire et la jonction se comporte comme une bobine d'inductance $L_J = \hbar^2/4e^2 E_J$.

Les jonctions habituellement utilisées dans les micro-circuits supraconducteurs qui servent à fabriquer des qubits ont une taille micrométrique et une épaisseur nanométrique. La capacité est donc au minimum de l'ordre de quelques femtofarads et l'énergie de Coulomb d'une paire de Cooper correspond à une fréquence au plus dans la gamme des dizaines de gigahertz, souvent moins. L'énergie Josephson est obtenue comme une fonction du courant $I_0 = 2eE_J/\hbar$, du *gap* supraconducteur Δ et de la résistance de la jonction R_T en régime normal :

$$I_0 = \frac{\pi}{2} \frac{\Delta}{eR_T}. \quad (10.61)$$

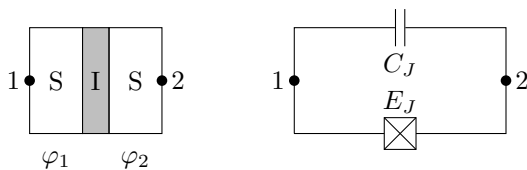


FIGURE 10.9 – À gauche : jonction Josephson formée d'une fine couche isolante entre deux supraconducteurs caractérisés par leur phase. À droite : circuit équivalent formé d'une jonction Josephson idéale en parallèle avec une capacité C_J .

En pratique, suivant la géométrie du qubit et sa réalisation physique, l'énergie Josephson E_J sera comparable à l'énergie de Coulomb. Notons qu'en augmentant la taille de la jonction, on ouvre plus de canaux de conduction, ce qui augmente $1/R_T$ et donc l'énergie Josephson. Par contre, augmenter la taille augmente aussi la capacité C_J et diminue donc l'énergie coulombienne. Le ratio E_J/E_C est donc une fonction croissante de l'aire de la jonction. Remarquons enfin qu'à ce stade, notre description ne prend en compte aucun effet dissipatif. Elle n'est donc valable que dans la mesure où nous ne créons pas d'excitation et que l'état du matériau est uniquement caractérisé par sa phase supraconductrice.

Jonction Josephson polarisée en tension

Regardons comment une jonction Josephson s'intègre dans un circuit électrique. Couplons un générateur de tension imposant une différence de potentiel $V_0(t)$ aux bornes de la jonction. Les deux bornes du générateur de tension sont connectées aux deux supraconducteurs de phases respectives φ_1 et φ_2 . En créant une différence de potentiel entre les deux supraconducteurs, ces porteurs se retrouvent décalés en énergie. Ceci implique que l'on force ces phases à ne pas tourner à la même vitesse. En conséquence :

$$\frac{d\varphi_{12}}{dt} = -\frac{2e}{\hbar} V_0(t). \quad (10.62)$$

Cette relation, due à Josephson, donne l'évolution de la phase d'une jonction Josephson polarisée en tension. Ainsi, une tension constante donne une différence de phase qui croît linéairement avec le temps $\varphi_{12}(t) = \varphi_0 - 2eV_0t/\hbar$. La relation (10.60) implique alors que la jonction est traversée par un courant oscillant $I(t) = I_c \sin(\varphi_0 - 2eV_0t/\hbar)$ qui dépend de la condition initiale φ_0 . Cet effet surprenant est appelé l'effet Josephson alternatif. La relation entre la fréquence de ce courant et la tension permet de mesurer le ratio $e/\hbar$, ou encore de relier le standard de temps au standard de tension par conversion d'une tension vers une fréquence. En pratique, la fréquence d'oscillation de ce courant est très élevée (484 GHz par mV appliqué).

Remarquons qu'à tension nulle, la jonction peut voir passer un courant continu $I = I_c \sin(\varphi_0)$ compris entre $-I_c$ et I_c . C'est l'effet Josephson continu qui est le pendant au niveau de la jonction du fait que, comme indiqué par l'équation de London (10.51), toute différence de phase supraconductrice induit un courant.

Jonctions Josephson en parallèle

Continuons à complexifier le circuit, et considérons maintenant deux jonctions Josephson, caractérisées par leur énergie E_{J_α} et leur capacité C_α ($\alpha = 1, 2$), montées en parallèle. Dans un tel montage, représenté sur la figure 10.10, on crée donc une boucle ayant sa propre auto-inductance L_S . Cependant, par souci de simplicité, nous allons la négliger [98]. Cette approximation a pour effet de contraindre les sauts de phase φ_2 et φ_1 : la différence $\varphi_1 - \varphi_2$ est reliée au flux magnétique $\Phi_B = \Phi_0 \varphi_B / 2\pi$ qui traverse la boucle par $\varphi_{1,2} = \varphi \pm \varphi_B / 2$. Le saut de phase moyen φ est donc la coordonnée décrivant l'ensemble des deux jonctions en parallèle. À champ magnétique constant, la différence de potentiel aux bornes de circuit est $-\frac{\hbar}{2e} \dot{\varphi}(t)$. Nous pouvons alors écrire le hamiltonien

$$H = \frac{C_1 + C_2}{2} \left(\frac{\hbar}{2e} \right)^2 \dot{\varphi}^2 - E_{J_1} \cos \left(\varphi + \frac{\varphi_B}{2} \right) - E_{J_2} \cos \left(\varphi - \frac{\varphi_B}{2} \right) \quad (10.63)$$

qui peut se réécrire comme le hamiltonien d'une jonction Josephson avec paramètres

$$C_{J_{\text{tot}}} = C_{J_1} + C_{J_2}, \quad (10.64a)$$

$$E_{J_{\text{tot}}} = \sqrt{E_{J_1}^2 + E_{J_2}^2 + 2E_{J_1}E_{J_2} \cos \varphi_B} \quad (10.64b)$$

pour un saut de phase effectif $\varphi - \phi(\varphi_B)$ donné par

$$\tan \phi(\varphi_B) = \frac{E_{J_1} - E_{J_2}}{E_{J_1} + E_{J_2}} \tan \frac{\varphi_B}{2}. \quad (10.65)$$

Ce décalage du saut de phase n'est présent que si les deux jonctions sont différentes et n'est observable qu'en plongeant l'ensemble des deux jonctions dans un circuit présentant une boucle.

Le point important est que l'on peut moduler l'énergie Josephson effective totale $E_{J_{\text{tot}}}$ au moyen d'un flux magnétique. Cette modulation provient de l'interférence quantique entre les deux chemins possibles, à travers l'une ou l'autre des jonctions, pour les paires de Cooper et ne peut donc être expliquée par un raisonnement classique. En effet, même dans le régime linéaire, les inductances associées aux jonctions Josephson ne se composent pas classiquement : $E_{J_{\text{tot}}} \neq E_{J_1}E_{J_2}/(E_{J_1} + E_{J_2})$. En revanche, l'addition des capacités en parallèle est conforme aux lois classiques de composition des impédances. Un

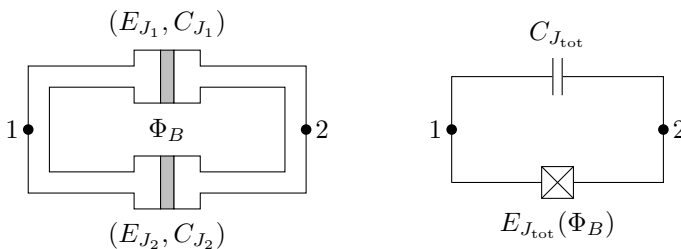


FIGURE 10.10 – À gauche : circuit formé par deux jonctions Josephson en parallèle qui délimitent une boucle traversée par un flux magnétique Φ_B . À droite : représentation équivalente finale sous la forme d'une jonction Josephson d'énergie donnée par l'équation (10.64b) et d'une capacité $C_{J_{tot}} = C_{J_1} + C_{J_2}$.

tel montage en parallèle de deux jonctions Josephson est appelé un SQUID (*superconducting quantum interference device*). Ce dispositif est au cœur des différents systèmes quantiques que l'on peut construire avec des circuits électriques et possède des applications technologiques comme magnétomètre par exemple.

10.2.3 Créer des qubits supraconducteurs

Disposer, à partir d'un circuit électrique quantique, d'un système harmonique est chose facile à l'aide d'un montage LC. Cependant, il n'est pas possible de construire un système à deux niveaux manipulable seulement à partir de condensateurs et de bobines puisqu'ils ne permettent que de fabriquer un circuit harmonique dont les écarts entre niveaux seront fortement dégénérés. Adresser des transitions de manière sélective nécessite donc des composants non linéaires et la jonction Josephson remplit parfaitement ce rôle.

Dans un dispositif contenant ces jonctions, la physique va dépendre de la compétition entre les différents termes présents dans le hamiltonien (équation (10.60)). Les capacités tendent à fixer le nombre N de paires de Cooper et à délocaliser la phase φ . De son côté, un terme d'énergie potentielle tend à fixer la phase et à délocaliser la charge. Le terme quadratique lié à l'inductance L confine cependant les fluctuations de la phase sur $(\Delta\varphi)_L/2\pi \simeq \sqrt{2L/C_J}/(h/4e^2)$.

C'est en jouant sur ces différents termes que sont construits les différents types de qubits développés en laboratoire. L'objectif est de réaliser des systèmes quantiques comportant quelques niveaux d'énergie (si possible deux pour avoir un qubit) dont l'évolution peut être contrôlée et qui, si possible, sont le moins sensibles aux sources de décohérence.

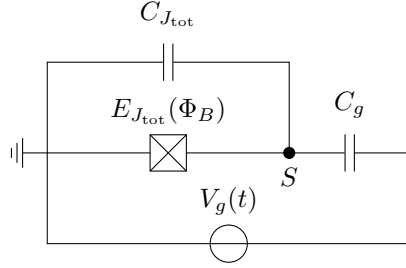


FIGURE 10.11 – Circuit électrique équivalent au qubit en configuration boîte à paires de Cooper, formé d'un ilot supraconducteur S connecté à la masse *via* un SQUID et couplé capacitivement à une électrode portée à la tension $V_g(t)$.

La boîte à paires de Cooper

Le premier type de système est basé sur un SQUID mis en série avec une capacité, que nous allons polariser avec une source de tension. Le circuit électrique correspondant est représenté sur la figure 10.11. Dans ce cas, la charge $Q_S(t)$ stockée au niveau du SQUID est la somme des charges de polarisations dues aux deux capacités C_J et à la capacité C_g :

$$Q_S(t) = -\frac{\hbar}{2e}C_{\text{tot}}\dot{\varphi} - C_g V_g(t), \quad (10.66)$$

où $C_{\text{tot}} = 2C_{J_{\text{tot}}} + C_g$, et φ est la phase donnée par l'équation (10.65). On peut alors réécrire le hamiltonien en termes de $N = -Q_S/2e$, qui vérifie $[\varphi, N] = i\mathbb{1}$. En introduisant $n_g(t) = -C_g V_g(t)/2e$, on obtient

$$H = E_C(N - n_g(t))^2 - E_J(\Phi_B) \cos \varphi, \quad (10.67)$$

où $n_g(t)$ s'interprète comme le nombre de paires de Cooper favorisées par le biais en tension.

Dans le cas indépendant du temps ($n_g(t) = n_g$ et $E_J(\Phi_B) = E_J$), nous allons chercher les états propres de ce hamiltonien dans la base de phase, c'est-à-dire en utilisant des fonctions d'ondes $\psi(\varphi)$, qui sont 2π -périodiques. On peut éliminer n_g en transformant la fonction d'onde $\chi(\varphi) = e^{-in_g\varphi}\psi(\varphi)$. Dans ce cas, les états d'énergie stationnaire sont donnés par des fonctions d'ondes $\chi(\varphi)$ qui obéissent à l'équation de Schrödinger indépendante du temps

$$-\frac{d^2\chi}{d\varphi^2} - \frac{E_J}{E_C} \cos \varphi \chi = \frac{E}{E_C} \chi, \quad (10.68)$$

avec la condition de quasi-périodicité $\chi(\varphi + 2\pi) = e^{-2\pi in_g} \chi(\varphi)$. Cette équation est connue dans la littérature sous le nom d'équation de Mathieu⁸ et a été

8. La forme canonique de l'équation de Mathieu est donnée par :

$$\frac{\partial^2 f}{\partial x^2} + (a - 2q) \cos(2x) f(x) = 0. \quad (10.69)$$

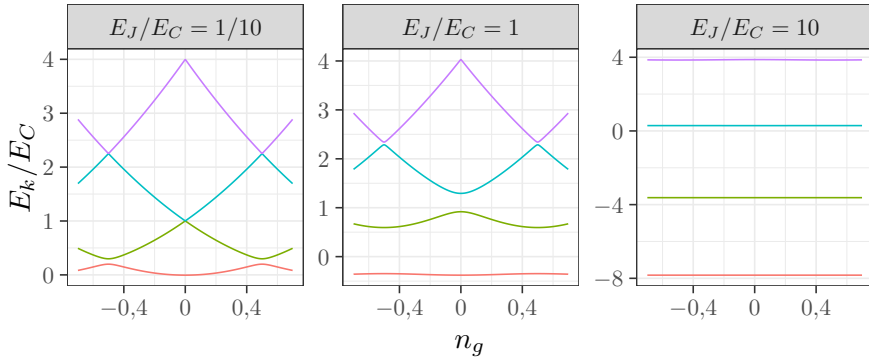


FIGURE 10.12 – Spectres de la boîte à paires de Cooper en fonction de n_g pour $E_J/E_C = 1/10$, 1 et 10.

introduite pour décrire le mouvement d'un pendule oscillant ou encore pour décrire le spectre d'énergie d'une particule libre dans un potentiel périodique sinusoïdal⁹. Les solutions de Mathieu avec conditions aux bords quasi périodiques nous permettent de déduire les énergies propres E du système qui sont organisées en bandes indexées par un nombre entier k [110]. Pour $n_g \in]0, 1/2[$, nous avons :

$$\frac{4E_k}{E_C} = \mathcal{M}_A \left(2 \left\lfloor \frac{k+1}{2} \right\rfloor + 2n_g(-1)^k, -\frac{2E_J}{E_C} \right), \quad (10.70)$$

où $\mathcal{M}_A(r, q)$ est appelée la caractéristique de Mathieu d'exposant r et de paramètre q . La figure 10.12 présente les spectres correspondants en fonction de l'*offset* de charge pour différentes valeurs du ratio E_J/E_C .

Lorsque l'énergie de charge domine ($E_J \ll E_C$), on trouve un spectre de paraboles caractéristiques d'une capacité, sans croisement de niveaux. L'ouverture du spectre autour de ces potentiels points de croisement, ce que l'on appelle un croisement de niveaux évité, est due à la traversée de paires de Cooper à travers l'îlot S . Entre deux croisements évités, les états propres sont des états de charge. L'hybridation entre états à nombre de paires de Cooper distincts n'a lieu qu'au voisinage du chaque croisement évité, c'est-à-dire rarement dans la limite $E_C \gg E_J$. Ceci justifie que ce dispositif soit appelé une boîte à paires de Cooper.

En isolant les deux niveaux proches de l'anticroisement, on dispose d'un système à deux niveaux que l'on peut contrôler en modulant d'une part n_g au moyen du biais en tension V_g et d'autre part en modulant E_J *via* le

9. Cette façon de résoudre une équation différentielle avec des paramètres périodiques repose sur le théorème de Floquet, connu aussi sous le nom de théorème de Bloch en matière condensée.

flux magnétique Φ_B traversant le SQUID. La boîte à paires de Cooper est historiquement le premier qubit supraconducteur développé [62].

Malheureusement, le point faible de ce type de qubit est la grande sensibilité au bruit en tension : si V_g et donc n_g varient, on induit décohérence et relaxation du qubit [25]. Une stratégie pour contourner ce problème consiste à augmenter le rapport E_J/E_C en diminuant l'énergie de charge. Pour cela, on place la jonction Josephson en parallèle avec une grande capacité C de manière à augmenter la capacité totale $C + C_J$ qui intervient dans l'énergie de charge. Lorsque l'on augmente le ratio E_J/E_C , les bandes du spectre deviennent plus plates, ce qui signifie que le dispositif est moins sensible au bruit de charge. C'est le principe du qubit appelé transmon [237, 209]. Toutefois, augmenter E_J/E_C se traduit par une anharmonicité moins élevée et donc moins d'écart entre les différentes transitions. Ceci rend le transmon plus difficile à faire fonctionner comme un simple système à deux niveaux. Néanmoins des techniques de contrôle de la dynamique par des impulsions radiofréquences adaptées permettent d'éviter d'exciter plus de deux niveaux [280].

Qubits de flux

Un autre type de qubit supraconducteur est obtenu en mettant une inductance en parallèle d'un SQUID (figure 10.13). En choisissant bien les rapports E_L/E_J , il est possible de façonner un paysage en énergie pour la phase φ qui présente deux minima, et donc d'obtenir un qubit effectif. Notons qu'ici l'énergie de charge liée à la partie capacitive du SQUID va introduire un terme de couplage entre les états correspondant à une localisation au fond des minima de potentiel. Ces deux états correspondent à des gradients de phase se faisant dans un sens opposé autour de l'anneau et donc, en vertu de la relation entre gradient de phase et courant électrique, à des courants permanents circulant dans des sens opposés autour de la boucle supraconductrice principale [369].

Ces qubits furent à l'origine proposés pour contourner la sensibilité aux fluctuations des charges des boîtes à paires de Cooper [291]. Le principal avantage des qubits de flux réside dans la possibilité de les coupler directement en mettant à profit l'inductance mutuelle entre boucles supraconductrices. Si on considère que les deux états de base canonique du qubit de flux correspondent à des sens de circulation du courant opposés, alors le couplage entre deux qubits de flux est décrit par un opérateur $\sigma_z^1 \sigma_z^2$. La perspective de constituer ainsi un ensemble de qubits aux couplages contrôlés motiva plusieurs études [195, 196, 220] avant d'être utilisée à plus grande échelle dans les appareils vendus par la société canadienne D-Wave.

10.2.4 L'électrodynamique sur circuit

Une percée décisive dans la création d'expériences utilisant ces systèmes a eu lieu en 2004, lorsque R. J. Schoelkopf et ses collaborateurs ont suggéré de réaliser des expériences d'électrodynamique en cavité sur un circuit intégré, en

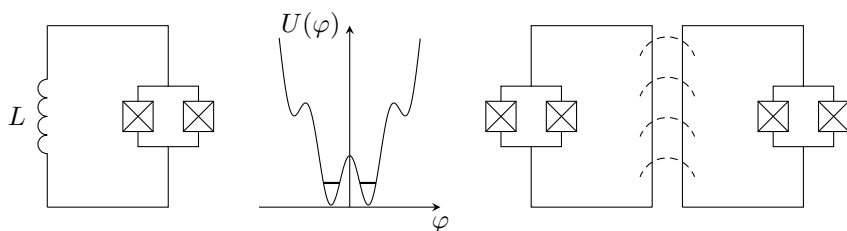


FIGURE 10.13 – À gauche : schéma du circuit correspondant à un qubit de flux, où un SQUID est intégré dans une boucle d'inductance L . Au milieu : représentation du potentiel pour le saut de phase φ à travers le SQUID. Il y a compétition entre le terme parabolique lié à l'inductance et le terme en cosinus des jonctions supraconductrices, ce qui donne deux minima de potentiels. À droite : représentation de deux qubits de flux couplés entre eux *via* l'inductance mutuelle des boucles considérées. Ce type de couplage est au cœur de systèmes plus complexes visant à avoir un grand nombre de qubits couplés.

combinant des atomes artificiels basés sur des nanostructures supraconductrices et des résonateurs également constitués de supraconducteurs [57]. La démonstration expérimentale suivit presque immédiatement [370], ouvrant ainsi le domaine de l'électrodynamique sur circuit. Ces expériences sur puce présentent deux avantages majeurs : d'une part, le temps d'interaction atome/champ n'est pas limité par le temps de transit au travers de la cavité, puisque les atomes sont fixes. D'autre part, l'utilisation de circuits intégrés permet d'espérer la conception de dispositifs plus complexes qu'avec des jets atomiques.

À l'origine, le résonateur était fabriqué au moyen de pistes supraconductrices formant ce qu'on appelle un résonateur en bande. Ce type de système est toujours utilisé, en particulier dans les projets de microprocesseur quantique utilisant les qubits supraconducteurs. Cependant, dans certaines expériences de physique fondamentale récentes, l'atome artificiel supraconducteur (figure 10.14, à gauche) est placé dans une cavité tridimensionnelle de fréquence caractéristique autour du gigahertz (figure 10.14, à droite). Il est couplé au champ de la cavité *via* deux antennes. La cavité est reliée à un câble coaxial, ce qui permet d'injecter ou de récupérer le champ. La mesure se fait grâce à un composant électronique qui permet d'accéder à des informations sur les deux quadratures du champ simultanément (figure 10.15).

Pour décrire ces expériences, la première étape consiste à modéliser l'environnement électromagnétique du qubit supraconducteur. Ce dernier est couplé au champ électromagnétique quantique par l'intermédiaire d'une cavité reliée au monde extérieur par une ligne de transmission. Dans la limite où la largeur spectrale des modes de la cavité est grande devant la largeur spectrale du qubit, on peut alors éliminer la cavité et tout se passe comme si le qubit était directement couplé aux modes de la ligne de transmission. Même si, en

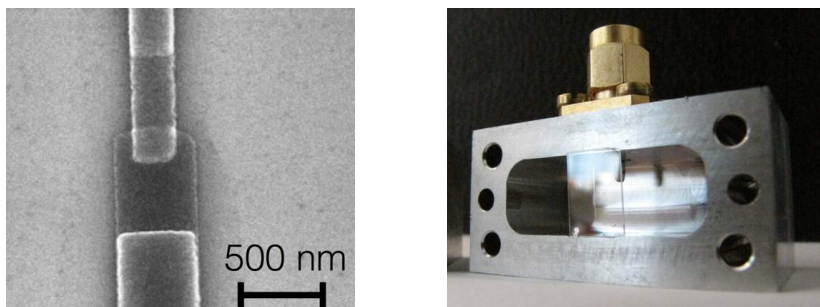


FIGURE 10.14 – Dispositif pour l'électrodynamique sur circuit 3D (reproduit avec l'accord de Benjamin Huard). À gauche : image en microscopie électronique de la nanostructure qui joue le rôle d'atome artificiel. Il s'agit d'une jonction Josephson munie de deux antennes capacitivement couplées. À droite : la cavité, qui abrite l'atome artificiel en son centre : on aperçoit l'atome qui est la nanostructure supraconductrice gravée sur un substrat isolant (plaque de saphir). Le connecteur que l'on voit en haut permet de lire ou de modifier le champ dans la cavité.

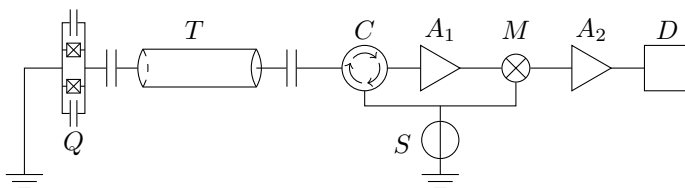


FIGURE 10.15 – Schéma du dispositif expérimental : l'atome artificiel Q est capacitivement couplé à une ligne de transmission T qui permet d'injecter et de capter le champ électromagnétique. Celui-ci, sélectionné par le circulateur C , est amplifié par un amplificateur A_1 à la limite quantique, puis rabaissé en fréquence par un mixeur micro-ondes M , puis de nouveau amplifié par A_2 et enfin analysé par une carte d'acquisition D travaillant au mégahertz.

principe, nous pourrions nous affranchir de cette approximation, nous allons dans la suite développer une modélisation simple basée sur celle-ci.

Ligne de transmission

Dans le régime considéré, l'environnement du qubit est principalement constitué par la ligne de transmission. Celle-ci est modélisée par un câble coaxial sans perte, dont l'état physique est décrit par la tension $V(z, t)$ et le courant $i(z, t)$ en un point z et un instant t donnés. En électrocinétique, un tel câble est décrit par deux ensembles d'oscillateurs harmoniques indépendants de fréquences positives, correspondant aux courants se déplaçant vers la droite et vers la gauche dont on note les opérateurs d'annihilation $a_R(\omega)$ et $a_L(\omega)$.

Néanmoins, cette représentation en fréquence n'est pas bien adaptée pour discuter les observables résolues en temps. Pour cela, on utilise la transformée de Fourier des opérateurs en fréquence :

$$a_{R/L}(t) = \int_0^\infty a_{R/L}(\omega) e^{-i\omega t} \frac{d\omega}{\sqrt{2\pi}}. \quad (10.71)$$

Les relations de commutation des $a_{R/L}(t)$ sont non triviales mais, si on suppose que le système émet dans une bande spectrale étroite autour d'une fréquence porteuse bien séparée des basses fréquences, on peut étendre formellement l'intégrale de l'équation (10.71) aux fréquences négatives. En faisant cela, les relations de commutation des opérateurs $a_{R/L}(t)$ sont modifiées. Ainsi, les modes L et les modes R commutent toujours entre eux, mais surtout, étendre l'ensemble des oscillateurs entraîne que les opérateurs pris à des temps différents commutent aussi entre eux. La relation de commutation canonique correspondante est alors donnée par :

$$\left[a_\alpha(t), a_{\alpha'}^\dagger(t') \right] = \delta_{\alpha,\alpha'} \delta(t - t'). \quad (10.72)$$

En conséquence, on peut voir cette ligne comme une collection d'oscillateurs harmoniques, un à chaque pas de temps. Bien entendu, d'un point de vue dynamique, même en l'absence de couplage au qubit (ligne isolée), les équations de Heisenberg décrivant l'évolution des opérateurs nous disent que ces oscillateurs sont couplés entre eux, ce qui traduit la propagation de l'information le long de la ligne.

Qubit et ligne de transmission

Considérons maintenant le couplage entre le qubit et la ligne de transmission dans le régime où l'on peut éliminer la cavité de manière adiabatique, c'est-à-dire dans le régime où le taux de fuite de la cavité est grand devant le couplage entre la cavité et le qubit. Dans ce régime, la cavité est presque transparente et l'interaction entre le qubit et la ligne de transmission est décrite par un hamiltonien effectif local.

En notant σ_- (resp. σ_+) l'opérateur de descente (resp. de montée) du qubit, $a(\omega) = a_R(\omega) + a_L(\omega)$ et en faisant l'approximation séculaire (section 10.1.1), le hamiltonien effectif décrivant ce système [380] est de nouveau le hamiltonien de Jaynes-Cummings :

$$H = H_S + H_{\mathcal{L}} + H_{S\mathcal{L}}, \quad (10.73a)$$

$$H_S = \hbar\omega_{eg} \sigma_+ \sigma_-, \quad (10.73b)$$

$$H_{\mathcal{L}} = \hbar \int_0^{+\infty} \omega a^\dagger(\omega) a(\omega) d\omega, \quad (10.73c)$$

$$H_{S\mathcal{L}} = \hbar \int_0^{+\infty} (g(\omega) a(\omega) \sigma_+ + g^*(\omega) a^\dagger(\omega) \sigma_-) \frac{d\omega}{\sqrt{2\pi}}. \quad (10.73d)$$

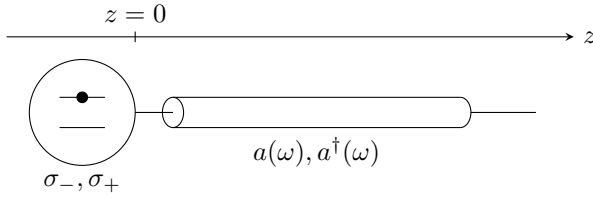


FIGURE 10.16 – L’environnement du qubit est constitué par une ligne de transmission couplée à son extrémité au qubit. Le couplage prend alors la forme de Jaynes-Cummings (voir équation (10.73d)).

où H_S et $\mathcal{H}_L$ désignent respectivement les hamiltoniens du qubit et de la ligne isolés. Comme le qubit émet dans une bande de fréquence proche de sa fréquence de résonance, il est raisonnable de supposer que $g(\omega)$ est constant, ce qui correspond à un couplage local en temps, sans effet mémoire. On pose alors $g(\omega) = i\sqrt{\gamma}$ où γ est le taux de relaxation du qubit, le facteur i étant purement conventionnel puisqu’on dispose de la symétrie de phase $\sigma_- \mapsto e^{i\vartheta}\sigma_-$. De plus, nous pouvons formellement étendre l’intégrale de $\mathbb{R}^+$ à $\mathbb{R}$. On se place alors en représentation d’interaction de manière à éliminer l’évolution libre de la ligne et du qubit. Le hamiltonien d’interaction prend alors une forme très simple en fonction des opérateurs temporels de la ligne et de l’opérateur $\sigma_-(t) = e^{i\omega_{eg}t}\sigma_-$:

$$H_I(t) = i\hbar\sqrt{\gamma} (a^\dagger(t)\sigma_-(t) - a(t)\sigma_+(t)) . \quad (10.74)$$

Examinons maintenant ce qui se passe lors de l’interaction qubit-ligne. À chaque instant, un oscillateur harmonique local en temps interagit avec le qubit (figure 9.3, page 376). À la suite de cette interaction, on a la formation d’un état intriqué entre cet oscillateur harmonique et le qubit. Le qubit a donc imprimé une information dans son environnement électromagnétique de manière locale en temps. En mesurant chaque oscillateur, on peut ainsi accéder à la dynamique du qubit conditionnée au signal expérimental enregistré.

10.2.5 Les trajectoires quantiques du qubit

Alors que, dans les expériences d’électrodynamique en cavité, nous avons extrait de l’information sur l’état du mode du champ en utilisant les atomes, ici nous allons utiliser le champ électromagnétique pour sonder l’état du qubit. Mais, en analysant les propriétés du rayonnement émis dans la ligne de transmission, nous allons agir sur le qubit. Ceci donne naissance aux trajectoires quantiques conditionnées aux signaux expérimentaux mesurés.

Suivant la mesure effectuée, on obtiendra une information et donc des trajectoires quantiques différentes. Par exemple, on peut commencer par considérer le rayonnement émis par le qubit par émission spontanée, ce qu’on

appelle la fluorescence en physique atomique. Celle-ci peut être observée en mesurant le nombre de photons de manière résolue en temps, ce qui nous donnera un premier exemple de trajectoires quantiques. Cependant, ce rayonnement étant émis dans le domaine micro-onde, une telle détection résolue en temps n'est pas réaliste.

En effet, dans le domaine optique, les détecteurs fonctionnent sur le principe du photomultiplicateur : un photon est absorbé, déclenchant l'émission d'un électron et un phénomène d'amplification (souvent une ionisation) permet ensuite d'avoir un signal macroscopique. Les photodétecteurs modernes sont des photo-diodes à avalanche dont le temps de réponse peut être de l'ordre de quelques dizaines de picosecondes. Mais il faut se rappeler que ce temps est long comparé à la période d'une onde dans le domaine optique qui est de l'ordre de la femtoseconde. Même si le couplage électromagnétique est faible, ce temps est assez long pour avoir une efficacité de détection de photons uniques proche de 100 %. On sait donc réaliser des détecteurs efficaces et résolus en temps. Mais dans le domaine micro-ondes, on cherche à détecter des photons dont la fréquence caractéristique est dans le domaine des gigahertz ou plus avec une résolution en temps de l'ordre de la nanoseconde ou moins. Le défi est tout autre. D'une part, il est nécessaire de fabriquer des systèmes dont les bandes d'absorption sont dans cette gamme, qui est bien plus basse que le *gap* naturel de nombreux semi-conducteurs. D'autre part, il est bien plus difficile d'assurer une efficacité voisine de l'unité sur des échelles de temps aussi courtes¹⁰. Pour cette raison, et bien que des recherches continuent activement pour mettre au point des détecteurs de photons micro-ondes uniques efficaces et résolus en temps [24, 215, 93, 50], la quasi-totalité des expériences conduites avec des circuits supraconducteurs repose encore sur des techniques de mesure électroniques, c'est-à-dire qu'elles reposent *in fine* sur des mesures de tension effectuées sur la ligne de transmission. Nous verrons comment elles permettent de reconstruire la projection du vecteur de Bloch associé à l'état du système sur son plan équatorial.

Comptage de photons

Afin de décrire la dynamique pour une mesure du nombre de photons, il faut développer l'opérateur d'évolution en représentation d'interaction :

$$U_I(t, t + dt) = \exp \left(-\frac{i}{\hbar} H_I(t) dt \right) = \exp \left[(a^\dagger(t) \sigma_- - a(t) \sigma_+) \sqrt{\gamma} dt \right]. \quad (10.75)$$

Ce développement contient une subtilité : en effet, il faut le pousser à l'ordre deux, car à cet ordre apparaissent des termes $a(t)a^\dagger(t)$ et $a^\dagger(t)a(t)$ que l'on va

10. Les dispositifs les plus performants sont des détecteurs à nano-fils supraconducteurs qui arrivent à des efficacités de détection de 60 % à 300 THz avec une résolution temporelle de 50 ps. Malheureusement, ces dispositifs sont pour le moment cantonnés à l'infrarouge lointain à cause de la température de transition supraconductrice du matériau et des contraintes de fabrication sur les nano-fils.

faire agir sur l'état $|0_t\rangle$. On doit donc utiliser la relation de commutation (10.72) à temps égaux. Il apparaît alors une singularité en $\delta(0)$ que l'on résout en utilisant l'astuce suivante : $[a(t), a^\dagger(t)] = \delta(0) = 1/dt$. Cela nous permet de développer l'opérateur d'évolution au premier ordre et donne l'évolution suivante, où $|\psi(t)\rangle$ est l'état du qubit :

$$|\psi(t)\rangle \otimes |0_t\rangle \rightarrow \left(\mathbb{1} - \frac{\gamma dt}{2} \sigma_+ \sigma_-(t) \right) |\psi(t)\rangle \otimes |0_t\rangle \quad (10.76a)$$

$$+ \sqrt{\gamma dt} \sigma_-(t) |\psi(t)\rangle \otimes \sqrt{dt} a^\dagger(t) |0_t\rangle . \quad (10.76b)$$

Selon que l'on détecte ou non un photon, on obtient deux états différents pour le système. Si on ne détecte pas de photon à l'instant t , ce qui correspond à l'état $|0_t\rangle$ en sortie de la ligne, on a l'état joint

$$|\psi(t+dt, n(t)=0)\rangle = \left(\mathbb{1} - \frac{\gamma dt}{2} \sigma_+ \sigma_-(t) \right) |\psi(t)\rangle . \quad (10.77)$$

Ceci implique que l'amplitude sur l'état $|e\rangle$ diminue. De même, si on détecte un photon à l'instant t , cela correspond cette fois à l'état $|1_t\rangle = \sqrt{dt} a^\dagger(t) |0_t\rangle$ côté environnement, et on obtient l'état joint :

$$|\psi(t+dt, n(t)=1)\rangle = \sqrt{\gamma dt} \sigma_-(t) |\psi(t)\rangle . \quad (10.78)$$

Il faut noter que l'état $|\psi(t+dt, n(t))\rangle$ n'est pas normé, car sa norme au carré correspond à la probabilité d'obtenir le résultat $n(t)$. Il s'agit ici directement de l'état joint introduit à la section 9.3.2. Nous obtenons finalement une équation différentielle stochastique pour l'état joint à l'historique de mesure $[n] : t \mapsto n(t)$:

$$|\psi(t+dt, [n])\rangle = \left(\mathbb{1} - \frac{\gamma dt}{2} \sigma_+ \sigma_-(t) + n(t) \left(\sqrt{\gamma dt} \sigma_-(t) - \mathbb{1} \right) \right) |\psi(t, [n])\rangle , \quad (10.79)$$

où $n(t) \in \{0, 1\}$ est une variable aléatoire stochastique. Sa distribution de probabilité dépend de l'état du qubit à l'instant t et est donnée par :

$$P(n(t)=0 | \psi(t)) = 1 - \gamma dt \langle \sigma_+(t) \sigma_-(t) \rangle_{|\psi(t)\rangle} , \quad (10.80a)$$

$$P(n(t)=1 | \psi(t)) = \gamma dt \langle \sigma_+(t) \sigma_-(t) \rangle_{|\psi(t)\rangle} . \quad (10.80b)$$

On peut alors s'intéresser à l'émission spontanée du qubit : le qubit initialement dans l'état excité se désexcite à l'instant t en émettant un photon. C'est ce que l'on voit sur les simulations de la figure 10.17.

Mesure de la tension

La détection de photons est couramment utilisée dans le domaine optique car on dispose de photodiodes à avalanche capables de détecter l'arrivée

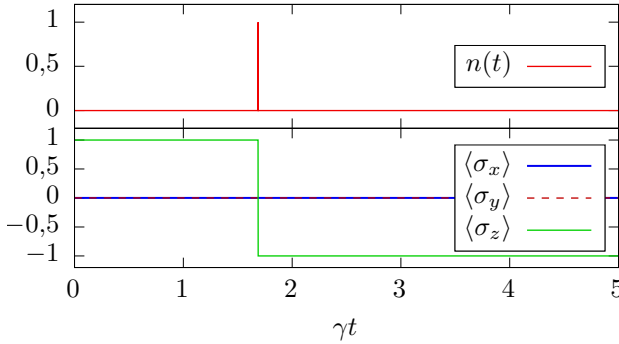


FIGURE 10.17 – Trajectoire associée au comptage des photons en représentant l'état du qubit sur la sphère de Bloch et *via* les trois observables σ_x , σ_y et σ_z . Avec cette méthode de mesure, on ne peut que détecter l'émission d'un photon par le qubit à un instant donné. À cet instant, l'état du qubit est donc modifié par un saut vers l'état fondamental, ce qui correspond à une valeur de σ_z qui passe de 1 à -1 .

d'un photon unique avec une très grande efficacité et un temps de réponse court devant l'écart entre deux arrivées de photon. Malheureusement, dans les expériences conduites avec des qubits en cavité, on n'a pas accès à de tels détecteurs et on se repose donc sur les outils de l'électronique micro-ondes. En pratique, on mesure donc la tension électrique en extrémité de ligne à chaque instant. La quantité mesurée est alors proportionnelle à la quadrature x de l'oscillateur harmonique au temps t , et non plus au nombre de photons. Ceci est encore une idéalisation de ce qui est réalisé dans les expériences mais elle va nous permettre de discuter d'un type de trajectoires radicalement différentes de celles considérées ci-dessus.

Pour trouver l'évolution conditionnée à une mesure de la tension, il suffit d'exprimer les états de la ligne de transmission dans la base orthogonale associée à cette mesure. On note $|x_t\rangle$ les états propres de l'opérateur $\hat{x}_t = (a + a^\dagger)\sqrt{\hbar}/\sqrt{2}$ associés à la valeur propre $x(t)$. Si on veut se servir de ce que l'on a déjà calculé, il faut alors décomposer l'état vide et l'état à un photon sur cette base, ce qui donne :

$$|0_t\rangle = \int \langle x_t|0_t\rangle |x_t\rangle dx_t \quad \sqrt{\hbar}a^\dagger(t)|0_t\rangle = \sqrt{2} \int \langle x_t|0_t\rangle |x_t\rangle x(t) dx_t. \quad (10.81)$$

Le produit scalaire $\langle x_t|0_t\rangle$ n'est rien d'autre que la fonction d'onde de l'état vide de l'oscillateur harmonique à l'instant t . Il s'agit donc d'une gaussienne centrée en zéro dont la largeur est déterminée par les relations de commutation section 4.2.2. Elle décrit les fluctuations de la tension dans l'état vide de la ligne de transmission. Si l'on décompose l'état de la ligne de transmission sur

cette base, on obtient alors l'équation de Schrödinger stochastique (10.76) :

$$\begin{aligned} |\psi(t + dt, [x])\rangle &= \frac{1}{\pi^{\frac{1}{4}}} e^{-x(t)^2/2} \\ &\times \left(\mathbb{1} - \frac{\gamma dt}{2} \sigma_+ \sigma_-(t) + \sqrt{2\gamma dt} \sigma_-(t) x(t) \right) |\psi(t, [x])\rangle . \end{aligned} \quad (10.82)$$

On déduit de cette équation la probabilité d'obtenir un certain $x(t)$ à l'instant t , conditionnée à l'état $|\psi(t)\rangle$:

$$P(x(t)|\psi(t)) = \frac{1}{\sqrt{\pi}} \exp \left(- \left(x - \sqrt{\frac{\gamma dt}{2}} \langle \sigma_- + \sigma_+ \rangle \right)^2 \right) . \quad (10.83)$$

Cette probabilité est donc une gaussienne décentrée de façon proportionnelle à $\langle \sigma_x \rangle$. Les fluctuations du signal mesuré reflètent les fluctuations de l'état vide de photons dans la ligne de transmission. Ici, l'information que l'on acquiert sur le qubit est donc la valeur moyenne $\langle \sigma_x(t) \rangle$ le long d'une trajectoire.

Pour cette observable, les trajectoires suivies par le qubit lors d'une émission spontanée sont plus complexes que dans le cas du comptage de photon. En effet, l'équation de Schrödinger stochastique (10.82) décrit un mouvement diffusif avec apparition de cohérences le long de la trajectoire ($\langle \sigma_{x,y}(t) \rangle \neq 0$). Cela provient du fait que le signal qui apparaît dans l'équation (10.83) possède des fluctuations gaussiennes décorréliées en temps : $\langle x(t) x(t') \rangle = \delta(t - t')$. Contrairement au cas du comptage de photons, la trajectoire quantique de l'état du qubit est continue. Cela reflète l'acquisition plus progressive d'information *via* la mesure de tension qu'au travers du photocomptage.

Mesure des deux quadratures

Ce schéma de mesure n'est cependant pas encore réaliste. En effet, il est difficile de mesurer une tension avec une résolution temporelle sub-nanoseconde. Or, ce qui nous intéresse ici, ce sont les fluctuations encodées dans les modes à des fréquences voisines de la fréquence de transition du qubit. Pour y accéder, on se ramène à des mesures à plus basse fréquence en mélangeant le signal expérimental avec une porteuse à la fréquence voulue. Dans le domaine des radiofréquences, cela se fait traditionnellement à l'aide d'un mixeur qui multiplie le signal expérimental par un signal sonde contrôlé.

Ainsi, considérons un signal classique $X(t)$ lentement modulé autour d'une porteuse à la pulsation ω_0 : $X(t) = x_R(t) \cos \omega_0 t + x_I(t) \sin \omega_0 t$ où les deux quadratures $x_R(t)$ et $x_I(t)$ sont des signaux à basse fréquence. En multipliant ce signal par une sonde à la pulsation $\omega = \omega_0 + \Omega$, nous obtenons un signal

$$X(t) \cos \omega t = \frac{1}{2} (x_R(t) \cos(\Omega + 2\omega_0)t + x_I(t) \sin(\Omega + 2\omega_0)t) \quad (10.84a)$$

$$+ \frac{1}{2} (x_R(t) \cos \Omega t + x_I(t) \sin \Omega t) \quad (10.84b)$$

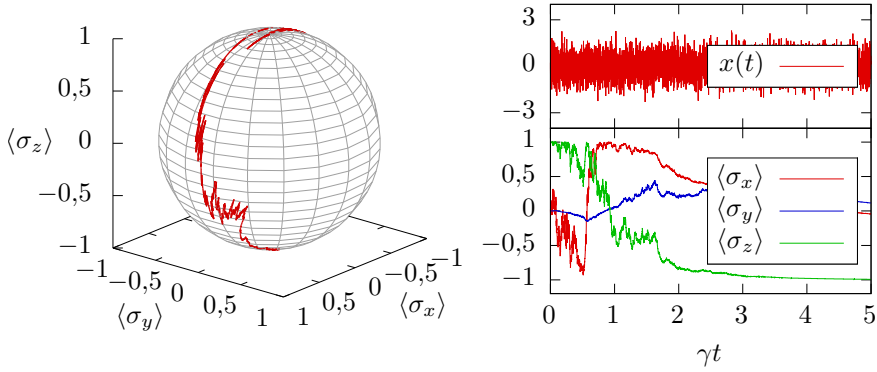


FIGURE 10.18 – Exemple de trajectoire $t \mapsto x(t)$ suivie lors d’une mesure de la tension représentée sur la sphère de Bloch ainsi que la position du qubit en coordonnées $\langle \sigma_{x,y,z}(t) \rangle$. On a cette fois une évolution continue entre l’état initial et l’état final.

qui possède une contribution (10.84a) haute fréquence autour de $2\omega_0$ et une contribution (10.84b) à basse fréquence contenant toute l’information sur l’évolution temporelle de $x_R(t)$ et $x_I(t)$. Numériser le signal avec un pas de temps long devant π/ω_0 permet de récupérer l’information sur les deux quadratures. Le composant qui réalise cette opération est appelé un démodulateur IQ ou *IQ mixer*. Cette technique, appelée *hétérodynage*, permet de récupérer potentiellement plus d’information que la détection directe d’une seule des deux quadratures.

En pratique, le démodulateur se situe dans la partie non cryogénique du dispositif expérimental. Cependant, les signaux émis par un qubit sont tellement faibles que l’on ne détecterait rien, même après amplification en sortie de ce composant. Le dispositif expérimental comporte donc une chaîne de pré-amplification placée en sortie du dispositif ou de la cavité 3D où sont placés les qubits. La démodulation est donc effectuée sur un signal au préalable amplifié comme indiqué sur la figure 10.15. Ce type de montage n’a été rendu possible qu’avec le développement d’amplificateurs supraconducteurs à la limite quantique [85, 339, 47, 46, 134].

Le modèle le plus simple d’amplificateur quantique possède deux ports d’entrée a_0, d_0 et deux ports de sortie a_1 et d_1 qui sont reliés par une transformation linéaire¹¹

$$\begin{pmatrix} a_1 \\ d_1^\dagger \end{pmatrix} = \begin{pmatrix} \sqrt{G} & \sqrt{G-1} \\ \sqrt{G-1} & \sqrt{G} \end{pmatrix} \begin{pmatrix} a_0 \\ d_0^\dagger \end{pmatrix} \quad (10.85)$$

11. C’est une transformation de Bogoliubov de paramètre r , tel que $\cosh r = \sqrt{G}$ où G est le gain de l’amplificateur.

qui, en sortie a_1 , amplifie le signal a_0 avec le gain¹² $G > 1$. La présence de la seconde voie d'entrée et de la seconde voie de sortie est nécessaire afin de préserver les relations de commutation canonique (10.72) sur la voie de sortie a_1 . Cet amplificateur préserve la phase du signal d'entrée : si on envoie en a_0 un état cohérent d'amplitude α et le vide de photons en d_0 , l'amplitude moyenne en sortie $\langle a_1 \rangle$ possède la même phase que α . Cependant, la relation (10.85) a deux autres conséquences physiques importantes.

Tout d'abord, le signal amplifié comporte un bruit ajouté qui est dû à la présence du terme en $d_0^\dagger$ dans l'expression de a_1 en fonction des signaux d'entrée. En effet, la moyenne

$$S^{(\text{in})} = \frac{1}{2} (\langle \Delta X^2 \rangle + \langle \Delta P^2 \rangle) = \langle a_0^\dagger a_0 \rangle - |\langle a_0 \rangle|^2 + \frac{1}{2} \quad (10.86)$$

des fluctuations des deux quadratures X et P associées à un mode du signal d'entrée contient une contribution $\langle a_0^\dagger a_0 \rangle - |\langle a_0 \rangle|^2 \geq 0$ liée au signal lui-même et une contribution des fluctuations de point zéro qui découle des relations de commutation canoniques entre les quadratures $[X, P] = -i$. En sortie de l'amplificateur, lorsque l'on envoie le vide de photons sur le mode d_0 , cette même quantité est donnée par

$$S^{(\text{out})} = \frac{G}{2} + G (\langle a_0^\dagger a_0 \rangle - |\langle a_0 \rangle|^2). \quad (10.87)$$

On y retrouve bien la contribution du bruit du signal amplifiée par le gain G mais la contribution quantique est $G/2$ au lieu de $1/2$. On a donc ajouté un bruit $(G - 1)/2$ aux fluctuations du vide attendues normalement. Tout se passe donc comme si on avait rajouté un bruit $(G - 1)/2G$ au signal en entrée, c'est-à-dire $1/2$ dans la limite $G \gg 1$.

L'autre point important est que ces voies supplémentaires entraînent une décohérence entre la voie d'entrée a_0 et la voie de sortie a_1 . En effet, l'équation (10.85) relie les opérateurs en entrée de l'amplificateur à ceux en sortie. En effet, nous pouvons montrer que l'opérateur unitaire

$$U_{\text{amp}} = \exp \left(\sqrt{\frac{G-1}{G}} (a_0^\dagger d_1^\dagger - a_0 d_1) \right) \quad (10.88)$$

décrit comment l'état des photons en amont de l'amplificateur est transformé en aval. Il crée manifestement de l'intrication entre les deux voies de sortie de l'amplificateur. Par exemple, l'état vide en entrée $|0, 0\rangle = |0\rangle_{a_0} \otimes |0\rangle_{d_1}$ est transformé en l'état intriqué¹³

$$U_{\text{amp}} |0, 0\rangle = \frac{1}{\sqrt{G}} \sum_{n=0}^{+\infty} \left(\frac{G-1}{G} \right)^{n/2} |n, n\rangle \quad (10.89)$$

12. Nous utilisons ici le gain en intensité et non en amplitude.

13. On note $|n, m\rangle = |n\rangle_{a_1} \otimes |m\rangle_{d_1}$.

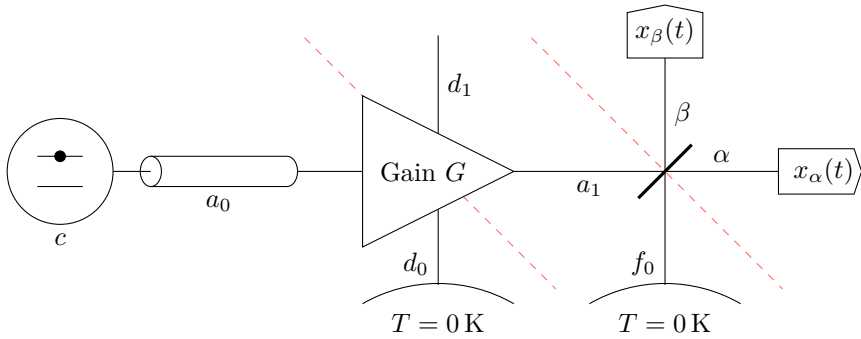


FIGURE 10.19 – Le signal de retour du qubit a_0 est envoyé dans un amplificateur quantique de gain $G > 1$ dont une des voies de sorties est envoyée sur un coupleur où il se mélange aux fluctuations du vide. On détecte une quadrature sur chaque voie de sortie afin de récupérer de l'information sur les deux quadratures de $a_1(t)$ en entrée.

dont la trace partielle sur le mode d_1 est un état thermique avec, pour un mode à la fréquence ω donné, une température effective $k_B T_\omega(G) = \hbar\omega / \log(G/(G-1))$. Elle est nulle pour $G = 1$ et tend vers $+\infty$ quand $G \rightarrow +\infty$.

La figure 10.19 présente une modélisation simplifiée de la chaîne de mesure : le signal en provenance du qubit est envoyé dans un amplificateur quantique dont une sortie est envoyée sur un coupleur. Les signaux $x_\alpha(t)$ et $x_\beta(t)$ sont alors numérisés et le signal mesuré est la combinaison linéaire

$$S(t) = x_\alpha(t) - ix_\beta(t) = \sqrt{G}dt \left(a_0(t) + \sqrt{\frac{G-1}{G}} d_0^\dagger(t) - i \frac{1}{\sqrt{G}} f_0^\dagger(t) \right) \quad (10.90)$$

qui permet une mesure de $\sqrt{G}a(t)$, auquel s'ajoutent d'une part les fluctuations issues de l'amplification et d'autre part les fluctuations de la seconde entrée de la séparatrice. C'est la présence de cette seconde voie d'entrée qui évite de violer les inégalités de Heisenberg sur les deux quadratures de $S(t)$.

Par un raisonnement analogue à celui présenté précédemment, on peut alors établir une équation de Schrödinger stochastique conditionnée cette fois à la mesure de l'observable $S(t)$:

$$\begin{aligned} |\psi(t+dt, [S])\rangle &= \frac{e^{-|S|^2/2G}}{\sqrt{\pi G}} \\ &\times \left(\mathbb{1} - \frac{\gamma dt}{2} \sigma_+ \sigma_-(t) + \sqrt{\gamma dt} \frac{S^*(t)}{\sqrt{G}} \sigma_-(t) \right) |\psi(t, [S])\rangle, \quad (10.91) \end{aligned}$$

où $S(t)$ joue le rôle de variable stochastique dont la densité de probabilité associée dépend de l'état $|\psi(t)\rangle$ du système :

$$P\left(S(t) \middle| \psi(t)\right) = \frac{1}{\pi G} \exp\left(-\left|\frac{S(t)}{\sqrt{G}} - \sqrt{\gamma} dt \langle \sigma_-(t) \rangle\right|^2\right). \quad (10.92)$$

La moyenne de $S(t)$ est donc proportionnelle à $\sqrt{G} \langle \sigma_-(t) \rangle$. En d'autres termes, la détection du rayonnement de fluorescence du qubit permet de reconstruire la projection sur le plan équatorial du vecteur de Bloch correspondant à l'état du qubit. À l'échelle de la trajectoire unique, cette détermination est entachée d'un bruit qui trouve son origine dans les fluctuations quantiques dans les voies d'entrées de la lame séparatrice et lors du processus d'amplification.

Mesure faible de l'état du qubit

Nous venons d'analyser le rayonnement de fluorescence émis par le qubit autour de sa fréquence de résonance. Mais il est possible de récupérer une information différente sur l'état du qubit en considérant une bande de fréquence hors résonance [168]. Dans ce régime dit dispersif, l'interaction entre le qubit et la ligne de transmission est décrite par le hamiltonien effectif (équation (10.5)) :

$$H_I = \hbar \chi \sigma_z \int_0^{+\infty} b_\omega^\dagger b_\omega d\omega, \quad (10.93)$$

où b_ω et $b_\omega^\dagger$ sont les opérateurs de création et destruction de photons dans la cavité. Suivant l'état du qubit, le mode de la cavité voit donc sa fréquence décalée de $\pm \chi$ indépendamment de ω . Le couplage entre la cavité et la ligne de transmission est décrit par un taux de relaxation κ . Le signal expérimental est collecté *via* la ligne de transmission et nous devons éliminer le mode de la cavité pour retomber sur une description n'impliquant plus que le qubit et la ligne. Pour cela, nous devons relier le nombre de photons $b_\omega^\dagger b_\omega$ aux modes de la ligne de transmission.

Supposons que le mode de la cavité contienne en moyenne $\bar{n}_c$ photons. Lorsqu'un état cohérent de paramètre α_t pour l'oscillateur associé aux opérateurs a_t et $a_t^\dagger$ est envoyé sur la cavité, $|\alpha_t|^2 dt$ représente le nombre moyen de photons qui arrivent *via* la ligne de transmission dans le mode associé à l'intervalle temporel $[t, t + dt]$. La bande spectrale de ce mode de la ligne est $1/dt$ de sorte que la fenêtre spectrale de la cavité ne représente qu'une fraction κdt de celle-ci. C'est seulement dans cette bande spectrale que le champ dans la ligne de transmission pénètre dans la cavité. L'amplitude du champ cohérent dans la cavité n'est donc pas la même que l'amplitude du champ incident ou en sortie de la cavité. Plus précisément, on s'attend à ce qu'autour de la fréquence de résonance, le ratio des énergies stockées dans la cavité et dans le champ incident reflète le rapport des largeurs des bandes spectrales :

$$|\alpha_t|^2 dt = \frac{\bar{n}_c}{\kappa dt}. \quad (10.94)$$

Cet argument qualitatif nous permet ainsi de relier l'amplitude $|\alpha_t|$ du champ incident au nombre de photons moyen dans la cavité $\bar{n}_c$.

À partir de l'équation (10.93) et de la relation (10.94), nous pouvons calculer l'évolution de l'état qubit-ligne lorsque l'on envoie un rayonnement cohérent dans la ligne de transmission. Supposons pour simplifier qu'à l'instant t le qubit et les degrés de liberté du champ associés à l'intervalle $[t, t + dt]$ soient dans l'état ¹⁴

$$|\Psi_t\rangle = (a_+ |+\rangle + a_- |-\rangle) \otimes |\alpha_t\rangle \quad (10.95)$$

et que, dans la représentation d'interaction, seul le hamiltonien (10.93) soit présent de sorte que le qubit, s'il est préparé initialement dans les états propres $|\pm\rangle$ de σ_z , y reste. Cet état factorisé devient alors un état intriqué

$$|\Psi'_t\rangle = a_+ |+\rangle |e^{-i\chi dt}\alpha_t\rangle + a_- |-\rangle |e^{i\chi dt}\alpha_t\rangle. \quad (10.96)$$

L'information sur l'état du qubit est donc imprimée dans la phase de l'onde cohérente. Tout l'enjeu consiste donc à trouver un moyen de voir les états cohérents $|e^{\mp i\chi dt}\alpha\rangle$ se séparer afin d'être capable de discerner entre les deux états $|\pm\rangle$ du qubit. Remarquons que ce point de vue est dual de celui utilisé dans la section 10.1.2 où des systèmes à deux niveaux traversant une cavité étaient utilisés pour sonder l'état quantique d'un mode du champ : ici c'est le rayonnement qui permet de sonder l'état d'un qubit confiné dans une cavité.

Pour cela, le choix de la quadrature mesurée ainsi que le processus d'amplification préalable est important mais, pour simplifier la discussion, nous allons considérer ici l'analogie de la situation décrite à la section 10.2.5, c'est-à-dire une mesure directe résolue en temps des quadratures du signal expérimental. Comme les composantes cohérentes se séparent *via* la phase, si nous mesurons la quadrature x_t , alors l'optimum pour discriminer entre les deux composantes est obtenu pour α_t imaginaire pur. Nous pouvons décomposer l'état $|\Psi'_t\rangle$ en utilisant la base $|x\rangle$, ce qui nous donne

$$|\Psi'_t\rangle = \int_{\mathbb{R}} |\psi_{\text{qb}}(t + dt, x)\rangle |x\rangle dx, \quad (10.97a)$$

$$|\psi_{\text{qb}}(t + dt, x)\rangle = \sum_{s=\pm} \langle x | e^{-is\chi dt} \alpha_t \rangle a_s |s\rangle. \quad (10.97b)$$

Comme $|\alpha_t|$ est donné par l'équation (10.94), la distribution de probabilité $|\langle x | e^{\mp i\chi dt} \alpha_t \rangle|^2$ est une gaussienne dont la moyenne est $\pm \sqrt{2\chi^2 \bar{n}_c dt / \kappa}$ et la variance est l'unité (équation (10.83)) et donc, elles se recouvrent fortement. Nous sommes exactement dans une situation de mesure faible (section 5.4.2).

L'équation (5.78) quantifie l'information acquise entre t et $t + dt$ lorsque l'on part d'une situation où notre ignorance est maximale ($|a_+|^2 = |a_-|^2 = 1/2$).

14. Nous ignorons les effets des degrés de liberté du champ bosonique à $t' < t$, ce qui revient à supposer que nous avons déjà effectué une mesure de ces derniers et que l'état $a_+ |+\rangle + a_- |-\rangle$ est l'état du qubit conditionné à ces résultats lus.

Dans le cas présent, cette information

$$I[E; R] = 4\bar{n}_c \frac{\chi^2}{\kappa} dt \quad (10.98)$$

est proportionnelle à dt , ce qui veut dire que la mesure donne un taux d'acquisition d'information $\Gamma_m = 4\bar{n}_c \chi^2 / \kappa$ qui est proportionnel au nombre moyen de photons dans la cavité. Le taux de fuite de la cavité étant κ , on peut donc interpréter χ^2 / κ^2 comme l'information véhiculée par photon. Nous travaillons dans le cadre d'une approximation markovienne qui suppose que le temps de relaxation de la cavité κ^{-1} est court devant le temps d'évolution du qubit¹⁵. Le qubit n'étant pas soumis à un forçage résonant qui introduirait une nouvelle échelle de temps, son échelle de temps caractéristique est le temps de décohérence dans la base $|\pm\rangle_z$ qui, comme montré dans la section 5.4.4, est inférieur au temps de mesure Γ_m^{-1} . La validité de l'approximation markovienne est donc assurée lorsque $1 \ll \kappa^2 / 4\bar{n}_c \chi^2$ soit $(\chi/\kappa)^2 \ll 1/\bar{n}_c$. Lorsque la cavité est peuplée par un champ quasi classique ($\bar{n}_c \gg 1$), chaque photon qui s'échappe emporte donc une quantité d'information faible. Plusieurs photons sont nécessaires pour séparer les deux états $|\pm\rangle_z$ du qubit mais, le taux d'émission du premier photon étant amplifié par la statistique bosonique, le temps de mesure est diminué. En revanche, dans la limite d'un champ quantique $\bar{n}_c \lesssim 1$ dans la cavité, on peut travailler dans un régime où $\chi/\kappa \sim 1$. Dans ce régime, une seule émission suffit car chaque photon emporte une information suffisante. Le temps de mesure est alors le temps κ^{-1} d'échappement d'un seul photon. Remarquons au passage que lorsque $\chi \gtrsim \kappa$, qui n'est pas le régime de mesure faible du qubit considéré ici, la largeur naturelle de la cavité est plus petite que le décalage spectral χ du qubit induit par l'ajout d'un photon dans la cavité. Dans ce régime, dit dispersif fort, on peut utiliser le qubit comme spectromètre pour résoudre le nombre de photons dans la cavité [169, 327].

Considérons maintenant le signal expérimental : la norme au carré de l'état joint $|\psi_{qb}(t + dt, x)\rangle$ donne la probabilité d'obtenir le résultat de mesure x entre t et $t + dt$. Au premier ordre en dt , elle est donnée par :

$$P(x|\psi_{qb}(t)) = \frac{1}{\sqrt{\pi}} e^{-(x - \sqrt{2\bar{n}_c} \chi^2 dt / \kappa \langle \sigma_z \rangle_{|\psi_{qb}(t)\rangle})^2}. \quad (10.99)$$

Au vu de cette équation, la quantité $X(t) = \sqrt{\kappa dt} \sum_{\tau \in [0, t]} x_\tau$ obéit à l'équation stochastique

$$dX(t) = \chi \sqrt{2\bar{n}_c} \langle \sigma_z(t) \rangle dt + \xi(t) dt, \quad (10.100)$$

où $\xi(t)$ est un bruit aléatoire gaussien de moyenne nulle. Les fluctuations statistiques sont données par $\mathbb{E}(\xi(t)\xi(t')) = \kappa \delta(t - t')$. Le signal expérimental $X(t)$ permet donc de remonter à $\langle \sigma_z(t) \rangle$ en considérant sa valeur moyenne

15. Rappelons que les fréquences propres du qubit et de la cavité sont désaccordées (régime dispersif) d'une quantité χ qui est supposée petite devant les fréquences propres de ces deux systèmes. La cavité possède un facteur de qualité $Q = \omega_{cav}/\kappa \gg 1$ de sorte que χ et κ sont tous deux petits devant ω_{cav} .

temporelle $\overline{X(t)}_{\Delta t}$ sur une durée Δt . L'indétermination quantique sur le signal expérimental se traduit par la présence d'une fluctuation gaussienne $(\delta X)_{\Delta t}$ de second moment $\kappa/\Delta t$ dans le signal expérimental :

$$\overline{X(t)}_{\Delta t} = \chi \overline{(\sigma_z(t))}_{\Delta t} + (\delta X)_{\Delta t}. \quad (10.101)$$

Le temps d'acquisition minimal pour séparer les deux états du qubit est donc de l'ordre de $t_{\text{mes}} = \kappa/8\chi^2\bar{n}_c$, qui comme attendu correspond à l'inverse du temps d'acquisition d'un bit classique. Remarquons qu'un couplage qubit/champ plus élevé ou une cavité de facteur de cavité plus élevé¹⁶ raccourcissent également le temps d'acquisition.

Les équations (10.97) et (10.99) permettent de calculer l'état joint le long d'une trajectoire à l'instant $t + dt$ en fonction de l'état à l'instant t conditionné au résultat de mesure x associé à l'intervalle correspondant. Le vecteur d'état $|\psi(t + dt|x)\rangle = a_+(t + dt|x)|+\rangle + a_-(t + dt|x)|-\rangle$ conditionné au résultat de mesure x est alors parfaitement spécifié une fois que l'on connaît le ratio

$$\frac{a_+(t + dt|x)}{a_-(t + dt|x)} = \frac{a_+}{a_-} \times \frac{\langle x|e^{-i\chi dt}\alpha_t\rangle}{\langle x|e^{i\chi dt}\alpha_t\rangle} = e^{-(8\chi/\kappa)\sqrt{2\bar{n}_c}dX(t)} \quad (10.102)$$

que nous avons exprimé en fonction de la variation $dX(t)$. Cette équation donne l'actualisation bayésienne quantique (section 9.3.2) du vecteur d'état le long de la trajectoire quantique conditionnée à l'acquisition du signal expérimental [239]. La phase relative de a_+ et de a_- n'étant pas modifiée, le mouvement du vecteur de Bloch associé à $|\psi_{\text{qb}}(t|[X])\rangle$ s'effectue le long d'un méridien de la sphère de Bloch contrairement au cas où on observe le rayonnement de fluorescence. L'équation (10.100) caractérise alors les propriétés statistiques de ce mouvement diffusif.

10.2.6 Les expériences de suivi des trajectoires

La première réalisation d'un suivi de trajectoires quantiques d'un qubit supraconducteur a été réalisée en effectuant une mesure faible en régime dispersif pour un transmon placé dans une cavité tri-dimensionnelle [281]. Afin de maximiser la séparation des deux composantes cohérentes du champ, un amplificateur paramétrique dégénéré supraconducteur fut utilisé [84]. Contrairement au cas discuté à la section 10.2.5, un tel amplificateur rend anisotrope les fluctuations quantiques du signal en sortie, les comprimant en deçà des fluctuations quantiques du vide dans certaines quadratures et les augmentant dans les quadratures orthogonales en vertu du principe d'incertitude de Heisenberg. Bien utilisé, un tel amplificateur permet ainsi de discerner plus facilement deux composantes cohérentes faiblement séparées avant amplification. Dans un premier temps, le suivi de la trajectoire quantique d'un qubit fut utilisé pour stabiliser les oscillations de Rabi [367] puis, au moyen de mesures

16. Qui permet de plus de se placer à un point de fonctionnement où χ est plus élevé.

stroboscopiques, pour stabiliser son évolution autour d'une trajectoire donnée sur la sphère de Bloch [81], démontrant ainsi la maturité des techniques de mesure et contrôle quantique des qubits supraconducteurs.

Plus récemment, la mesure du rayonnement de fluorescence directement émis par le qubit a été réalisée en utilisant un amplificateur non dégénéré, réalisant ainsi le protocole à la section 10.2.5. Ces mesures ont confirmé l'évolution vers l'état fondamental du qubit au travers d'un processus de diffusion de l'état sur la sphère de Bloch [82]. Comme elles opèrent dans des bandes de fréquences bien séparées, la mesure du rayonnement de fluorescence et celui diffusé en régime dispersif ont pu être combinées [159], donnant ainsi accès à des ensembles de trajectoires quantiques associés à la mesure de quantités incompatibles σ_z et σ_- . En jouant sur l'intensité des couplages sous-jacents à ces deux observations effectuées dans l'environnement électromagnétique du qubit, différents régimes de trajectoires purent être obtenus et étudiés.

Ces résultats illustrent comment en théorie quantique, une même dynamique dissipative peut donner lieu à plusieurs ensembles statistiques de trajectoires quantiques qui correspondent à différents choix de mesures effectuées dans l'environnement. L'ambiguïté des trajectoires reflète celle dans l'écriture de l'état intriqué entre le système et son environnement (section 5.1.2).

Les recherches actuelles s'orientent vers l'analyse de la statistique des trajectoires quantiques. En effet, nous savons d'une part quel est le poids statistique d'un ensemble de résultats de mesures $X(t)$ et d'autre part, comment l'état du système change pour un poids de mesure donné, cette dernière équation étant déterministe. On peut alors écrire un formalisme d'intégrale de chemins pour calculer le poids statistique des trajectoires stochastiques quantiques qui partent d'un état initial $|\psi_i\rangle$ à l'instant t_i et qui aboutissent à l'état $|\psi_f\rangle$ à l'instant t_f [89]. Dans cette intégrale fonctionnelle, l'élément stochastique est fourni par la distribution statistique fonctionnelle des résultats de mesure $[X(t)]$ dont les fluctuations sont d'origine purement quantique. Dans le cas markovien qui nous intéresse ici, le poids de chaque trajectoire est obtenu comme l'exponentielle d'une fonctionnelle « action » qui est elle-même une intégrale temporelle. Ceci permet alors de caractériser les trajectoires qui contribuent le plus au moyen d'un principe de minimisation analogue à celui discuté à la section 1.2.3 mais avec un poids associé à chaque trajectoire qui est réel¹⁷. Cette technique théorique, largement développée puis approfondie [90, 91, 282], a été appliquée avec succès à l'étude des trajectoires quantiques d'un circuit supraconducteur sous l'influence combinée d'une force extérieure et d'une mesure faible [373].

10.2.7 Applications des circuits supraconducteurs

Même si les qubits supraconducteurs ont pour motivation originelle la réalisation du calcul quantique selon le modèle par circuits discuté dans le

17. On parle alors d'une méthode du col plutôt que d'une méthode de la phase stationnaire.

chapitre 8, leur potentiel applicatif va bien au-delà [377]. En effet, ils constituent un modèle extrêmement intéressant pour discuter de l'interaction matière-lumière dans le domaine des micro-ondes au point qu'ils ont véritablement fait exploser le domaine de l'optique quantique dans cette partie du spectre électromagnétique [188].

Sans prétendre à l'exhaustivité, surtout dans un domaine aussi évolutif, nous mentionnerons deux applications des circuits supra-conducteurs : la première est la simulation quantique de modèles pertinents en physique de la matière condensée et la seconde est la réalisation d'un calculateur analogique spécialisé dans la résolution de problèmes d'optimisation.

✱ Exemple de simulation quantique

Ainsi, on sait maintenant réaliser des chaînes de jonctions Josephson qui permettent de réaliser des lignes de transmission dont l'impédance caractéristique est voisine de la résistance de von Klitzing $R_K = e^2/h = 25,8 \text{ k}\Omega$ alors que des lignes de transmission basées sur des pistes supraconductrices ne dépassent pas quelques centaines d'ohms. Ces dispositifs permettent alors de réaliser un système qui simule un système à deux niveaux très fortement couplé à une ligne de transmission, un problème largement étudié depuis les années 1980 [253]. Les régimes de fort couplage sont intéressants car, d'une part, leur évolution temporelle ne suit pas une dynamique markovienne effective (section 9.5.3) et, d'autre part, des phénomènes nouveaux peuvent y être observés comme la transition entre oscillations quantiques et localisation du qubit sous l'influence de l'environnement [87]. Aller au-delà de l'approximation des champs tournants qui sous-tend le modèle de Jaynes-Cummings donne aussi accès à divers effets exotiques comme un état fondamental dans lequel le champ électromagnétique est excité [176]. Enfin, ce modèle est lui-même équivalent au problème Kondo d'une impureté magnétique interagissant avec une bande de conduction dans un métal [376], ce qui ouvre une nouvelle fenêtre pour aborder ce problème classique de la physique de la matière condensée [45, 44, 344, 59, 265]. On peut aussi voir la ligne de transmission artificielle comme un matériau non linéaire pour l'optique micro-ondes [241]. Dans ces travaux, les circuits supraconducteurs sont utilisés pour réaliser un modèle particulier de couplage matière/lumière dans le domaine micro-ondes, modèle qui se trouve être en correspondance avec un modèle emblématique de la physique de la matière condensée.

D'autres travaux visent à réaliser des circuits supraconducteurs complexes de manière à simuler plus directement encore des modèles sur réseau étudiés en physique de la matière condensée comme le modèle de Bose-Hubbard [175] introduit pour décrire un supraconducteur granulaire. Du fait de la compétition entre énergie de charge de chaque grain et l'effet Josephson entre les grains, ce modèle présente des transitions de phase à température nulle entre d'une part des phases isolantes où les paires de Cooper sont localisées sur chaque grain et donc le transport d'excitation est bloqué et une phase supraconductrice dans

laquelle une cohérence de phase existe à grande distance dans le réseau [162]. Là encore, le couplage entre rayonnement micro-ondes et circuits supraconducteurs fournit une manière intéressante d'exciter et de sonder le système que l'on cherche à modéliser.

✱ Les machines D-Wave

Les machines construites par l'entreprise canadienne D-Wave sont basées sur des qubits de flux couplés par des inductances mutuelles. Elles sont construites pour trouver le fondamental d'un hamiltonien de type verre de spin d'Ising considéré au chapitre précédent. Dans ces machines, le degré de liberté $s_i = \pm 1$ n'est autre que la chiralité du courant circulant dans la boucle du qubit de flux i .

Les machines D-wave trouvent le fondamental d'un hamiltonien de type Ising en effectuant un recuit, c'est-à-dire en laissant le système évoluer sous l'action des fluctuations et la dissipation vers l'état de plus basse énergie. Dans un système classique, les fluctuations sont d'origine thermique et permettent de franchir des barrières avec un taux donné par la loi d'Arrhenius. Une question actuellement fortement débattue consiste à savoir si une machine D-wave bénéficie d'une quelconque accélération quantique qui proviendrait de l'effet tunnel entre deux configurations de spins pas trop éloignées [224, 318]. En effet, une approche semi-classique de l'effet tunnel montre qu'à la différence du taux de franchissement de barrière d'Arrhenius, le taux de traversée par effet tunnel dépend également de l'épaisseur de la barrière et pourrait donc permettre une accélération significative du franchissement de petites barrières [309]. Malheureusement, les effets de décohérence quantique sont présents et ont sûrement une influence significative bien que mal comprise, même s'il semble établi que la cohérence quantique ne soit pas totalement supprimée au niveau local [247]. Enfin, à ce jour, aucune preuve d'une accélération quantique pour la machine D-wave n'a été apportée.

Les travaux les plus récents visent donc, outre tout ce qui concerne la recherche d'une accélération quantique pour les nouvelles machines D-wave, à comprendre plus quantitativement comment et pour quels types de problèmes l'effet tunnel peut accélérer l'exécution des algorithmes de recuit sur une machine de type D-wave en se concentrant justement sur l'effet de la dépendance en l'épaisseur de la barrière [127]. Sur le front expérimental, des recherches visent à réaliser un dispositif préservant mieux la cohérence quantique et où donc l'accélération quantique serait plus forte [374].

Quoi que nous réserve l'avenir, l'existence des machines D-wave aura eu le mérite de braquer les projecteurs sur les potentialités du calcul par recuit pour de nombreux problèmes d'optimisation combinatoire dans des domaines aussi variés que le spatial [33, 56], la cosmologie [28] ou encore l'optimisation du trafic routier [284].

Chapitre 11

Caractériser les corrélations quantiques

L'étude des systèmes composés en physique quantique a permis de mettre en évidence un de ses aspects les plus fondamentaux et les moins intuitifs, à savoir l'intrication. En effet, elle est à l'origine de corrélations non triviales entre les mesures effectuées sur les parties d'un système bipartite (chapitre 5). L'étude des conséquences de l'intrication, au chapitre 6 avec les différents théorèmes d'impossibilité ainsi qu'au chapitre 7 avec la théorie de la communication et de l'information quantique, nous a permis de commencer à comprendre à quel point l'information quantique est différente de l'information classique.

Peut-on caractériser cette différence de manière plus quantitative et rigoureuse ? Il s'avère que c'est en effet possible. Nous allons voir que les corrélations quantiques sont d'une nature très différente des corrélations permises par une large classe de théories physiques, en particulier les théories classiques locales. Ce résultat dû à J. S. Bell [36] fait suite au travail pionnier d'Einstein, Podolski et Rosen [147] qui, trente ans auparavant, montrèrent que les trois propriétés de localité, réalisme et complétude ne peuvent être compatibles avec les prédictions de la théorie quantique. Il est d'autant plus remarquable que ces résultats amènent à des prédictions qui furent et sont toujours testées dans les laboratoires. L'expérience historique de A. Aspect [20] a permis de valider les prédictions de la théorie quantique et a montré que les corrélations quantiques ne peuvent pas être comprises dans un cadre de pensée classique.

Avec un regard plus théorique (ou philosophique), ces travaux, qui sont très imprégnés du point de vue de la théorie de l'information, sont très singuliers en physique. En effet, ce sont des résultats qui ne traitent pas d'un modèle particulier mais du cadre théorique lui-même. Cet aspect général et abstrait exprime toute la portée, la beauté mais aussi les difficultés conceptuelles du contenu de ce chapitre¹.

1. Pour approfondir l'étude des résultats EPR et Bell, de leurs implications et généralisations, nous recommandons le livre de Laloë [245].

11.1 Le raisonnement EPR

11.1.1 Hypothèses

L'intrication est l'une des propriétés les plus remarquables de la physique quantique et va à l'encontre de notre intuition forgée sur la physique classique. Einstein fut particulièrement troublé par cela, allant jusqu'à parler d'une étrange action à distance associée au fait qu'une mesure sur un sous-système semble affecter instantanément son partenaire intriqué et cela indépendamment de leur séparation. Ceci amena Einstein, Podolski et Rosen (EPR) à l'élaboration de leur fameux article sur l'incomplétude de la théorie quantique [147], article qui historiquement a attiré l'attention sur l'intrication quantique.

Le *raisonnement EPR* est un raisonnement logique fondé sur des hypothèses claires. Il s'apparente ainsi plus à un théorème mathématique. Les hypothèses de départ sont au nombre de trois. Deux portent sur la structure de la théorie considérée :

- la première hypothèse est celle de *réalisme* : si on peut prédire avec certitude la valeur d'une quantité physique, alors la théorie contient un « *élément de réalité* » correspondant à cette quantité ;
- la deuxième hypothèse est celle de *localité* : les prédictions relatives à une région de l'espace-temps ne sont pas affectées par les éléments relatifs à une région distincte de l'espace ².

Enfin, la troisième hypothèse porte sur la comparaison entre la théorie et l'expérience et fait explicitement référence à la théorie quantique :

- les prédictions de la physique quantique sont confirmées par les expériences.

À partir de ces hypothèses, Einstein, Podolski et Rosen arrivent à la conclusion suivante : la physique quantique est une théorie *incomplète*, c'est-à-dire qu'elle ne permet pas de rendre compte de tous les éléments de réalité possibles.

11.1.2 Le raisonnement

Focalisons-nous sur Alice qui choisit de faire une mesure de l'observable A dans des conditions notées α (comme une direction du polariseur dans l'exemple discuté plus bas). Les résultats d'Alice seront parfaitement corrélés avec ceux de Bob si ce dernier choisit un polariseur orienté dans la même direction. En vertu de l'hypothèse de réalisme, cette corrélation parfaite conduit Alice à assigner un élément de réalité à son résultat et à celui de Bob. Alice et Bob étant spatialement éloignés, tandis que toutes les causes et influences sont locales suivant l'hypothèse de localité, il doit donc exister des causes communes locales λ au niveau de la source qui déterminent les résultats expérimentaux et expliquent les corrélations parfaites. Toujours à partir de la localité et de la

2. Cette notion de localité, purement spatiale, est plus élémentaire que celle introduite en relativité restreinte *via* la structure de cône de lumière.

véracité des prédictions de la théorie quantique, les variables λ doivent être indépendantes des choix (α, β) des expérimentateurs et doivent à elles seules permettre de déterminer toutes les réponses aux questions qu'Alice ou Bob peuvent poser. L'argument EPR permet ainsi de conclure que :

- la théorie est déterministe au sens où, pour chaque paire de particule, la variable λ détermine la valeur de n'importe quelle quantité physique. Ce déterminisme est une conséquence et non une hypothèse du raisonnement EPR, un point souligné avec force par Bell [37] ;
- les résultats de mesure d'Alice $A(\alpha, \lambda)$ ou de Bob $B(\beta, \lambda)$ sont fonctions des variables λ et des paramètres expérimentaux (α, β) . Remarquons encore une fois le rôle crucial de l'hypothèse de localité : sans elle nous ne pourrions affirmer que le résultat de la mesure côté Alice n'est fonction que de α et λ et pas de β .

Exposons maintenant en détail le raisonnement EPR reformulé par Bohm qui utilise un système bipartite formé de systèmes à deux niveaux au lieu des variables continues de position et d'impulsion de l'article original. Ce système est partagé entre Alice et Bob qui sont spatialement très éloignés. Chacun a le choix de faire une mesure locale sur son sous-système d'une observable pouvant prendre deux valeurs ± 1 . Le schéma de principe d'une telle expérience est illustré par la figure 11.1. Concrètement, nous pouvons penser à une paire de photons intriqués en polarisation partagée par Alice et Bob. Chacun choisit de mesurer la polarisation de son photon avec des polariseurs linéaires orientés comme il le souhaite. Les valeurs ± 1 correspondent respectivement au cas où le photon passe ou non le polariseur.

Or, dans la théorie quantique, il existe des observables qui ne sont pas compatibles : la mesure d'une de ces quantités avec une précision arbitraire ne nous permet plus de déterminer la valeur de l'autre aussi précisément. Mais l'argument EPR montre que les variables λ déterminent les valeurs de toutes les observables. En conséquence, l'argument EPR montre l'incomplétude de la théorie quantique. Si nous admettons les trois hypothèses du raisonnement EPR, alors il devrait exister une théorie plus fondamentale permettant de prédire les valeurs de toutes les grandeurs physiques côté Alice comme côté Bob.

11.2 L'inégalité de Bell

Le caractère très général et désincarné du raisonnement EPR rend difficile sa confrontation à l'expérience. La contribution de J. Bell fut décisive pour extraire du résultat EPR une prédiction quantitative vérifiable expérimentalement. Son idée repose sur une analyse des corrélations entre mesures faites par Alice et Bob qui permet d'établir l'*inégalité de Bell* qui est violée par la physique quantique standard. Le théorème de Bell [36] est donc une conséquence directe de celui de EPR dont la confrontation à l'expérience permet d'invalidier ou non les idées mises en avant par EPR.

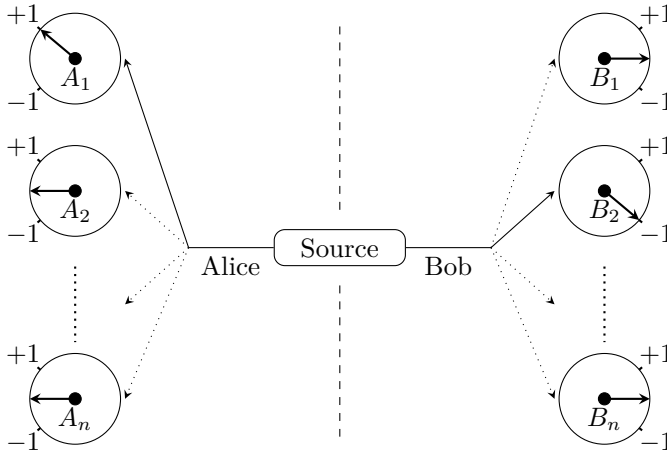


FIGURE 11.1 – Schéma type d’une expérience de Bell : une source engendre un état intriqué à deux quantas qui sont envoyés à Alice et Bob. Chacun effectue dans son laboratoire un choix aléatoire de la quantité qu’il va mesurer (ici dénotée par les A_i pour Alice et les B_i pour Bob) et peut obtenir les résultats ± 1 . On va ensuite comparer les résultats obtenus par les deux expérimentateurs.

11.2.1 Établissement de l’inégalité

On reprend le même cadre expérimental que pour le raisonnement EPR. Cependant, au lieu de choisir une seule configuration de leur appareil de mesure, Alice et Bob en choisissent deux, respectivement (α, α') et (β, β') . Pour l’un des observateurs, par exemple Bob, deux cas de figures se présentent : soit les deux mesures donnent des résultats de même signe, auquel cas $B(\beta, \lambda) - B(\beta', \lambda) = 0$, soit les mesures donnent des résultats opposés. Ainsi, $B(\beta, \lambda) + B(\beta', \lambda) = \pm 2$ suivant le signe obtenu. En considérant une combinaison particulière $\mathcal{B}$ de ces résultats, nous avons

$$\mathcal{B} \equiv A(\alpha, \lambda) [B(\beta, \lambda) - B(\beta', \lambda)] + A(\alpha', \lambda) [B(\beta, \lambda) + B(\beta', \lambda)] = \pm 2 \quad (11.1)$$

car au moins un des deux termes est nul tandis que l’autre vaut ± 2 . Un point clé est de séparer suffisamment les dispositifs de mesure d’Alice et Bob pour qu’ils ne puissent pas s’influencer mutuellement.

À une paire de photons qu’Alice et Bob choisissent d’analyser correspond une valeur de λ . En itérant l’expérience un très grand nombre de fois, Alice et Bob échantillonnent la distribution statistique de cette variable cachée. Leurs données expérimentales permettent d’accéder à la valeur moyenne $C(A_\alpha, B_\beta)$

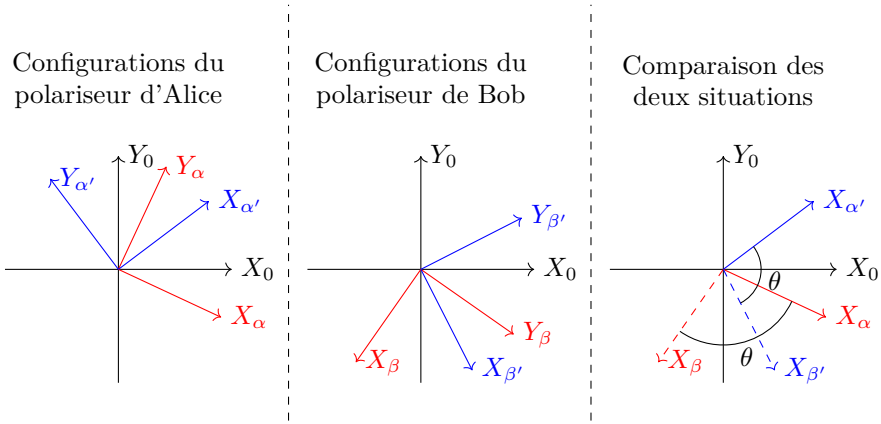


FIGURE 11.2 – Représentation schématique des choix de mesures pour Alice et Bob. Alice se donne deux polariseurs faisant les angles α et α' avec l'horizontale tandis que Bob possède deux polariseurs d'angles β et β' . On suppose pour simplifier que les configurations d'Alice et Bob sont simplement tournées l'une de l'autre par un angle θ .

des corrélations $A(\alpha, \lambda) B(\beta, \lambda)$,

$$C(A_\alpha, B_\beta) = \langle A(\alpha, \lambda) B(\beta, \lambda) \rangle_\lambda, \quad (11.2)$$

ainsi qu'à la valeur moyenne de la quantité $\mathcal{B}$ de Bell qui satisfait à

$$\langle \mathcal{B} \rangle_\lambda \leq 2. \quad (11.3)$$

C'est l'inégalité de Bell sous une forme particulière, due à Clauser, Holt, Shimony et Horne [99], et parfois appelée inégalité CHSH. Cette inégalité est la conséquence du réalisme local sur les corrélations entre des mesures qu'Alice et Bob effectuent. Rappelons de nouveau que cette inégalité est la conséquence de deux hypothèses, celle du réalisme et celle de localité, qui se traduisent par l'existence de variables cachées locales.

11.2.2 Prédictions de la théorie quantique

L'inégalité de Bell est un résultat majeur permettant de tester expérimentalement les fondements conceptuels de la théorie quantique. Pour le comprendre, il faut analyser le même problème dans le cadre quantique.

Spécifions la situation. Alice et Bob se partagent une paire de photons dans l'état maximalement intriqué $|B_{00}\rangle$ de la base de Bell défini par l'équation (5.22) et mesurent la polarisation de leur photon à l'aide de polariseurs. Chacun

choisira deux orientations à sa guise. Les quatre observables correspondent alors à quatre opérateurs hermitiens agissant sur les espaces de Hilbert $\mathcal{H}_A$ et $\mathcal{H}_B$. La séparation spatiale assure une nouvelle fois que les observables de l'un ne dépendent pas de ce que fait l'autre expérimentateur. Les corrélations qui nous intéressent sont de la forme

$$C(A_\alpha, B_\beta) = \langle A_\alpha \otimes B_\beta \rangle_{|\Psi\rangle} \quad (11.4)$$

et on souhaite évaluer la quantité $\mathcal{B}(\alpha, \beta, \alpha', \beta')$ de Bell définie par

$$\mathcal{B}(\alpha, \beta, \alpha', \beta') = C(A_\alpha, B_\beta) - C(A_\alpha, B_{\beta'}) + C(A_{\alpha'}, B_\beta) + C(A_{\alpha'}, B_{\beta'}) . \quad (11.5)$$

Les corrélations (11.4) valent donc

$$\begin{aligned} C(A_\alpha, B_\beta) &= p(A_\alpha = 1, B_\beta = 1) + p(A_\alpha = -1, B_\beta = -1) \\ &\quad - p(A_\alpha = -1, B_\beta = 1) - p(A_\alpha = 1, B_\beta = -1) . \end{aligned} \quad (11.6)$$

Pour calculer ces probabilités, imaginons qu'Alice commence par faire ses mesures. Nous avons donc, avec la règle de Bayes,

$$p(A_\alpha = \epsilon, B_\beta = \epsilon') = p(A_\alpha = \epsilon) p(B_\beta = \epsilon' | A_\alpha = \epsilon) , \quad (11.7)$$

où ϵ, ϵ' représentent n'importe quel résultat de mesure. Compte tenu de l'état que nous avons choisi $p(A_\alpha = \epsilon) = 1/2$. Il s'agit donc maintenant de calculer les probabilités conditionnelles. Après une mesure d'Alice, l'état que possède Bob est projeté par la mesure soit sur $|X_\alpha X_\alpha\rangle$ si Alice mesure 1, soit sur $|Y_\alpha Y_\alpha\rangle$ si Alice mesure -1 , où X et Y représentent les axes de la base de mesure. La figure 11.2 schématise la situation. En fonction de l'angle noté (α, β) entre les axes de mesures des deux observateurs, les probabilités conditionnelles s'écrivent

$$p(B_\beta = 1 | A_\alpha = 1) = p(B_\beta = -1 | A_\alpha = -1) = \cos^2(\alpha, \beta) \quad (11.8a)$$

$$p(B_\beta = -1 | A_\alpha = 1) = p(B_\beta = 1 | A_\alpha = -1) = \sin^2(\alpha, \beta) . \quad (11.8b)$$

Les probabilités conditionnelles sont facilement calculées en utilisant l'envairance qui permet d'écrire

$$|\Psi\rangle = \frac{|X_\theta X_\theta\rangle + |Y_\theta Y_\theta\rangle}{\sqrt{2}} . \quad (11.9)$$

Finalement, l'expression (11.6) se résume à

$$\begin{aligned} C(A_\alpha, B_\beta) &= \frac{1}{2} (\cos^2(\alpha, \beta) + \cos^2(\alpha, \beta)) - \frac{1}{2} (\sin^2(\alpha, \beta) + \sin^2(\alpha, \beta)) \\ &= \cos(2(\alpha, \beta)) . \end{aligned} \quad (11.10)$$

Calculons maintenant l'expression de Bell (11.5), qui est égale à

$$\begin{aligned} \mathcal{B}(\alpha, \beta, \alpha', \beta') &= \cos(2(\alpha, \beta)) - \cos(2(\alpha, \beta')) \\ &\quad + \cos(2(\alpha', \beta)) + \cos(2(\alpha', \beta')) . \end{aligned} \quad (11.11)$$

Cette expression dépend de trois angles indépendants (α, β) , (α, β') et (α', β') car le dernier est fonction des autres *via* la relation $(\alpha', \beta') = (\alpha', \beta) - (\alpha, \beta) + (\alpha, \beta')$. Rechercher le maximum de $\mathcal{B}$ en fonction des trois angles est un problème compliqué mais nous pouvons faire un choix intelligent. Nous allons prendre un dispositif symétrique où l'angle entre les deux axes de référence entre Alice et Bob est le même pour les deux jeux d'observables considérés : $(\alpha, \beta) = (\alpha', \beta') = \theta$ (figure 11.2). Ceci nous donne donc

$$\mathcal{B}(\alpha, \beta, \alpha', \beta') = 2 \cos(2\theta) - \cos(2(\alpha, \beta')) + \cos(2(\alpha', \beta)) . \quad (11.12)$$

Les deux derniers angles sont déterminés par l'angle entre les axes α et α' puisque

$$(\alpha', \beta) = \theta - (\alpha, \alpha') , \quad (11.13a)$$

$$(\alpha, \beta') = \theta + (\alpha, \alpha') . \quad (11.13b)$$

Comme nous voulons n'avoir qu'un seul paramètre, nous avons deux choix naturels, à savoir $(\alpha, \alpha') = \theta$ et $(\alpha, \alpha') = 2\theta - \pi$. Ces choix nous donnent deux expressions pour l'expression de Bell qui sont respectivement

$$\mathcal{B}_\theta(\theta) = 2 - 2 \cos 2\theta (\cos 2\theta - 1) , \quad (11.14a)$$

$$\mathcal{B}_{2\theta-\pi}(\theta) = 3 \cos 2\theta \cos 6\theta . \quad (11.14b)$$

L'étude de $\mathcal{B}_\theta$ montre que les extremum sont atteints pour $\theta = \pm \frac{\pi}{6}$, $\theta = \pm \frac{5\pi}{6}$ et $\theta = \frac{\pi}{2}$. Pour $\mathcal{B}_\theta$, nous avons les valeurs extrémales

$$\mathcal{B}_\theta\left(\pm \frac{\pi}{6}\right) = \mathcal{B}_\theta\left(\pm \frac{5\pi}{6}\right) = \frac{5}{2} \geq 2 , \quad (11.15a)$$

$$\mathcal{B}_\theta\left(\pm \frac{\pi}{2}\right) = -2 . \quad (11.15b)$$

L'étude de $\mathcal{B}_{2\theta-\pi}(\theta)$ montre quant à elle que les extrema sont atteints pour $\theta = \pm \frac{\pi}{8}$, $\theta = \frac{\pi}{2} \pm \frac{\pi}{8}$ et $\theta = \pi - \frac{\pi}{8}$, d'où

$$\mathcal{B}_{2\theta-\pi}\left(\pm \frac{\pi}{8}\right) = \mathcal{B}_{2\theta-\pi}\left(\pm \pi \mp \frac{\pi}{8}\right) = 2\sqrt{2} > 2 , \quad (11.16a)$$

$$\mathcal{B}_{2\theta-\pi}\left(\frac{\pi}{2} \pm \frac{\pi}{8}\right) = \mathcal{B}_{2\theta-\pi}\left(\pm \frac{\pi}{2} \pm \frac{\pi}{8}\right) = -2\sqrt{2} < -2 . \quad (11.16b)$$

La physique quantique prévoit donc qu'il existe des corrélations qui violent l'inégalité de Bell d'un facteur $\sqrt{2}$ (40 % de violation). En d'autres termes, les corrélations quantiques sont plus fortes que n'importe quelles corrélations classiques.

La possibilité d'un test expérimental de l'inégalité est claire. Si elles sont violées, cela veut dire que l'hypothèse qui consiste à affecter aux observables non mesurées une valeur bien définie ne tient pas. Des expériences furent réalisées dans les années 1970 mais il fallut attendre les années 1980 pour obtenir un résultat sans ambiguïté [20]. La conclusion expérimentale est que l'inégalité de Bell est violée.

11.3 Les tests expérimentaux du réalisme local

11.3.1 Historique

Après les articles de Bell puis celui de Clauser *et al.*, les physiciens ont cherché à mettre en œuvre des expériences permettant de tester l'inégalité de Bell. Cependant, une des difficultés de ces expériences était de disposer d'une source de photons intriqués assez lumineuse et capable d'émettre des paires de manière fiable. C'était un défi considérable car disposer d'une source capable d'émettre des paires convenablement préparées avec une fréquence d'émission assez importante était nécessaire pour diminuer le temps d'acquisition permettant une bonne précision statistique sur l'estimation des corrélateurs.

Ce sont les expériences du groupe d'Alain Aspect, proposées dans les années 1970 [18], qui ont permis d'obtenir les premières violations nettes de l'inégalité de CHSH [21]. La troisième génération de leurs expériences permit de plus d'assurer que les mesures étaient faites dans des régions de l'espace-temps suffisamment séparées [20]³.

11.3.2 Les difficultés expérimentales

Compte tenu des enjeux derrière le test du théorème de Bell, il est particulièrement important de s'assurer que toutes les précautions soient prises pour ne pas laisser de place au réalisme local. Or, dans ces expériences, plusieurs problèmes se posent qui peuvent compromettre les conclusions que l'on tire des données expérimentales. Ces *échappatoires* (*loopholes*) font encore l'objet de recherches actives et sont intéressantes à comprendre compte tenu des enjeux épistémologiques qu'elles soulèvent.

Le problème de l'efficacité de détection

Un premier problème est que les détecteurs de photons ne sont pas efficaces à 100 %. Or, en principe, cela pourrait fausser les conclusions du raisonnement. Cela avait déjà été pris en compte dans l'interprétation des expériences en corrigeant la borne de Bell afin de prendre en compte les imperfections expérimentales tant au niveau de la source de paires de photons que de la détection [19]. Mais pour ce qui concerne la détection, l'analyse se basait sur

3. Par un intervalle de genre espace.

une hypothèse sous-jacente forte, à savoir supposer que les paires de photons détectées forment un ensemble représentatif. Cependant, on peut toujours objecter que, pour des raisons mystérieuses, cette hypothèse de représentativité de l'échantillon expérimental pouvait être mise en question. Il était évidemment souhaitable de réaliser des expériences avec une efficacité de détection aussi proche que possible de 100 %. De plus, au niveau théorique, il est possible de trouver une inégalité [144] qui remplace l'inégalité de Bell en présence d'une détection imparfaite et surtout sans faire l'hypothèse d'un échantillonnage représentatif.

Un gros progrès dans cette direction a été réalisé avec une expérience conduite sur des ions piégés [316]. Cette expérience présente une efficacité de détection très proche de 100 % mais ne pouvait, par construction, échapper au second problème qui est lié à la localité. En 2013, Zeilinger et ses collaborateurs résolurent le problème de la détection pour les photons en améliorant les sources de paires de photons et surtout en utilisant des détecteurs de photons supraconducteurs basés sur la transition de phase d'un petit îlot supraconducteur [177].

Le problème de la communication

Dans une expérience de test de l'inégalité CHSH, il est crucial de s'assurer que les mesures soient effectuées en évitant toute influence du système de mesure d'Alice sur celui de Bob (et réciproquement). Cela implique de réaliser les mesures dans des régions séparées par un intervalle de genre espace mais aussi de totalement déconnecter les chaînes de mesure. De plus, dans les expériences avec choix aléatoire, il est indispensable que le choix des quantités mesurées côté Alice et côté Bob soit totalement décorrélié. Le diagramme d'espace-temps 11.3 résume toutes les contraintes pour pallier ce problème.

Technologiquement, c'est un véritable défi et il a fallu attendre la fin du XX^e siècle pour que des expériences échappent totalement au problème de la communication et effectuent un choix purement aléatoire juste avant la détection [375].

Ce n'est que très récemment, en août 2015, qu'un groupe a annoncé clore simultanément les deux *loopholes* liés à la détection et à la localité. Cette expérience [201] combine à la fois une efficacité de détection élevée ($\gtrsim 96\%$) pour éviter de faire appel à l'hypothèse d'échantillonnage représentatif et une séparation spatiale suffisante (1,3 km) pour éviter toute communication entre détecteurs.

Le « problème » du libre arbitre

Il existe une dernière échappatoire que l'on appelle le « problème du libre arbitre » ou l'échappatoire par le « super-déterminisme ». L'idée est que comme les expériences sont effectuées par des détecteurs qui sont des objets physiques dont le cône de lumière passé a forcément eu un recouvrement non vide, il est

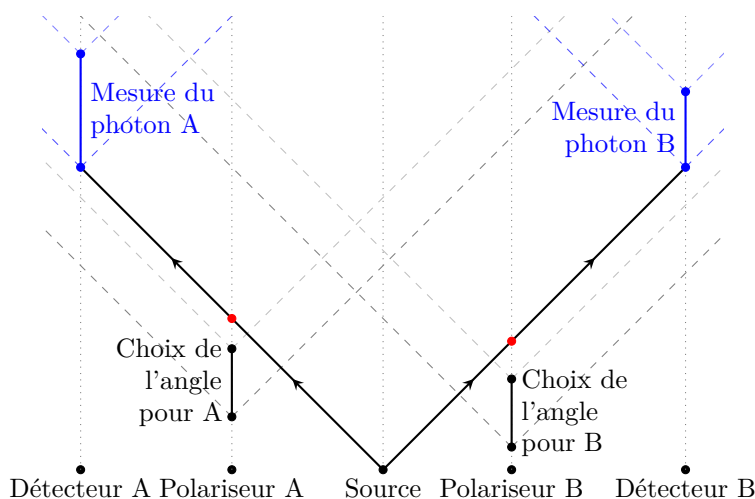


FIGURE 11.3 – Le diagramme d'espace-temps d'une expérience de test de l'inégalité de Bell. Il est absolument nécessaire que la mesure ou le choix du type de polarisation d'un des expérimentateurs ne puisse pas influencer sur la mesure de l'autre : la mesure effectuée par A doit être dans un intervalle de genre espace par rapport à la mesure et au choix de polarisation du côté B.

parfaitement possible qu'un élément de causalité soit à l'origine des choix de mesures et donc de l'apparition des corrélations supérieures à celles données par l'inégalité de Bell alors que, si nous avons véritablement eu le choix, nous n'aurions pas observé de violation de l'inégalité de Bell⁴.

Il faut reconnaître que l'objection est assez capillotractée : il s'agit ni plus ni moins d'invoquer une conspiration du Cosmos induisant les physiciens en erreur. La physique quantique est certes bizarre mais invoquer une telle explication, outre le fait qu'elle met l'ensemble des physiciens au chômage et retire toute valeur épistémologique aux sciences, paraît tellement alambiquée qu'il semble assez légitime d'invoquer l'argument du rasoir d'Ockham pour s'en débarrasser.

4. Comme l'a expliqué John Bell lui-même dans une interview à la BBC en 1985 : « *There is a way to escape the inference of superluminal speeds and spooky action at a distance. But it involves absolute determinism in the universe, the complete absence of free will. Suppose the world is super-deterministic, with not just inanimate nature running on behind-the-scenes clockwork, but with our behavior, including our belief that we are free to choose to do one experiment rather than another, absolutely predetermined, including the "decision" by the experimenter to carry out one set of measurements rather than another, the difficulty disappears. There is no need for a faster than light signal to tell particle A what measurement has been carried out on particle B, because the universe, including particle A, already "knows" what that measurement, and its outcome, will be.* »

Néanmoins, cela n'a pas empêché des physiciens de s'interroger sur les moyens de contrer cette objection. Dans une expérience récente [201], les auteurs montrent qu'ils ne peuvent rejeter ces effets de conspiration cosmique qu'à 690 ns dans le passé. Pour faire mieux, il faut que les générateurs aléatoires qui déterminent le choix des quantités mesurées par Alice et Bob soient suffisamment éloignés pour rejeter l'intersection de leurs cônes de lumières passés assez loin dans le passé (voir figure 11.3). Une proposition originale récente [167] vise donc à utiliser des observations de photons en provenance de quasars lointains ou de régions angulairement éloignées du fond cosmologique comme germe dans le choix aléatoire des quantités mesurées dans une expérience de corrélations à la Alain Aspect. Pour que de tels événements d'émission de photons conspirer à induire les expérimentateurs en erreur, il faudrait donc qu'ils aient été corrélés des milliards d'années dans le passé, repoussant de 20 ordres de grandeur dans le passé la grande conspiration cosmique.

11.4 Compléments sur les inégalités

Avant d'examiner les conséquences conceptuelles de la violation expérimentale de l'inégalité de Bell, nous allons explorer quelques points complémentaires. Dans un premier temps, nous vérifierons que les états mélanges d'états factorisés vérifient l'inégalité de Bell alors qu'un état qui n'est pas maximalelement intriqué la viole. Nous discuterons ensuite un autre test qui met en jeu trois systèmes intriqués. Dans ce cas, une seule mesure permet de mettre en évidence un comportement qui ne peut pas être décrit par une théorie réaliste locale.

11.4.1 Le cas d'un état mélange

Jusqu'à maintenant, tous les calculs ont été effectués avec un état quantique bien spécifique, de la forme $(|00\rangle + |11\rangle)/\sqrt{2}$. On peut se demander quelles sont les valeurs que l'on obtiendrait pour d'autres types d'états. Ainsi, prenons cette fois un état non intriqué, mélange équiprobable des états $|XX\rangle$ et $|YY\rangle$. Nous allons calculer le signal de corrélation de Bell. Pour cela, nous allons regarder la corrélation (11.6). Notons que

$$|X_\alpha\rangle = \cos \theta_\alpha |X\rangle + \sin \theta_\alpha |Y\rangle, \quad (11.17a)$$

$$|Y_\alpha\rangle = \cos \theta_\alpha |Y\rangle - \sin \theta_\alpha |X\rangle. \quad (11.17b)$$

Nous avons alors les amplitudes suivantes en fonction des angles θ_α et θ_β

$$\langle X_\alpha X_\alpha | XX \rangle = \cos \theta_\alpha \cos \theta_\beta, \quad (11.18a)$$

$$\langle X_\alpha Y_\beta | XX \rangle = -\cos \theta_\alpha \sin \theta_\beta, \quad (11.18b)$$

$$\langle Y_\alpha X_\alpha | XX \rangle = -\sin \theta_\alpha \cos \theta_\beta, \quad (11.18c)$$

$$\langle Y_\alpha Y_\alpha | XX \rangle = \sin \theta_\alpha \sin \theta_\beta. \quad (11.18d)$$

Grâce à ces amplitudes, la corrélation (11.6) dans l'état $|XX\rangle$ s'écrit

$$\langle A_\alpha B_\beta \rangle_{|XX\rangle} = \cos 2\theta_\alpha \cos 2\theta_\beta. \quad (11.19)$$

Nous pouvons vérifier que nous avons la même corrélation pour l'état $|YY\rangle$. Ainsi, dans un mélange statistique équiprobable⁵ de ces deux états, nous avons

$$C(A_\alpha, B_\beta) = \cos 2\theta_\alpha \cos 2\theta_\beta \quad (11.20)$$

alors que pour un état intriqué, nous avons

$$C_{\text{Bell}}(A_\alpha, B_\beta) = \cos 2(\theta_\alpha - \theta_\beta). \quad (11.21)$$

L'expression que nous avons obtenue pour un mélange donne des corrélations qui sont factorisées. En conséquence, le raisonnement de Bell peut être repris quasiment à l'identique et cela nous montre que pour un état mélange

$$|\mathcal{B}(\alpha, \beta, \alpha', \beta')| \leq 2. \quad (11.22)$$

Le mélange statistique d'états factorisés ne viole donc pas l'inégalité de Bell. Attention il ne faut cependant pas en conclure trop hâtivement que tous les mélanges statistiques vérifient l'inégalité de Bell : certains mélanges d'états intriqués peuvent la violer.

11.4.2 Le cas d'un état pur général

Pour préciser encore ces résultats, regardons maintenant le cas d'un état pur qui n'est pas maximalelement intriqué, c'est-à-dire de la forme générique

$$|\psi_\lambda\rangle = \sqrt{\lambda}|XX\rangle + \sqrt{1-\lambda}|YY\rangle, \quad \lambda \in [0, 1]. \quad (11.23)$$

Il est alors facile grâce à ce que nous avons fait dans la section précédente de calculer les amplitudes de probabilité analogues à (11.18) pour l'état $|\psi_\lambda\rangle$. Ceci nous permet alors de calculer la corrélation $C_\lambda(A_\alpha, B_\beta)$ pour cet état qui vaut alors

$$\begin{aligned} C_\lambda(A_\alpha, B_\beta) &= \lambda C_1(A_\alpha, B_\beta) + (1-\lambda) C_0(A_\alpha, B_\beta) \\ &\quad + 2\sqrt{\lambda(1-\lambda)} \cos \theta_\alpha \cos \theta_\beta \sin \theta_\alpha \sin \theta_\beta, \end{aligned} \quad (11.24)$$

où les corrélations $C_0(A_\alpha, B_\beta)$ et $C_1(A_\alpha, B_\beta)$ correspondent à celles calculées dans la section précédente. Ceci nous permet d'écrire notre corrélation comme

$$C_\lambda(A_\alpha, B_\beta) = \cos 2\theta_\alpha \cos 2\theta_\beta + 2\sqrt{\lambda(1-\lambda)} \sin 2\theta_\alpha \sin 2\theta_\beta. \quad (11.25)$$

5. Ce résultat est complètement indépendant des proportions des états.

Nous cherchons toujours à calculer l'indicateur de Bell $\mathcal{B}_\lambda(\alpha, \beta, \alpha', \beta')$. Pour regarder si on peut obtenir une violation, considérons $\theta_\beta = 0$ et $\theta_{\beta'} = \pi/4$ de sorte que

$$C_\lambda(A_\alpha, B_\beta) = \cos 2\theta_\alpha, \quad (11.26a)$$

$$C_\lambda(A_{\alpha'}, B_\beta) = \cos 2\theta_{\alpha'}, \quad (11.26b)$$

$$C_\lambda(A_\alpha, B_{\beta'}) = 2\sqrt{\lambda(1-\lambda)} \sin 2\theta_\alpha, \quad (11.26c)$$

$$C_\lambda(A_{\alpha'}, B_{\beta'}) = 2\sqrt{\lambda(1-\lambda)} \sin 2\theta_{\alpha'}. \quad (11.26d)$$

Si nous supposons de plus que $\theta_{\alpha'} = -\theta_\alpha = \theta$, nous obtenons l'expression suivante pour l'indicateur de Bell

$$\mathcal{B}(\alpha, \beta, \alpha', \beta') = 2 \cos(2\theta) - 4\sqrt{\lambda(1-\lambda)} \sin 2\theta. \quad (11.27)$$

En développant le membre de droite à petit θ , nous voyons immédiatement que dès que $0 < \lambda < 1$, il est possible de choisir θ de manière à avoir une violation de l'inégalité de Bell. En conséquence, même des états qui ne sont pas maximalelement intriqués donnent lieu à des violations de l'inégalité de Bell.

11.4.3 Théorème de Tsirelson

La physique quantique prévoit une violation de l'inégalité de Bell d'un facteur $\sqrt{2}$. Une question que l'on peut se poser est de savoir si cette borne est la borne supérieure autorisée par la théorie quantique. Le théorème de Tsirelson répond par l'affirmative à cette question. Il existe des preuves directes de ce résultat dans de nombreux livres d'information quantique. Ici, nous allons exposer une idée de preuve à travers un résultat plus général dû à Grothendieck [186, 6].

Reprenons le problème de l'inégalité de Bell. Cette dernière est construite au travers de corrélations $C_{\alpha\beta}$ entre Alice et Bob dans un état donné⁶. Dans une théorie à variables cachées, nous avons $C_{\alpha\beta}^{\text{cl}} = \int A_\alpha(\lambda) B_\beta(\lambda) P(\lambda) d\lambda$ où λ représente les variables cachées et $P(\lambda)$ leur distribution de probabilité. Dans le cas quantique, nous avons $C_{\alpha\beta} = \text{tr}(\rho A_\alpha B_\beta) = \langle \psi_{AB} | A_\alpha B_\beta | \psi_{AB} \rangle$. Comme dans les exemples précédents, les résultats de mesures sont binaires. L'inégalité de Bell porte sur des quantités qui sont de la forme $\mathcal{B}(a) = \sum_{(\alpha,\beta)} a_{\alpha\beta} C_{\alpha\beta}$ où

6. On supposera dans la suite que les observables sont de moyenne nulle.

a est une matrice réelle. Dans le cas classique, nous avons donc

$$\begin{aligned} |\mathcal{B}_{\text{cl}}(a)| &= \left| \sum_{\alpha, \beta} \int a_{\alpha\beta} A_{\alpha}(\lambda) B_{\beta}(\lambda) P(\lambda) d\lambda \right| \\ &\leq \underbrace{\int_{=1} P(\lambda)}_{=1} \times \underbrace{\left(\max_{\substack{A_{\alpha} = \pm 1 \\ B_{\beta} = \pm 1}} \left| \sum_{\alpha, \beta} a_{\alpha\beta} A_{\alpha} B_{\beta} \right| \right)}_{B[a]}. \end{aligned} \quad (11.28)$$

La borne $B[a]$ est appelée *borne de Bell*. Nous avons vu précédemment comment la calculer dans le cas particulier $a = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. L'inégalité de Bell peut alors se réécrire un peu différemment comme $\left| \sum_{\alpha, \beta} a_{\alpha\beta} C_{\alpha\beta}^{\text{cl}} \right| \leq B[a]$ soit encore

$$\sup_{A_{\alpha}, B_{\beta}} \left| \sum_{\alpha, \beta} a_{\alpha\beta} A_{\alpha} B_{\beta} \right| \leq B[a]. \quad (11.29)$$

Dans le cas quantique, la formulation est bien différente car nous avons

$$\langle \mathcal{B}(a) \rangle = \sum_{\alpha, \beta} a_{\alpha\beta} \langle \psi_{AB} | A_{\alpha} B_{\beta} | \psi_{AB} \rangle = \sum_{\alpha, \beta} a_{\alpha\beta} \langle X_{\alpha} | Y_{\beta} \rangle, \quad (11.30)$$

où $X_{\alpha} = A_{\alpha} | \psi_{AB} \rangle$, $Y_{\beta} = B_{\beta} | \psi_{AB} \rangle$ et où $\langle x_{\alpha} | y_{\beta} \rangle$ désigne le produit scalaire de deux vecteurs dans un espace de Hilbert. Un théorème dû à Grothendieck énonce le résultat suivant [186].

Si pour (x_{α}) et (y_{β}) deux N -uplets de réels de norme inférieure à l'unité, nous avons

$$\left| \sum_{\alpha, \beta} M_{\alpha\beta} x_{\alpha} y_{\beta} \right| \leq 1$$

pour $M \in \mathcal{M}_n(\mathbb{R})$, alors pour tout espace de Hilbert $\mathcal{H}$, $(|X_{\alpha}\rangle)$, $(|Y_{\beta}\rangle)_{\beta}$ deux familles de vecteurs de $\mathcal{H}$, nous avons

$$\left| \sum_{\alpha, \beta} M_{\alpha\beta} \langle X_{\alpha} | Y_{\beta} \rangle \right| \leq K_N \sup_{\alpha} (\| |X_{\alpha}\rangle \|) \sup_{\beta} (\| |Y_{\beta}\rangle \|)$$

où K_N est une constante, appelée *constante de Grothendieck*, qui ne dépend que de N .

Nous voyons donc immédiatement la pertinence du théorème en question. Il suffit de prendre $M_{\alpha\beta} = a_{\alpha\beta}/B[a]$, $X_{\alpha} = A_{\alpha} | \psi_{AB} \rangle$ et $Y_{\beta} = B_{\beta} | \psi_{AB} \rangle$ pour

avoir

$$\left| \sum_{\alpha, \beta} a_{\alpha\beta} \langle X_\alpha | Y_\beta \rangle \right| \leq K_N B[a] \sup_{\alpha} (\|X_\alpha\|) \sup_{\beta} (\|Y_\beta\|) . \quad (11.31)$$

Comme le spectre des opérateurs en présence est égal à ± 1 , nous avons

$$\langle X_\alpha | X_\alpha \rangle = \langle A_\alpha^\dagger A_\alpha \rangle_{|\psi_{AB}\rangle} \leq 1 . \quad (11.32)$$

Les bornes supérieures présentes dans l'inégalité de Grothendieck sont donc inférieures à l'unité. Ceci nous permet d'affirmer que, si les corrélations obéissent à une inégalité de Bell de la forme

$$\left| \sum_{\alpha, \beta} a_{\alpha\beta} C_{\alpha\beta}^{\text{cl}} \right| \leq B[a] , \quad (11.33)$$

alors les corrélations quantiques impliquant N observables vérifient

$$\left| \sum_{\alpha, \beta} a_{\alpha\beta} \langle A_\alpha B_\beta \rangle_{|\psi_{AB}\rangle} \right| \leq K_N B[a] . \quad (11.34)$$

Dans le cas $N = 2$, la constante de Grothendieck vaut $\sqrt{2}$ [240, 58], ce qui est bien inférieur à la limite mathématique de 4 que l'on trouve pour une somme de 4 variables valant ± 1 . Nous venons d'obtenir le *théorème de Tsirelson* stipulant que la borne supérieure des corrélations de Bell autorisées par la théorie quantique est effectivement $2\sqrt{2}$.

11.4.4 Contradiction GHZ

La configuration expérimentale fut jusqu'ici composée de seulement deux sous-systèmes. Greenberger, Horne et Zeilinger (GHZ) comprirent que l'on pouvait obtenir de nouveaux tests des hypothèses EPR en considérant des états intriqués à trois particules [183]. La contradiction entre les résultats GHZ et la physique quantique est plus impressionnante que celle mise en évidence par Bell puisqu'ici les résultats seront strictement opposés. Du point de vue conceptuel, on teste la même chose dans une expérience de Bell ou GHZ, mais il est toujours bienvenu d'avoir différentes expériences pour étudier la même physique. Bien qu'étant plus difficile à mettre en œuvre au laboratoire, l'avantage significatif du schéma GHZ est que la violation peut être obtenue avec des mesures sur un seul état alors que l'expérience de Bell nécessite d'accumuler de la statistique⁷.

7. En pratique, de la statistique doit toujours être faite puisque les expériences ne sont jamais parfaites.

Détaillons la situation GHZ. L'expérience consiste à faire une série de mesures compatibles sur un état maximalement intriqué de trois particules partagées entre Alice, Bob et Carole. Un état maximalement intriqué est un état ayant des corrélations parfaites à trois corps tandis qu'à deux corps on a seulement un mélange statistique. Un *état GHZ* est un état de la forme

$$|\psi_{\text{GHZ}}\rangle = \frac{|\uparrow_z \uparrow_z \uparrow_z\rangle - |\downarrow_z \downarrow_z \downarrow_z\rangle}{\sqrt{2}}. \quad (11.35)$$

Si chaque observateur fait une mesure de l'observable σ_x et que nous faisons le produit des résultats de ces trois mesures, le résultat sera toujours égal à -1 ,

$$\sigma_x^A \sigma_x^B \sigma_x^C = -1. \quad (11.36)$$

De manière similaire, nous pouvons développer une paire sur la base des y et la dernière partie sur x ⁸. Un calcul direct montre alors que le produit des résultats de mesures des observables $(\sigma_y^A, \sigma_y^B, \sigma_x^C)$ donnera un. Trois relations sont alors à disposition

$$\sigma_y^A \sigma_y^B \sigma_x^C = 1, \quad (11.37a)$$

$$\sigma_x^A \sigma_y^B \sigma_y^C = 1, \quad (11.37b)$$

$$\sigma_y^A \sigma_x^B \sigma_y^C = 1. \quad (11.37c)$$

La physique quantique nous fournit donc quatre relations. Regardons maintenant ce qu'un raisonnement EPR implique pour un état avec les corrélations GHZ. Tout d'abord, toutes les mesures faites par nos trois expérimentateurs sont locales. De par les corrélations parfaites, une mesure d'Alice lui permet de déduire les orientations des spins de Bob et Carole, sans les perturber. Selon EPR, il doit donc exister des éléments de réalité associés à ces quantités, qui doivent alors vérifier les relations (11.37). La localité implique que les valeurs $\sigma_{(x,y)}^{(A,B,C)}$ mesurées sont les mêmes pour les trois relations. Or, le carré de chacun de ces nombres étant égal à l'unité, le produit des trois relations (11.37) nous donne

$$\sigma_x^A \sigma_x^B \sigma_x^C = 1. \quad (11.38)$$

Ceci est en contradiction directe avec le résultat obtenu par la physique quantique (11.36). Comme annoncé, un seul état suffit pour obtenir une violation des hypothèses EPR. Encore une fois le résultat expérimental est en accord avec les prédictions de la physique quantique (11.36).

11.5 Conséquences conceptuelles

La violation de l'inégalité de Bell permet de conclure qu'au moins une des hypothèses du théorème EPR doit être abandonnée. Comme les prédictions

8. Par exemple Alice et Bob sur la base y et Carole sur x .

de la théorie quantique n'ont jusqu'à présent jamais été mises en défaut, les hypothèses restantes sont la localité et le réalisme.

Dans une perspective où la physique quantique constitue une description valide et complète de l'univers, la question de la localité dans cette théorie est déjà subtile et il serait trop naïf d'affirmer la non-localité de la théorie sans nuances. En effet, d'un côté, tous les modèles utilisés actuellement pour décrire les interactions fondamentales sont des théories des champs locales : l'interaction entre champs ne couple pas des champs en des points différents de l'espace-temps. De même, toujours dans le cadre des théories quantiques relativistes des champs, des observables localisées dans des régions séparées par un intervalle de genre espace commutent. En ces deux sens (non exclusifs), les théories quantiques décrivant les interactions fondamentales sont locales sans conflits entre la théorie quantique et la localité. Enfin, comme nous l'avons vu dans le chapitre 6, la physique quantique ne permet pas de communiquer de l'information plus vite que la lumière : il n'y a donc pas non plus conflit entre causalité relativiste⁹ et théorie quantique.

Cependant, ce que nous révèle la violation de l'inégalité de Bell, c'est que la notion d'état quantique possède une forme de non-localité. Dans la description de l'expérience de pensée EPR, l'état quantique associé à une paire de photons intriqués n'est pas factorisable en un produit d'états associés chacun à un corpuscule. L'intrication de l'état n'est donc pas localisée dans l'espace. Cette nécessaire non-localité est d'ailleurs parfaitement illustrée dans la reformulation équivalente de Bohm de la mécanique quantique non relativiste en termes de variables cachées non locales¹⁰. Quoi qu'il en soit, il est de mise de rester prudent sur cette forme de non-localité : en effet, il faut toujours garder en tête qu'un état quantique n'est qu'une notion relative à un observateur interagissant localement sur le système et ne pouvant faire que des prédictions sur des mesures locales.

Pour résumer, bien que les états quantiques, qui sont des objets essentiels dans la théorie quantique, soient non locaux, la théorie quantique ne permet pas l'existence d'effets manifestement non locaux comme la communication instantanée. Elle permet cependant l'existence de corrélations supérieures à celles permises par une théorie réaliste et locale, corrélations qui ne sont révélées seulement que si les différents observateurs communiquent entre eux [184, 185].

Ceci nous amène donc à considérer plus précisément la question du réalisme dans la physique quantique. Et, là aussi, la situation est subtile. Le raisonnement d'EPR conduit à l'existence d'éléments de réalité sous la forme de variables λ qui déterminent les valeurs de toutes les quantités physiques (à savoir les mesures faites par n'importe quel polariseur installé chez Alice ou chez Bob). En pratique, Alice et Bob choisiront chacun une direction de polarisation qui est effectivement mesurée. Mais pour toutes les autres orienta-

9. Parfois simplement aussi appelée localité, raison supplémentaire pour faire attention aux qualificatifs que l'on emploie !

10. Cependant, une telle approche rencontre des difficultés avec la théorie de la relativité.

tions, la mesure n'est pas effectuée. Le réalisme local d'EPR conduit à affecter une valeur bien définie à ces quantités qui, répétons-le, ne sont pas mesurées. Le raisonnement EPR conduit donc à des théories contrefactuelles, c'est-à-dire qui affectent des valeurs bien définies à des quantités qui ne sont pas mesurées.

La violation expérimentale des inégalités peut être vue comme la réfutation de la contrefactualité et donc des théories réalistes locales. Dans le cadre de la physique quantique, cette idée est naturelle : dans une expérience, une fois une mesure projective effectuée, l'état du système est irrémédiablement affecté. Parler de la valeur d'une quantité non mesurée n'est tout simplement pas possible. Comme l'a exprimé avec force Asher Peres [297] : « Les expériences non effectuées n'ont pas de résultats » (« *Unperformed experiments have no results* »).

Pouvons-nous dire pour autant que la théorie quantique remet totalement en cause le postulat de réalisme ? Pas totalement car l'état quantique est précisément un élément de réalité. En effet, prenons l'exemple d'un photon polarisé préparé dans un état pur. L'état quantique du photon est l'élément de réalité qui prédit cela avec certitude. Dans le cas de la paire de photons intriquée, préparée dans un état de Bell, nous avons vu au chapitre 7 comment réaliser un appareil de mesure qui donnera toujours le même résultat. L'état intriqué, bien que non factorisable et donc non local, constitue donc un élément de réalité associé à ce résultat de mesure certain. Par contre, dans un univers quantique, il existe toujours des quantités physiques dont la valeur ne peut être prédite avec certitude même lorsque l'on connaît parfaitement l'état quantique. La physique quantique n'est donc pas déterministe au sens où toute quantité potentiellement mesurable devrait avoir une valeur bien déterminée. Même si on peut voir un état quantique comme un élément de réalité, il ne s'agit pas d'un élément de réalité déterministe au sens d'Einstein, Podolski et Rosen.

Nous voyons donc émerger deux caractéristiques conceptuelles de la physique quantique : tout d'abord, c'est une théorie réaliste mais de manière limitée. En effet, d'une part, elle ne permet pas de considérer des quantités physiques contrefactuelles et, d'autre part, les éléments de réalité que sont les états quantiques sont inaccessibles expérimentalement sur une réalisation unique du système physique considéré. Ils ne peuvent être que reconstruits à partir de l'information statistique que l'on peut extraire d'un grand nombre d'expériences.

Ensuite la théorie quantique a des caractéristiques non locales mais, là encore, pas trop non locale. Il n'y a pas incompatibilité entre la localité, la causalité et la théorie quantique : d'une part, les interactions fondamentales sont décrites par des interactions locales dans l'espace-temps et, d'autre part, les corrélations non classiques qui découlent de la non-localité des états quantiques ne permettent pas la communication instantanée.

Nous commençons ainsi à entrevoir comment la théorie quantique bouscule nos conceptions de notions aussi ancrées dans notre expérience quotidienne que le sont la localité et le réalisme. Face à cela, deux attitudes sont possibles : nous

pouvons nous accrocher aux concepts forgés par notre intuition quotidienne comme l'espace, le temps, la notion de réalité objective et, finalement, considérer que la théorie quantique n'est pas une théorie complète pour décrire l'univers. Ou bien, nous pouvons prendre de la distance avec notre intuition quotidienne et plonger dans l'inconnu pour explorer les conséquences de la théorie quantique jusqu'au bout.

Conclusion

Après cette exploration des relations entre théorie quantique, théorie de l'information et théorie du calcul, où sommes-nous arrivés ?

Le premier enseignement que nous souhaitons transmettre au lecteur est que la théorie quantique n'est plus seulement le socle permettant une modélisation précise de la physique microscopique (particules élémentaires, noyaux, atomes et molécules) mais qu'elle est devenue une ressource à part entière permettant de développer de nouvelles technologies basées sur des concepts comme les interférences quantiques et l'intrication.

Nous espérons donc que l'ingénieur ou le chercheur débutant dans ce nouveau domaine retiendra de ce premier volume le rôle essentiel joué par l'intrication. Celle-ci permet en effet de définir les mesures les plus générales (section 5.4) ainsi que la notion de canal de communication quantique (section 7.1). Elle détermine également la dynamique des systèmes quantiques ouverts (chapitre 9). Dans les technologies quantiques, l'intrication quantique devient une ressource permettant de communiquer (chapitre 7) ou de calculer (chapitre 8) différemment et potentiellement plus efficacement. Une des leçons essentielles à retenir du présent volume est que le calcul quantique n'est ni une baguette magique permettant de calculer l'incalculable au sens du Turing, ni une parallélisation qui donnerait accès à un nombre exponentiel de résultats : c'est juste une manière de concentrer astucieusement l'information relative à la solution d'un problème au moyen des interférences quantiques.

Bien entendu, l'expert en technologies quantiques aura remarqué que nous n'avons pas traité l'intégralité des développements récents du domaine. En particulier, nous n'avons pas développé la métrologie quantique qui est l'art de concevoir de nouveaux instruments de mesures exploitant les interférences quantiques. De même, les notions de base sur les codes correcteurs d'erreurs présentées à la section 8.4.3 ne prétendent pas couvrir l'ensemble de ce thème. Cependant, le lecteur du présent ouvrage devrait être bien armé pour approfondir ces deux sujets notamment *via* l'excellent cours sur la métrologie quantique au Collège de France [119] et le livre [287] pour les codes correcteurs d'erreurs quantiques.

Enfin et surtout, nous avons limité notre exploration des systèmes physiques utilisés pour les technologies quantiques à deux incarnations de l'électrodynamique en cavité (chapitre 10) afin d'illustrer simplement et efficacement les

concepts de base présentés dans ce volume. Mais les dispositifs à base d'ions piégés, d'atomes neutres refroidis dans des pièges magnéto-optiques, ou encore les systèmes optiques ou électroniques quantiques auraient tout autant mérité d'être développés. Cependant, cela nous aurait conduit à consacrer un volume entier aux systèmes physiques pour les technologies quantiques. Les excellents cours au Collège de France de Claude Cohen-Tannoudji, Serge Haroche, Michel Devoret puis maintenant Jean Dalibard constituent une excellente base pour découvrir la physique de ces systèmes [191, 104, 117].

Le second enseignement majeur que nous voulons transmettre se situe sur un plan plus conceptuel. Pour nous, un des apports majeurs de la théorie quantique que nous avons commencé à entrevoir est le caractère relatif de l'état quantique. Pour bien comprendre le changement de paradigme que cela représente, revenons un instant sur le développement de la physique au cours des derniers siècles.

La physique classique, de la mécanique jusqu'à l'électromagnétisme de Maxwell ou la gravitation universelle de Newton, est d'une efficacité redoutable pour décrire la majorité des phénomènes physiques allant des échelles mésoscopiques aux échelles astrophysiques¹¹. Mais la physique classique suppose que chaque système est décrit par un état objectif, indépendant de l'observateur et de tout processus physique impliquant ce dernier. Dans l'univers classique de Newton, les phénomènes physiques ne sont que le reflet de l'essence des systèmes, qui pré-existent en dehors de toute interaction. Leurs propriétés physiques sont révélées à l'observateur par une mesure dont le processus précis n'a aucune importance : l'observateur accède à la réalité qui, existant indépendamment de lui, est objective.

À la lumière de ce que nous avons appris dans ce volume, cette vision du monde est remise en cause par la théorie quantique : l'état d'un système n'a pas de sens en dehors du contexte dans lequel on le définit. Parce que toute mesure implique une interaction avec l'observateur qui altère l'état du système, il n'existe pas d'état quantique absolu. Il n'existe que des états relatifs à d'autres sous-systèmes. Bien que nous ayons déjà commencé à explorer les conséquences conceptuelles de ce changement de paradigme à travers la notion d'intrication et ses diverses conséquences, cette rupture ontologique appelle la question suivante : quelles sont les conséquences de la relativité des états quantiques sur notre vision du monde ? Telle est la question centrale à laquelle sera consacré le second volume de cet ouvrage.

Pour l'aborder, notre attitude consistera à rester humbles face à ce que les meilleures théories physiques du moment ont à nous dire sur le monde. Nous essaierons donc de minimiser l'introduction de préjugés sur comment le monde doit être et nous laisserons, autant que possible, ces théories nous dicter leur propre interprétation et remettre en question nos certitudes. Nous verrons donc comment l'ensemble des histoires abordées dans ce volume et le suivant

11. La notion d'échelle est à prendre dans un sens très large. Elle peut correspondre à une distance, une énergie, une pression, une température, etc.

dessinent la perspective d'une pensée cohérente d'un monde complètement régi par les lois de la théorie quantique, mais aussi nous permettent d'en appréhender les limites et de mieux cerner les questions qui restent à bien poser, notamment sur la nature quantique de l'espace et du temps.

Bibliographie

- [1] Aaronson S. *Nature Physics*, 11(4):291, 2015. 336
- [2] Aaronson S. $P \stackrel{?}{=} NP$. <https://www.scottaaronson.com/papers/npnp.pdf>, 2017. 125, 338
- [3] Aaronson S. and Arkhipov A. *Theory of Computing*, 9(1):143–252, 2013. 335
- [4] Abeyesinghe A., Devetak I., Hayden P., and Winter A. *Proc. Roy. Soc. A*, 465:2537, 2009. 268, 275, 290
- [5] Acín A., Bagan E., Baig M., Masanes L., and Muñoz Tapia R. *Phys. Rev. A*, 71:032338, 2005. 225
- [6] Acín A., Gisin N., and Toner B. *Physical Review A*, 73(6):062105, 2006. 461
- [7] Adleman L. M., DeMarrais J., and Huang M.-D. A. *SIAM Journal on Computing*, 26(5):1524–1540, 1997. 331, 333
- [8] Agrawal M., Kayal N., and Saxena N. *Annals of mathematics*, pages 781–793, 2004. 122
- [9] Aharonov D. *arXiv preprint quant-ph/0301040*, 2003. 345
- [10] Aharonov D., Ben-Or M., Impagliazzo R., and Nisan N. *arXiv preprint quant-ph/9611028*, 1996. 353
- [11] Aharonov D., van Dam W., Kempe J., Landau Z., Lloyd S., et al. *SIAM Rev.*, 50:755, 2008. 360
- [12] Albash T. and Lidar D. A. *Rev. Mod. Phys.*, 90:015002, 2018. 360, 363
- [13] Altshuler B., Krovi H., and Roland J. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28):12446–12450, 2010. 362, 363
- [14] Ambainis A. *Journal of Computer and System Sciences*, 64(4):750–767, 2002. 303
- [15] Anderson P. W. *Phys. Rev.*, 109:1492–1505, 1958. 362
- [16] Arndt M., Nairz O., Vos-Andreae J., Keller C., van der Zouw G., et al. *Nature*, 401:680, 1999. 14
- [17] Artjuhov M. *Acta Arith*, 12(355-364):67, 1966. 130
- [18] Aspect A. *Phys. Rev. D*, 14:1944–1951, 1976. 168, 456
- [19] Aspect A. *Trois tests expérimentaux des inégalités de Bell par mesure de corrélation de polarisation de photons*. PhD thesis, Université de

- Paris Sud, 1983. [456](#)
- [20] Aspect A., Dalibard J., and Roger G. *Phys. Rev. Lett.*, 49:1804, 1982. [168](#), [449](#), [456](#)
 - [21] Aspect A., Grangier P., and Roger G. *Phys. Rev. Lett.*, 49:91–94, 1982. [456](#)
 - [22] Aspuru-Guzik A., Dutoi A. D., Love P. J., and Head-Gordon M. *Science*, 309:1704–1707, 2005. [358](#)
 - [23] Assalini A., Dalla Pozza N., and Pierobon G. *Phys. Rev. A*, 84:022342, 2011. [225](#)
 - [24] Astafiev O., Komiyama S., Kutsuwa T., and Antonov V. *Appl. Phys. Lett.*, 80:4250, 2002. [434](#)
 - [25] Astafiev O., Pashkin Y. A., Nakamura Y., Yamamoto T., and Tsai J. S. *Phys. Rev. Lett.*, 93:267007, 2004. [429](#)
 - [26] Baeurle S. A. and Nogovitsin E. A. *Polymer*, 48(16):4883–4899, 2007. [356](#)
 - [27] Ballentine L. *American Journal of Physics*, 54(10):883–889, 1986. [13](#)
 - [28] Bao N., Bouusso R., Jordan S., and Lackey B. *Phys. Rev. D*, 96:103512, 2017. [447](#)
 - [29] Barahona F. *Journal of Physics A: Math and General*, 15:3241, 1982. [361](#)
 - [30] Bardeen J., Cooper L., and Schrieffer J. *Phys. Rev.*, 108:1175, 1957. [419](#)
 - [31] Barnum H. and Knill E. *J. Math. Phys.*, 43:2097, 2002. [221](#)
 - [32] Barthélémy P., Rolland R., et Véron P. *Cryptographie : principes et mises en œuvre - 2e édition revue et augmentée*. Informatique. Hermès Sciences Publications, 2012. [230](#)
 - [33] Bass G., Tomlin C., Kumar V., Rihaczek P., and Dulny III J. *Quantum Science and Technology*, 3(2):024010, 2018. [447](#)
 - [34] Bauer M. and Bernard D. *Phys. Rev. A*, 84:044103, 2011. [404](#)
 - [35] Beigi S. and Shor P. W. 2007. ArXiv:0709.2090. [288](#)
 - [36] Bell J. S. *Physique Physique Fizika*, 1(3):195, 1964. [449](#), [451](#)
 - [37] Bell J. S. *Speakable and unspeakable in quantum mechanics. Collected papers on quantum philosophy*. 1987. [451](#)
 - [38] Bennett C. H. *IBM J. Res. Dev.*, 17(6):525–532, 1973. [331](#)
 - [39] Bennett C. H., Bernstein H. J., Popescu S., and Schumacher B. *Phys. Rev. A*, 53:2046–2052, 1996. [281](#)
 - [40] Bennett C. H., Bessette F., Brassard G., Salvail L., and Smolin J. *Journal of cryptology*, 5(1):3–28, 1992. [235](#)
 - [41] Bennett C. H. and Brassard G. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, vol. 175, pages 175–179. IEEE, 1984. [232](#)
 - [42] Bennett C. H., Brassard G., Popescu S., Schumacher B., Smolin J. A., et al. *Phys. Rev. Lett.*, 76:722–725, 1996. [268](#)
 - [43] Bennett C. H., Shor P. W., Smolin J. A., and Thapliyal A. V. *Physical Review Letters*, 83(15):3081, 1999. [290](#)
 - [44] Bera S., Baranger H. U., and Florens S. *Phys. Rev. A*, 93:033847, 2016. [446](#)

- [45] Bera S., Nazir A., Chin A. W., Baranger H. U., and Florens S. *Phys. Rev. B*, 90:075110, 2014. [446](#)
- [46] Bergeal N., Schackert F., Metcalfe M., Vijay R., Manucharyan V. E., et al. *Nature*, 465(7294):64–68, 2010. [438](#)
- [47] Bergeal N., Vijay R., Manucharyan V. E., Siddiqi I., Schoelkopf R. J., et al. *Nat Phys*, 6(4):296–302, 2010. [438](#)
- [48] Bernstein D. J. In *International Workshop on Post-Quantum Cryptography*, pages 73–80. Springer, 2010. [308](#)
- [49] Bernstein E. and Vazirani U. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. [331](#)
- [50] Besse J.-C., Gasparinetti S., Collodo M. C., Walter T., Kurpiers P., et al. *Phys. Rev. X*, 8:021003, 2018. [434](#)
- [51] Bhatia R. *Positive Definite Matrices*. Princeton University Press, 2007. [382](#)
- [52] Biamonte J. D. and Love P. J. *Phys. Rev. A*, 78:012352, 2008. [360](#)
- [53] Biham E., Huttner B., and Mor T. *Phys. Rev. A*, 54:2651–2658, 1996. [243](#)
- [54] Binder K. *Monte Carlo and Molecular Dynamics simulations in polymer science*. Oxford University Press (New York, Oxford), 1995. [356](#)
- [55] Birman J. *Braids, links and mapping class groups*. Princeton University Press, 1975. [31](#)
- [56] Biswas R., Jiang Z., Kechezhi K., Knysh S., Mandra S., et al. *Parallel Computing*, 64:81–98, 2017. [447](#)
- [57] Blais A., Huang R.-S., Wallraff A., Girvin S., and Schoelkopf R. *Phys. Rev. A*, 69:062320, 2004. [430](#)
- [58] Blei R. *Proc. Am. Math. Soc.*, 100:58, 1987. [463](#)
- [59] Blunden-Codd Z., Bera S., Bruognolo B., Linden N.-O., Chin A. W., et al. *Phys. Rev. B*, 95:085104, 2017. [446](#)
- [60] Bogdanov A., Khovratovich D., and Rechberger C. *Advances in cryptology—ASIACRYPT 2011*, pages 344–371, 2011. [308](#)
- [61] Bookatz A. *Quantum Information and Computation*, 14:6, 2014. [357](#)
- [62] Bouchiat V., Vion D., Joyez P., Estève D., and Devoret M. *Physica Scripta*, 1998(T76):165, 1998. [429](#)
- [63] Bouwmeester D., Ekert A., and Zeilinger A. *The Physics of Quantum Information*. Springer Berlin Heidelberg, 2000. [242](#)
- [64] Boyer M., Brassard G., Høyer P., and Tapp A. *Fortschritte der Physik*, 46(4-5):493–505, 1998. [303](#), [310](#)
- [65] Brassard G., Hoyer P., Mosca M., and Tapp A. *Contemporary Mathematics*, 305:53–74, 2002. [309](#), [310](#)
- [66] Brassard G., Lütkenhaus N., Mor T., and Sanders B. C. *Phys. Rev. Lett.*, 85:1330–1333, 2000. [235](#)
- [67] Braunstein S. L. and Pati A. K. *Phys. Rev. Lett.*, 98:080502, 2007. [215](#)
- [68] Bravyi S. B. and Kitaev A. Y. *Annals of Physics*, 298(1):210–226, 2002. [358](#)

- [69] Breuer H.-P., Laine E.-M., Piilo J., and Vacchini B. *Rev. Mod. Phys.*, 88:021002, 2016. [383](#), [392](#)
- [70] Breuer H.-P., Petruccione F., et al. *The theory of open quantum systems*. Oxford University Press, 2002. [380](#)
- [71] Brezger B., Hackermuller L., Uttenthaler S., Petschinka J., Arndt M., et al. *Phys. Rev. Lett.*, 88:100404, 2002. [14](#), [367](#)
- [72] Briegel H.-J., Dür W., Cirac J. I., and Zoller P. *Phys. Rev. Lett.*, 81:5932–5935, 1998. [281](#)
- [73] Brodutch A., Datta A., Modi K., Rivas A., and Rodríguez-Rosario C. A. *Phys. Rev. A*, 87:042301, 2013. [392](#)
- [74] de Broglie L. *Recherches sur la théorie des quantas*. PhD thesis, Université de Paris, 1924. [14](#)
- [75] Brune M., Bernu J., Guerlin C., Deléglise S., Sayrin C., et al. *Phys. Rev. Lett.*, 101:240402, 2008. [416](#)
- [76] Brune M., Hagley E., Dreyer J., Maître X., Maali A., et al. *Phys. Rev. Lett.*, 77:4887–4890, 1996. [416](#)
- [77] Buscemi F. *Phys. Rev. Lett.*, 113:140502, 2014. [392](#)
- [78] Cai N., Winter A., and Teung R. *Problem of Information Transmission*, 40:318, 2004. [289](#)
- [79] Calderbank A. R. and Shor P. W. *Physical Review A*, 54(2):1098–1105, 1996. [352](#)
- [80] Camino F. E., Zhou W., and Goldman V. J. *Phys. Rev. B*, 72:075342, 2005. [29](#)
- [81] Campagne-Ibarcq P., Flurin E., Roch N., Darson D., Morfin P., et al. *Phys. Rev. X*, 3:021008, 2013. [445](#)
- [82] Campagne-Ibarcq P., Six P., Bretheau L., Sarlette A., Mirrahimi M., et al. *Phys. Rev. X*, 6:011002, 2016. [445](#)
- [83] Carnot S. *Reflexions sur la puissance motrice du feu*. Bachelier (Paris), 1824. [111](#)
- [84] Castellanos-Beltran M. A., Irwin K. D., Hilton G. C., Vale L. R., and Lehnert K. W. *Nature Physics*, 4:929–931, 2008. [444](#)
- [85] Caves C. M. *Physical Review D*, 26(8):1817, 1982. [438](#)
- [86] Cerf N. J. and Adami C. *Phys. Rev. Lett.*, 79:5194–5197, 1997. [272](#)
- [87] Chakravarty S. and A. L. *Phys. Rev. Lett.*, 52:4–8, 1984. [446](#)
- [88] Chambers R. G. *Phys. Rev. Lett.*, 5:3–5, 1960. [26](#)
- [89] Chantasri A., Dressel J., and Jordan A. N. *Phys. Rev. A*, 88:042110, 2013. [445](#)
- [90] Chantasri A. and Jordan A. N. *Phys. Rev. A*, 92:032125, 2015. [445](#)
- [91] Chantasri A., Kimchi-Schwartz M. E., Roch N., Siddiqi I., and Jordan A. N. *Phys. Rev. X*, 6:041052, 2016. [445](#)
- [92] Chaturvedi S. and Shibata F. *Zeitschrift für Physik B Condensed Matter*, 35(3):297–308, 1979. [390](#)
- [93] Chen Y.-F., Hover D., Sendelbach S., Maurer L., Merkel S. T., et al. *Phys. Rev. Lett.*, 107:217401, 2011. [434](#)

- [94] Choi V. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7):E19–E20, 2011. 363
- [95] Chu J. *Biometrika*, 42:263, 1955. 223
- [96] Church A. *American Journal of Mathematics*, 58:345, 1935. 106, 111
- [97] Clader B. D., Jacobs B. C., and Sprouse C. R. *Phys. Rev. Lett.*, 110:250504, 2013. 336
- [98] Claudon J. *Oscillations cohérentes dans un circuit quantique supraconducteur : le SQUID dc*. PhD thesis, Université Joseph Fourier, 2005. 425
- [99] Clauser J., Horne M., Shimony A., and Holt R. *Phys. Rev. Lett.*, 23:880, 1969. 453
- [100] Clerk A., Girvin S., and Stone A. *Phys. Rev. B*, 67:165324, 2003. 202
- [101] Cleve R., Ekert A., Macchiavello C., and Mosca M. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998. 320, 321
- [102] Coester F. *Nuclear Physics*, 7:421, 1958. 359
- [103] Coester F. and Kümmel H. *Nuclear Physics*, 17:477, 1960. 359
- [104] Cohen-Tannoudji C. Physique atomique et moléculaire. <http://www.phys.ens.fr/~cct/college-de-france/>, 1973. 470
- [105] Cohen-Tannoudji C., Dupont-Roc J., et Grynberg G. *Introduction à l'électrodynamique quantique*. InterEditions/CNRS, 1988. 146, 156
- [106] Colless J. I., Ramasesh V. V., Dahlen D., Blok M. S., Kimchi-Schwartz M. E., et al. *Phys. Rev. X*, 8:011021, 2018. 359
- [107] Cook R. L., Martin P. J., and Geremia J. M. *Nature*, 446:774, 2007. 225
- [108] Cook S. A. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, 1971. 124
- [109] Cooley J. W. and Tukey J. W. *Mathematics of computation*, 19(90):297–301, 1965. 314
- [110] Cottet A. *Implementation of a quantum bit in a superconducting circuit*. PhD thesis, Université Pierre et Marie Curie, 2002. 428
- [111] Cronin A. D., Schmiedmayer J., and Pritchard D. E. *Rev. Mod. Phys.*, 81:1051–1129, 2009. 149
- [112] Csizsar I. and Körner J. *IEEE Trans. Inform. Theory*, 24:339, 1978. 289
- [113] Cubitt T. and Montanaro A. *Siam. J. Comp.*, 45:268, 2013. 360
- [114] Cummings F. *Phys. Rev.*, 140:A1051, 1965. 397
- [115] Curry H. B. *Proceedings of the National Academy of Sciences*, 20(11):584–590, 1934. 106
- [116] Curry H. B., Feys R., Craig W., and Craig W. *Combinatory logic*, vol. 1. North-Holland Publ., 1958. 106
- [117] Dalibard J. Atomes et rayonnement. <https://www.college-de-france.fr/site/jean-dalibard/>, 2012. 470
- [118] Danilin S., Lebedev A., Vepsäläinen A., Lesovik G., Blatter G., et al. *npj Quantum Information*, 4(1):29, 2018. 318
- [119] Davidovich L. *L'annuaire du Collège de France. Cours et travaux*,

- (116):677–678, 2018. [469](#)
- [120] Davis M. *Applied Mathematics and Computation*, 178:4, 2006. [107](#)
 - [121] Davisson C. *Bell System Tech. J.*, 7:90, 1928. [14](#)
 - [122] Dawson C. M. and Nielsen M. A. *arXiv preprint quant-ph/0505030*, 2005. [345](#)
 - [123] Delahaye J. *Pour la Science*, (375):88, 1009. [108](#)
 - [124] Delaye J.-P. *Pour la Science*, (295):98, 2002. [108](#)
 - [125] Deléglise S. *Reconstruction d'états non classiques du champ en électrodynamique en cavité*. PhD thesis, Université Pierre et Marie Curie, 2006. [408](#)
 - [126] Deléglise S., Dotsenko I., Sayrin C., Bernu J., Brune M., et al. *Nature*, 455:510, 2008. [408](#), [416](#)
 - [127] Denchev V. S., Boixo S., Isakov S. V., Ding N., Babbush R., et al. *Phys. Rev. X*, 6:031015, 2016. [447](#)
 - [128] Deutsch D. *Proceedings of the Royal Society (London)*, 400:97, 1985. [112](#), [113](#), [331](#)
 - [129] Devetak I. *IEEE Trans. Inform. Theory*, 51:44, 2005. [289](#)
 - [130] Devetak I. *Phys. Rev. Lett.*, 97:140503, 2006. [268](#)
 - [131] Devetak I., Harrow A. W., and Winter A. *Phys. Rev. Lett.*, 93:230504, 2004. [244](#), [284](#)
 - [132] Devetak I., Harrow A. W., and Winter A. *IEEE Trans. Inf. Th.*, 54:4587, 2008. [244](#), [247](#), [275](#)
 - [133] Devetak I. and Winter A. *IEEE Trans. Inf. Th.*, 50(3183), 2004. [244](#)
 - [134] Devoret M. and Roy A. *Comptes Rendus Physique*, 17:740, 2016. [438](#)
 - [135] Dickson N. G. *New Journal of Physics*, 13(7):073011, 2011. [363](#)
 - [136] Dickson N. G. and Amin M. H. S. *Phys. Rev. Lett.*, 106:050502, 2011. [363](#)
 - [137] Dieks D. *Physics Letters A*, 126:303, 1988. [219](#)
 - [138] Dirac P. *Lectures on quantum mechanics*. Dover Books on Physics. Dover Publications Inc., 2003. [146](#)
 - [139] Dolinar S. *Res. Lab. of Electronics MIT Quarterly Progress Rep.*, 111:115, 1973. [224](#), [225](#)
 - [140] Dolinar S., Moision B., and Erkmen B. *Fundamentals of free-space optical communications*. Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2012. [221](#), [226](#)
 - [141] Dumitrescu E. F., McCaskey A. J., Hagen G., Jansen G. R., Morris T. D., et al. *Phys. Rev. Lett.*, 120:210501, 2018. [359](#)
 - [142] Dupuis F., Hayden P., and Ke L. *IEEE Transactions on Information Theory*, 56:2946, 2010. [268](#), [284](#)
 - [143] Dutta P. and Horn P. *Rev. Mod. Phys.*, 53:497–516, 1981. [389](#)
 - [144] Eberhard P. H. *Physical Review A*, 47(2):R747, 1993. [457](#)
 - [145] Edwards S. *Proceedings of the Physical Society*, 85(4):613, 1965. [356](#)
 - [146] Einstein A. *Deutsche Physikalische Gesellschaft*, 18, 1916. [409](#)
 - [147] Einstein A., Podolski B., and Rosen N. *Phys. Rev.*, 47:777–780, 1935. [449](#), [450](#)

- [148] Ekert A. K. *Phys. Rev. Lett.*, 67:661–663, 1991. [236](#), [281](#)
- [149] Eldar Y., Megretski G., and Verghese G. *IEEE Trans. Inform. Theory*, 49:1017, 2003. [221](#)
- [150] Ettinger M., Høyer P., and Knill E. *Information Processing Letters*, 91(1):43–48, 2004. [326](#), [327](#)
- [151] Everett III H. *The Everett interpretation of quantum mechanics*. Princeton University Press, 2012. [3](#)
- [152] Fahri E., Goldstone J., Gutmann S., and Sipser M. Quantum computation by adiabatic evolution. ArXiv:quant-ph/0001106, 2000. [360](#), [362](#)
- [153] Feynman R. *Rev. of Mod. Phys.*, 20:367, 1948. [10](#), [20](#)
- [154] Feynman R. *La thèse de Feynman : une nouvelle approche de la mécanique quantique*. Pearson Education, 2007. [10](#), [15](#), [20](#)
- [155] Feynman R. and Hibbs A. *Quantum Mechanics and Path Integrals*. McGraw-Hill, 1965. [151](#)
- [156] Feynman R., Leighton R., and Sands M. *The Feynman Lectures on Physics*. Pearson-Addison Wesley, 2006. [11](#), [110](#), [135](#), [147](#), [150](#), [409](#)
- [157] Feynman R. and Vernon F. *Annals of Physics*, 24:118–173, 1963. [370](#)
- [158] Feynman R. P. *International journal of theoretical physics*, 21(6):467–488, 1982. [336](#), [355](#)
- [159] Ficheux Q., Jezouin S., Leghtas Z., and Huard B. *Nature Communications*, 9(1):1926, 2018. [445](#)
- [160] Filiol E. *Les virus informatiques : théorie, pratique et applications*. Springer, 2009. [101](#)
- [161] Fischer D. G., Kienle S. H., and Freyberger M. *Phys. Rev. A*, 61:032306, 2000. [225](#)
- [162] Fisher M., Weichman P., Grinstein G., and Fisher D. *Phys. Rev. B*, 40:546–570, 1989. [447](#)
- [163] Flory P. J. *Statistical Mechanics of Chain Molecules*. Hanser Gardner Pubns, 1989. [356](#)
- [164] Fowler A. G., Mariantoni M., Martinis J. M., and Cleland A. N. *Physical Review A*, 86(3), 2012. [352](#)
- [165] Fröhlich J. and Studer U. *Communications in mathematical physics*, 148(3):553–600, 1992. [31](#)
- [166] Galles D. Data structure visualization. <http://www.cs.usfca.edu/~galles/visualization/java/download.html>. [61](#)
- [167] Gallicchio J., Friedman A. S., and Kaiser D. I. *Phys. Rev. Lett.*, 112:110405, 2014. [459](#)
- [168] Gambetta J., Blais A., Boissonneault M., Houck A. A., Schuster D. I., et al. *Phys. Rev. A*, 77:012112, 2008. [441](#)
- [169] Gambetta J., Blais A., Schuster D., Wallraff A., Frunzio L., et al. *Phys. Rev. A*, 74:042318, 2006. [443](#)
- [170] Garey M. R. and Johnson D. S. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990. [124](#)

- [171] Gauss C. F. *Carl Friedrich Gauss Werke*, 3:265–327, 1866. 314
- [172] de Gennes P. *Scaling concepts in polymer physics*. Cornell University Press (Ithaca, N.Y.), 1979. 356
- [173] Georgescu I., Ashhab S., and Nori F. *Review of Modern Physics*, 86:153, 2014. 355
- [174] Gerlich S., Eibenberger S., Tomandl M., Nimmrichter S., Hornberger K., et al. *Nature Communications*, 2(1), 2011. 14
- [175] Gersch H. A. and Knollman G. C. *Phys. Rev.*, 129:959–967, 1963. 446
- [176] Gheeraert N., Zhang X. H. H., Sépulcre T., Bera S., Roch N., et al. *Phys. Rev. A*, 98:043816, 2018. 446
- [177] Giustina M., Mech A., Ramelow S., Wittmann B., Kofler J., et al. *Nature*, 497:227–230, 2013. 457
- [178] Gleyzes S., Kuhr S., Guerlin C., Bernu J., Deléglise S., et al. *Nature*, 446:297, 2007. 402, 416
- [179] Gödel K. Letter to J. von Neumann. 1956. 126
- [180] Gottesman D. *Physical Review A*, 54(3):1862–1868, 1996. 351
- [181] Gottesman D. *Stabilizer codes and quantum error correction*. PhD thesis, California Institute of Technology, 1997. 351
- [182] Government Communications Headquarters. Quantum key distribution. <https://www.ncsc.gov.uk/whitepaper/quantum-key-distribution>, 2016. 237
- [183] Greenberger D., Horne M., Shimony A., and Zeilinger A. *Am. J. Phys.*, 58:1131, 1990. 463
- [184] Griffiths R. B. *American Journal of Physics*, 79(9):954–965, 2011. 465
- [185] Griffiths R. B. *Foundations of Physics*, 41(4):705–733, 2011. 465
- [186] Grothendieck A. *Bol. Soc. Mat. Sao Paulo*, 8:1, 1953. 461, 462
- [187] Grover L. K. In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA, 1996. ACM. 303
- [188] Gu X., Frisk Kickum A., Miranowicz A., Liu Y.-X., and Nori F. *Physics Reports*, 718:1, 2017. 446
- [189] Guerlin C., Bernu J., Deléglise S., Sayrin C., Gleyzes S., et al. *Nature*, 448:889, 2007. 405, 406
- [190] Gyongyosi L., Imre S., and Nguyen H. *IEEE Communications Surveys and Tutorials*, 20:1149, 2018. 268, 286
- [191] Haroche S. Physique quantique. https://www.college-de-france.fr/site/serge-haroche/_course.htm, 2001. 470
- [192] Haroche S. *Physique quantique*. Leçons inaugurales du Collège de France. Collège de France, 2013. 1
- [193] Haroche S., Brune M., and Raimond J.-M. *Journal de Physique II*, 2(4):659–670, 1992. 405
- [194] Haroche S. and Raimond J.-M. *Exploring the quantum*. Oxford University Press, 2006. 135, 156, 380, 382, 396, 405
- [195] Harris R., Johansson J., Berkley A. J., Johnson M. W., Lanting T., et al.

- Phys. Rev. B*, 81:134510, 2010. [429](#)
- [196] Harris R., Lanting T., Berkley A. J., Johansson J., Johnson M. W., et al. *Phys. Rev. B*, 80:052506, 2009. [429](#)
- [197] Harrow A. W., Hassidim A., and Lloyd S. *Phys. Rev. Lett.*, 103:150502, 2009. [336](#)
- [198] Hastings M. *Nature Physics*, 5:255, 2009. [288](#)
- [199] Hayden P., Horodecki M., Winter A., and Yard J. *Open Syst. Inf. Dyn.*, 15:7, 2008. [289](#)
- [200] Helstrom C. *Information and Control*, 10:254, 1967. [220](#)
- [201] Hensen B., Bernien H., Dréau A. E., Reiserer A., Kalb N., et al. *Nature*, 526(7575):682, 2015. [457](#), [459](#)
- [202] Hirvensalo M. *Quantum Computing*. Natural computing series. Springer, 2004. [310](#), [313](#), [320](#), [325](#)
- [203] Hoffstein J., Pipher J., and Silverman J. H. In *International Algorithmic Number Theory Symposium*, pages 267–288. Springer, 1998. [328](#)
- [204] Holevo A. S. *Problems of Information Transmission*, 9:177, 1973. [252](#)
- [205] Holevo A. S. *IEEE Trans. Info. Theor.*, 44:269, 1998. [287](#)
- [206] Hornberger K., Uttenthaler S., Bzgerger B., Kackermuller L., Arndt M., et al. *Phys. Rev. Lett.*, 90:160401, 2003. [367](#), [368](#)
- [207] Horodecki M., Oppenheim J., and Winter A. *Nature*, 436:673 EP –, 2005. [268](#), [277](#), [278](#), [281](#)
- [208] Horodecki M., Oppenheim J., and Winter A. *Comm. Math. Phys.*, 269:107, 2007. [278](#)
- [209] Houck A. A., Koch J., Devoret M. H., Girvin S. M., and Schoelkopf R. J. *Quantum Information Processing*, 8(2):105–115, 2009. [429](#)
- [210] Howard W. A. *To HB Curry: essays on combinatory logic, lambda calculus and formalism*, 44:479–490, 1980. [106](#)
- [211] Hsieh M. and Wilde M. *IEEE Trans. Inform. Theory*, 56:4705, 2010. [249](#)
- [212] Hughston L. P., Jozsa R., and Wootters W. K. *Physics Letters A*, 183(1):14–18, 1993. [184](#)
- [213] Hwang W.-Y. *Phys. Rev. Lett.*, 91:057901, 2003. [235](#)
- [214] Impagliazzo R. In *Proceedings of the 10th Annual Structure in Complexity Theory Conference (SCT'95)*. IEEE Computer Society, 1995. [231](#)
- [215] Inomata K., Lin Z., Koshino K., Oliver W. D., Tsai J.-S., et al. *Nature Communications*, 7:12303 EP –, 2016. [434](#)
- [216] Itzykson C. and Zuber J. *Quantum Field Theory*. McGraw-Hill, 1980. [146](#)
- [217] Ivanovic I. D. *Physics Letters A*, 123(6):257–259, 1987. [219](#)
- [218] Jao D. and De Feo L. In *International Workshop on Post-Quantum Cryptography*, pages 19–34. Springer, 2011. [328](#)
- [219] Ji Y., Chung Y., Sprinzak D., Heiblum M., Mahalu D., et al. *Nature*, 422:415, 2003. [150](#)
- [220] Johnson M., Bunyk P., Maibaum F., Tolkacheva E., Berkley A., et al. *Superconductor Science and Technology*, 23(6):065004, 2010. [429](#)

- [221] Jönsson C. *Zeitschrift für Physik*, 161:454, 1961. [14](#)
- [222] Jordan P. and Wigner E. *Zeitschrift für Physik*, 47:631, 1928. [358](#)
- [223] Joyez P. Introduction to quantum circuits. Notes de cours Master Dispositifs Quantiques, 2016. [418](#)
- [224] Kadowaki T. and Nishimori H. *Phys. Rev. E*, 58:5355–5363, 1998. [447](#)
- [225] Kalb N., Reiserer A. A., Humphreys P. C., Bakermans J. J. W., Kamberling S. J., et al. *Science*, 356(6341):928–932, 2017. [244](#)
- [226] Kandala A., Mezzacapo A., Temme K., Takita M., Brink M., et al. *Nature*, 549:242 EP –, 2017. [359](#)
- [227] Kennedy R. *MIT Res. Lab. of Electronics Quarterly Progress Reports*, 108:219, 1973. [223](#)
- [228] Kitaev A. Y. *arXiv preprint quant-ph/9511026*, 1995. [317](#), [327](#)
- [229] Kitaev A. Y. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997. [339](#)
- [230] Kitaev A. Y. *Annals of Physics*, 303(1):2–30, 2003. [31](#), [352](#)
- [231] Kitaev A. Y., Shen A. H., and Vyalı M. N. *Classical and Quantum Computation (Graduate Studies in Mathematics)*. Amer Mathematical Society, 2002. [339](#), [342](#), [344](#), [345](#), [346](#), [357](#)
- [232] Kitaev A. Y. and Watrous J. In *Proc. 32nd ACM Symposium on Theory of Computing*, page 608. ACM New York, 2000. [357](#)
- [233] Kleene S. C. *Transactions of the American Mathematical Society*, 53(1):41–73, 1943. [111](#)
- [234] Knill E. *arXiv preprint quant-ph/9610012*, 1996. [357](#)
- [235] Knill E. and Laflamme R. *Phys. Rev. A*, 55:900–911, 1997. [351](#)
- [236] Koblitz N. *Math. Comp.*, 48:203–209, 1987. [328](#)
- [237] Koch J., Yu T. M., Gambetta J., Houck A., Schuster D., et al. *arXiv preprint cond-mat/0703002*, 2007. [429](#)
- [238] Koiran P., Nesme V., and Portier N. *Theoretical Computer Science*, 380(1):115–126, 2007. Automata, Languages and Programming. [312](#)
- [239] Korotkov A. In Devoret M., Huard B., Schoelkopf R., and Cugliandolo L., editors, *Quantum Machines: Measurement and control of engineered quantum systems*, vol. 96 of *Lectures Notes of the Les Houches Summer School*. Oxford University Press, 2011. [444](#)
- [240] Krivine J. Constantes de Grothendieck et fonctions de type positif sur les sphères. In *Séminaire d'analyse fonctionnelle (Polytechnique)*. École Polytechnique, 1977. [463](#)
- [241] Krupko Y., Nguyen V. D., Weiß T., Dumur E., Puertas J., et al. *Phys. Rev. B*, 98:094516, 2018. [446](#)
- [242] Kuhr S., Gleyzes S., Guerlin C., Bernu J., Hoff U., et al. *Applied Physics Letters*, 90:164101, 2007. [397](#), [405](#)
- [243] Ladner R. E. *Journal of the ACM (JACM)*, 22(1):155–171, 1975. [126](#)
- [244] Lagendijk A., van Tiggelen B., and Wiersma D. *Physics Today*, 62:24, 2009. [362](#)
- [245] Lalœ F. *Comprenons-nous vraiment la mécanique quantique?* Savoirs Actuels. EDP sciences, 2013. [449](#)

- [246] Landauer R. *IBM. J. Res. Dev.*, 53:183, 1961. 111, 212
- [247] Lanting T., Przybysz A. J., Smirnov A. Y., Spedalieri F. M., Amin M. H., et al. *Phys. Rev. X*, 4:021041, 2014. 447
- [248] Laskar J. *Solar System: Stability*, page 2198. 2000. 40
- [249] Laughlin R. *Phys. Rev. Lett.*, 50:1395–1398, 1983. 29
- [250] Laumann C., Moessner R., Scardicchio A., and Sondhi S. *The European Physical Journal Special Topics*, 224(1):75–88, 2015. 363
- [251] Le Bellac M. *Des phénomènes critiques aux champs de jauge*. Savoirs Actuels. EDP Sciences, 1998. 23, 356
- [252] Le Bellac M. *Physique quantique*. EDP Sciences, 2003. 359, 409
- [253] Leggett A., Chakravarty S., Dorsey A., Fisher M., Garg A., et al. *Rev. Mod. Phys.*, 59:1–85, 1987. 446
- [254] Leighton F. T. *Introduction to parallel algorithms and architectures: Arrays, trees, hypercubes*. Elsevier, 2014. 338
- [255] Levitin L. *Quantum communication and measurement*, page 439. New York Plenum, 1995. 198
- [256] Lévy L. *Magnétisme et supraconductivité*. InterEditions/CNRS. 419
- [257] Li X., Shen K., Huang M. C., and Chu L. A memory soft error measurement on production systems. <http://www2.ece.rochester.edu/~xinli/usenix07/>. 346
- [258] Liao S.-K., Cai W.-Q., Handsteiner J., Liu B., Yin J., et al. *Phys. Rev. Lett.*, 120:030501, 2018. 235
- [259] Lloyd S. *Phys. Rev. A*, 55:1613–1622, 1997. 289
- [260] Lo H.-K. and Popescu S. *Phys. Rev. Lett.*, 83:1459–1462, 1999. 281
- [261] London F. and London H. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 149(866):71–88, 1935. 420
- [262] Lutterbach L. G. and Davidovich L. *Phys. Rev. Lett.*, 78:2547–2550, 1997. 408
- [263] Lydersen L., Wiechers C., Wittmann C., Elser D., Skaar J., et al. *Nature Photonics*, 4(10):686–689, 2010. 236
- [264] Markov I. L. and Saeedi M. *Quantum Info. Comput.*, 12(5-6):361–394, 2012. 324
- [265] Martínez J. P., Léger S., Gheeraert N., Dassonneville R., Planat L., et al. *npj Quantum Information*, 5(1):19, 2019. 446
- [266] Massar S. and Popescu S. *Phys. Rev. Lett.*, 74:1259, 1995. 162
- [267] McClean J. R., Romero J., Babbush R., and Aspuru-Guzik A. *New Journal of Physics*, 18(2):023023, 2016. 358
- [268] McEliece R. J. *Deep Space Network Progress Report*, 44:114–116, 1978. 328
- [269] Meyer A. R. and Stockmeyer L. J. In *SWAT (FOCS)*, pages 125–129, 1972. 123
- [270] Mézard M. and Montanari A. *Information, Physics and Computation*. Oxford Graduate Texts. Oxford University Press, 2009. 361

- [271] Miller G. L. In *Proceedings of the seventh annual ACM symposium on Theory of computing*, pages 234–239. ACM, 1975. [130](#), [322](#)
- [272] Miller V. S. *Use of Elliptic Curves in Cryptography*, pages 417–426. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986. [328](#)
- [273] Montambaux G. et Akkermans E. *Physique mésoscopique des électrons et des photons*. Savoirs Actuels. EDP Sciences, 2004. [27](#)
- [274] Montanaro A. In *Proc. 2008 IEEE Information Theory workshop*, page 378. Institute of Electrical and Electronics Engineers (IEEE), 2008. [221](#)
- [275] Montanaro A. *Npj Quantum Information*, 2:15023, 2016. Review Article. [309](#)
- [276] Montanaro A. and Pallister S. *Physical Review A*, 93(3):032324, 2016. [336](#)
- [277] Moore G. and Read N. *Prog. Theor. Phys. Suppl.*, 107:157–166, 1992. [31](#)
- [278] Moser R. A. and Scheder D. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 245–252. ACM, 2011. [308](#)
- [279] Mostafavi M. *Cinétique et dynamique des réactions chimiques*. EDP Sciences, 2015. [356](#)
- [280] Motzoi F., Gambetta J. M., Reberntrost P., and Wilhelm F. K. *Phys. Rev. Lett.*, 103:110501, 2009. [429](#)
- [281] Murch K. W., Weber S. J., Macklin C., and Siddiqi I. *Nature*, 502:211 EP –, 2013. [444](#)
- [282] Naghiloo M., Tan D., Harrington P. M., Lewalle P., Jordan A. N., et al. *Phys. Rev. A*, 96:053807, 2017. [445](#)
- [283] National Institute of Standards and Technology. Post-quantum cryptography standardization. <https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization>, 2017. [329](#)
- [284] Neukart F., Compostella G., Seidel C., von Dollen D., Yarkoni S., et al. *Frontiers in ICT*, 4:29, 2017. [447](#)
- [285] von Neumann J. First draft of a report on the EDVAC. Technical report, University of Pennsylvania, 1945. [91](#)
- [286] von Neumann J. and Wigner E. *Physikalische Zeitschrift*, 30:465, 1929. [362](#)
- [287] Nielsen M. and Chuang I. *Quantum computation and quantum information*. Cambridge University Press, 2000. [304](#), [339](#), [344](#), [346](#), [469](#)
- [288] Nishimura H. and Ozawa M. *Theoretical Computer Science*, 276(1-2):147–181, 2002. [331](#)
- [289] Noguès G., Rauschenbeutel A., Osnaghi S., Brune M., Raimond J.-M., et al. *Nature*, 400:239, 1999. [405](#)
- [290] O’Malley P. J. J., Babbush R., Kivlichan I. D., Romero J., McClean J. R., et al. *Phys. Rev. X*, 6:031007, 2016. [359](#)
- [291] Orlando T., Mooij J., Tian L., van der Wal C., Levitov L., et al. *Phys. Rev. B*, 60:15398–15413, 1999. [429](#)

- [292] Osakabe N., Matsuda T., Kawasaki T., Endo J., Tonomura A., et al. *Phys. Rev. A*, 34:815–822, 1986. 26
- [293] Ozawa M. In *International Conference on Unconventional Methods of Computation*, pages 58–65. Springer, 2002. 331
- [294] Pan J.-W., Bouwmeester D., Weinfurter H., and Zeilinger A. *Phys. Rev. Lett.*, 80:3891–3894, 1998. 244
- [295] Peaudecerf B. *Mesure adaptative non destructive du nombre de photons dans une cavité*. PhD thesis, Université Pierre et Marie Curie, 2013. 406
- [296] Peaudecerf B., Rybarczyk T., Gerlich S., Gleyzes S., Raimond J. M., et al. *Phys. Rev. Lett.*, 112:080401, 2014. 406
- [297] Peres A. *Am. J. Phys.*, 46:746, 1978. 466
- [298] Peres A. *Physics Letters A*, 128:19, 1988. 219
- [299] Peruzzo A., McClean J., Shadbolt P., Yung M.-H., Zhou X.-Q., et al. *Nature Communications*, 5:4213 EP –, 2014. 358
- [300] Pollard J. M. *Mathematics of computation*, 25(114):365–374, 1971. 314
- [301] Popov L. and Faddeev L. *Phys. Lett. B.*, 25:29, 1967. 146
- [302] Post E. L. *Bulletin of the American Mathematical Society*, 50(5):284–316, 1944. 99
- [303] Pratt V. R. *SIAM Journal on Computing*, 4(3):214–220, 1975. 122
- [304] Preskill J. Lecture notes on quantum computation. <http://www.theory.caltech.edu/people/preskill/ph229/>, 1999. 346
- [305] Qiu D. *Phys. Rev. A*, 77:012328, 2008. 221
- [306] Rabin M. O. *Journal of number theory*, 12(1):128–138, 1980. 130, 322
- [307] Rains E. M. *IEEE Transactions on Information Theory*, 45(7):2361–2366, 1999. 351
- [308] Ramsey N. *Phys. Rev.*, 78:695, 1950. 150
- [309] Ray P., Chakrabarti B. K., and Chakrabarti A. *Phys. Rev. B*, 39:11828–11832, 1989. 447
- [310] Razborov A. A. In *Soviet Math. Dokl.*, vol. 31, pages 354–357, 1985. 338
- [311] Rempe J., Kitaev A. Y., and Regev O. *SIAM Journal on Computing*, 35:1070, 2006. 358
- [312] Richardson R. and Stone L. *Naval Research Logistics Quarterly*, 18:141, 1971. 74
- [313] de Riedmatten H., Marcikic I., van Houwelingen J. A. W., Tittel W., Zbinden H., et al. *Phys. Rev. A*, 71:050302, 2005. 244
- [314] Rivas Á., Huelga S. F., and Plenio M. B. *Reports on Progress in Physics*, 77(9):094001, 2014. 383
- [315] Roussel B. *Autopsie d’un courant électrique quantique*. PhD thesis, Université de Lyon, 2017. 340
- [316] Rowe M. A., Kielpinski D., Meyer V., Sackett C. A., Itano W. M., et al. *Nature*, 409(6822):791–794, 2001. 457
- [317] Saminadayar L., Glattli D., Jin Y., and Etienne B. *Phys. Rev. Lett.*, 79:2526–2529, 1997. 29

- [318] Santoro G. E., Martoňák R., Tosatti E., and Car R. *Science*, 295(5564):2427–2430, 2002. 447
- [319] Savage J. *Journal of the ACM (JACM)*, 19(4):660–674, 1972. 338
- [320] Sayrin C., Dotsenko I., Xingxing Z., Peaudecerf B., Rybarczyk T., et al. *Nature*, 477:73, 2011. 417
- [321] Schneier B. Quantum cryptography. https://www.schneier.com/blog/archives/2008/10/quantum_cryptog.html, 2008. 237
- [322] Schönhage A. and Strassen V. *Computing*, 7(3-4):281–292, 1971. 314, 324
- [323] Schroeder B., Pinheiro E., and Weber W.-D. In *ACM SIGMETRICS Performance Evaluation Review*, vol. 37, pages 193–204. ACM, 2009. 346
- [324] Schumacher B. *Phys. Rev. A*, 51:2738–2747, 1995. 256
- [325] Schumacher B. and Nielsen M. A. *Phys. Rev. A*, 54:2629–2635, 1996. 285
- [326] Schumacher B. and Westmoreland M. D. *Phys. Rev. A*, 56:131–138, 1997. 287
- [327] Schuster D., Houck A., Schreier J., Wallraff A., Gambetta J., et al. *Nature*, 445:515, 2007. 443
- [328] Seeley J., Richard M., and Love P. *The Journal of Chemical Physics*, 137:224109, 2012. 358
- [329] Shannon C. E. *Automata studies*, 34:157–165, 1956. 97
- [330] Shannon C. E. and Weaver W. *Théorie mathématique de la communication*. Retz - CEPL, 1975. 49, 51, 54, 63, 75, 80, 81
- [331] Sharvin D. and Sharvin Y. *JETP Lett.*, 34:272, 1981. 26
- [332] Shi Y. *arXiv preprint quant-ph/0205115*, 2002. 345
- [333] Shibata F., Takahashi Y., and Hashitsume N. *Journal of Statistical Physics*, 17(4):171–187, 1977. 390
- [334] Shimizu F., Shimizu K., and Takuma H. *Phys. Rev. A*, 46:R17 – R20, 1992. 14, 367
- [335] Shor P. W. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994. 311, 317, 322
- [336] Shor P. W. *Siam. J. Comp.*, 26:1484–1509, 1997. 322, 324, 327
- [337] Shor P. W. In *Lecture notes, MSRI Workshop on Quantum Computation*, 2002. 289
- [338] Shor P. W. *Communications in Mathematical Physics*, 246(3):453–472, 2004. 288
- [339] Siddiqi I., Vijay R., Pierre F., Wilson C., Frunzio L., et al. *Phys. Rev. Lett.*, 94:027005, 2005. 438
- [340] Simon D. R. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, SFCS '94*, pages 116–123, Washington, DC, USA, 1994. IEEE Computer Society. 311
- [341] Simon S. H. *The Oxford solid state basics*. Oxford Univ. Press, Oxford, UK, 2013. 417
- [342] Sin Q.-C., Jiang Y.-F., Mao Y.-L., You L.-X., Zhang W., et al. *Optica*, 4:1214, 2017. 244

- [343] Sisper M. *Introduction to the theory of computation*. Wadsworth Publishing Co Inc, 3rd edition, 2012. [87](#)
- [344] Snyman I. and Florens S. *Phys. Rev. B*, 92:085131, 2015. [446](#)
- [345] Solovay R. and Yao A. manuscript, 1996. [331](#)
- [346] Specker E. *Journal of Symbolic Logic*, 14(145), 1949. [102](#)
- [347] Sridharan V. and Liberty D. In *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference for*, pages 1–11. IEEE, 2012. [346](#)
- [348] Stannett M. *Applied Mathematics and Computation*, 178:8, 2006. [107](#)
- [349] Steane A. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 452(1954):2551–2577, 1996. [352](#)
- [350] Steudtner M. and Wehner S. *New Journal of Physics*, 20(6):063010, 2018. [358](#)
- [351] Stone L. *Theory of optimal search*. Mathematics in Science and Engineering (Book 118). Elsevier Science, 1975. [73](#)
- [352] Stone L., Royset J., and Washburn A. *Optimal Search for Moving Targets*. International Series in Operations Research & Management Science (Book 237). Springer, 2016. [73](#)
- [353] Sun Q.-C., Mao Y.-L., Jiang Y.-F., Zhao Q., Chen S.-J., et al. *Phys. Rev. A*, 95:032306, 2017. [244](#)
- [354] Takeoka M. and Sasaki M. *Phys. Rev. A*, 78:022320, 2008. [223](#)
- [355] Tanenbaum A. S. and Austin T. *Structured computer organization*. Pearson, 6th edition, 2012. [87](#)
- [356] Tao B. and Wu H. *Improving the Biclique Cryptanalysis of AES*, pages 39–56. Springer International Publishing, Cham, 2015. [308](#)
- [357] Taylor G. *Proc. Cam. Phil. Soc.*, 15:114, 1909. [14](#)
- [358] Tonomura A., Osakabe N., Matsuda T., Kawasaki T., Endo J., et al. *Phys. Rev. Lett.*, 56:792–795, 1986. [26](#)
- [359] Touchette H. *Physics Reports*, 478(1-3):1–69, 2009. [53](#)
- [360] Townsend P. D., Phoenix S. J. D., Blow K. J., and Barnett S. M. *Electronics Letters*, 30(22):1875–1877, 1994. [242](#)
- [361] Tsui D. C., Stormer H. L., and Gossard A. C. *Phys. Rev. Lett.*, 48:1559, 1982. [29](#)
- [362] Turing A. *The Journal of Symbolic Logic*, 2:153, 1937. [97](#), [106](#)
- [363] Turing A. *Proc. London Math. Soc.*, 42:230, 1937. [111](#)
- [364] Van Meter R. and Itoh K. M. *Phys. Rev. A*, 71:052320, 2005. [324](#)
- [365] Vassalo C. *Théorie des guides d’ondes électromagnétiques*. Eyrolles, 1985. [156](#)
- [366] Vazirani U. and Vidick T. *Phys. Rev. Lett.*, 113:140501, 2014. [236](#)
- [367] Vijay R., Macklin C., Slichter D. H., Weber S. J., Murch K. W., et al. *Nature*, 490:77 EP –, 2012. [444](#)
- [368] Wagner N. R. and Magyarik M. R. *A Public-Key Cryptosystem Based on the Word Problem*, pages 19–36. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985. [328](#)

- [369] van der Wal C. H., ter Haar A. C. J., Wilhelm F. K., Schouten R. N., Harmans C. J. P. M., et al. *Science*, 290:773, 2000. 429
- [370] Wallraff A., Schuster D., Blais A., Frunzio L., Huang R.-S., et al. *Nature*, 431:162, 2004. 430
- [371] Watrous J. Lecture notes. <https://cs.uwaterloo.ca/~watrous/LectureNotes.html>, 2006. Accessed: 2017-09-14. 313, 320
- [372] Webb R., Washburn S., Umbach C., and Laibowitz R. *Phys. Rev. Lett.*, 54:2696–2699, 1985. 26
- [373] Weber S. J., Chantasri A., Dressel J., Jordan A. N., Murch K. W., et al. *Nature*, 511:570 EP –, 2014. 445
- [374] Weber S. J., Samach G. O., Hover D., Gustavsson S., Kim D. K., et al. *Phys. Rev. Applied*, 8:014004, 2017. 447
- [375] Weihs G., Jennewein T., Simon C., Weinfurter H., and Zeilinger A. *Phys. Rev. Lett.*, 81:5039–5043, 1998. 457
- [376] Weiss U. *Quantum Dissipative Systems*. World Scientific, 1999. 389, 446
- [377] Wendin G. *Reports on Progress in Physics*, 80(10):106001, 2017. 446
- [378] Wigderson A. Knowledge, creativity, and P versus NP, 2006. 126
- [379] Wilde M. *Quantum information theory*. Cambridge University Press, 2nd edition, 2017. 244, 249, 257, 281, 286
- [380] Wiseman H. *Quantum measurement and control*. Cambridge University Press, 2014. 432
- [381] Wittmann C., Takeoka M., Cassemiro K. N., Sasaki M., Leuchs G., et al. *Phys. Rev. Lett.*, 101:210501, 2008. 224
- [382] Wolper P. *Introduction à la calculabilité : Cours et exercices corrigés*. Dunod, 2006. 87
- [383] Yao A. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 352–361. IEEE, 1993. 331
- [384] Yung M. H., Casanova J., Mezzacapo A., McClean J., Lamata L., et al. *Scientific Reports*, 4:3589 EP –, 2014. 358
- [385] Yurke B. *Phys. Rev. A*, 29:1419, 1984. 418
- [386] Zeilinger A. *Physica Scripta*, 92:072501, 2017. 244
- [387] Zhou X., Dotsenko I., Peaudecerf B., Rybarczyk T., Sayrin C., et al. *Physical review letters*, 108(24):243602, 2012. 417
- [388] Ziman J. *Principles of the theory of solids*. Cambridge University Press, 1972. 417
- [389] Zinn-Justin J. *Quantum Field Theory and critical phenomena*. Oxford Science Publications, 1992. 146
- [390] Zinn-Justin J. *Intégrale de chemin en mécanique quantique : introduction*. Savoirs Actuels. EDP Sciences, 2003. 23
- [391] Zurek W. H. *Rev. Mod. Phys.*, 75:715, 2003. 179
- [392] Zurek W. H. *Physical review letters*, 90(12):120404, 2003. 176
- [393] Zurek W. H. *Physical Review A*, 71(5):052105, 2005. 176
- [394] Zurek W. H., Habib S., and Paz J. *Phys. Rev. Lett.*, 70:1187–1190, 1993. 413

Index

- accélération quantique, 294, 295, 311, 330
- action, 15–17, 39
 - principe de moindre action, 16, 22
- algorithme
 - amplification d’amplitude, 309
 - d’estimation de phase quantique, 317–321
 - de Deutsch-Jozsa, 295–301
 - de Grover, 302–311
 - de Shor, 321–326
 - de Simon, 311–313
 - de Solovay-Kitaev, 345–346
- alphabet, 86, 92
- amplitude de probabilité, 11–15, 18, 22, 31–38, 41, 43
- anyon, 29
- approche fréquentiste, 11
- approximation
 - de Born, 370
 - de Born-Markov, 388
 - markovienne, 388, 402, 443
- base de Bell, 238
- bit, 52, 53
- borne
 - de Bell, 462
 - de Hamming quantique, 351
 - de Helstrom, 220, 223–225
 - de Holevo, 252–253
 - de Massard-Popescu, 162, 214, 225
 - de Rains, 351
 - de Singleton quantique, 351
 - de Slepian-Wolf classique, 83
 - de Slepian-Wolf quantique, 277
- boson, 28
- calcul quantique adiabatique, 360
- calculabilité, 85–87, 97
 - des nombres réels, 101
- canal
 - classique, 78
 - quantique, 229
- capacité
 - assistée par intrication, 290
 - d’un canal classique, 51, 80, 81
 - d’un canal quantique, 285–286, 289
- circuit
 - classique, 87–90, 121, 337–338
 - P-uniforme, 338
 - quantique, 237, 294, 298–301, 303–304, 336–355
- classes de complexité, 117–118
 - BPP, 128–133, 332–333
 - BQP, 331–336, 345
 - NP, 118–127, 308–309, 334
 - NP-complet, 124, 308–309, 334
 - NP-dur, 123–124
 - P, 118, 124–127, 132–133, 338
 - PH, 123
 - PP, 127–128, 332–333
 - QMA, 357
 - QMA-complet, 360
- codage, 58
 - de Huffman, 60
 - de Vernam, 231, 236, 263
 - régulier, 58
- code, 58
- code correcteur d’erreurs, 294, 346–355
 - bit flip*, 347–348
 - phase flip*, 348
 - code topologique, 352
 - de Shor, 348–349
 - de Steane, 352, 354
 - de surface, 352, 354
 - stabilisateur, 351–352
- cohérence, 173
- complexité, 85, 114, 116, 117

- arithmétique, 116
- en circuit quantique, 294, 295
- en espace, 114
- en requêtes, 294, 295, 333
- en temps, 114, 115, 295, 333
- constante
 - de Grothendieck, 462
 - de Planck, 16
- contradiction
 - GHZ, 464
- cryptographie, 230, 328–329
 - asymétrique, 231
 - symétrique, 231
- cryptosystème
 - à courbes elliptiques, 328
 - post-quantique, 328–329
 - RSA, 328
- décohérence, 22, 47, 200, 365–367, 370, 413–417
- décomposition de Schmidt, 176, 178, 188, 216
- déplacement Lamb, 399
- détecteur
 - de Dolinar, 224
 - de Kennedy, 223
- détection homodyne, 223, 224
- distribution d'intrication, 237–238
- divergence de Kullback-Leibler, 392
 - classique, 67–68, 76, 77
 - quantique, 255, 261
- échappatoire, 456–459
 - de la communication, 457
 - du libre arbitre, 457
 - efficacité de détection, 456
- effet
 - Aharonov-Bohm, 25, 150, 420
 - Hall quantique fractionnaire, 29
 - Josephson alternatif, 424
 - Josephson continu, 425
 - Meissner, 421
 - Stark dynamique, 399
- électrodynamique
 - en cavité, 395
 - sur circuit, 429
- élément de réalité, 450
- empreinte quantique, 369
- entropie
 - conditionnelle, 75, 263
 - conditionnelle quantique, 263
 - de Boltzmann, 55
 - de Shannon, 51, 54, 57, 177, 187, 251
 - de von Neumann, 188, 251, 259–261
- envariance, 179–181
- équation
 - d'Euler-Lagrange, 17
 - de Hamilton-Jacobi, 40
 - de Lindblad, 377–378, 390, 391, 410
 - de Mathieu, 427
 - de Schrödinger, 39, 153
 - de Schrödinger stochastique, 378–380, 435, 437
 - markovienne, 375–380
- espace
 - de Hilbert, 36
 - des états, 34–36
 - des configurations, 16, 40
 - des phases, 40
- estimateur, 160, 163
 - qualité de, 160
- état
 - cohérent, 154, 157–158, 411
 - factorisé, 170, 175
 - GHZ, 464
 - intriqué, 170, 171, 176
 - joint, 435
 - pointeur, 413
 - quantique, 35, 47
 - relatif, 171, 177, 189, 240, 372–375, 399, 401
- facteur topologique, 16, 27, 28, 30, 31
- fermion, 28
- fonction
 - de Hamilton, 22, 39
 - de Wigner, 407
- fonctionnelle d'influence, 370
- grande déviation, 164
- groupe fondamental, 30
- groupeïde fondamental, 30
- hamiltonien, 39, 40, 152, 419
 - de Jaynes-Cummings, 397–399, 433
- hétérodynage, 438
- hyper-calcul, 106–110

- impulsion, 40
- incomplétude, 450, 451
- indépendance, 64
- inégalité
 - d'Heisenberg, 45
 - de Bell, 451, 461
 - de CHSH, 453
- inférence bayésienne, 70, 76–78
- information
 - accessible, 250–252
 - de préparation, 250, 251
 - manquante, 250, 251
 - mutuelle, 73, 76, 81, 196, 252, 443
 - mutuelle quantique, 254
- intégrale de chemins, 14, 20
- interférences, 14
- interférométrie, 367
 - de Mach-Zehnder, 149
 - de Ramsey, 149, 397, 400
- intrication, 170, 450
- jonction Josephson, 422
- lagrangien, 16, 20, 23, 418
- λ -calcul, 106
- langage, 87, 92, 93
 - décidable, 93, 97
 - indécidable, 98–101
 - reconnaissable, 93
- limite
 - classique, 22
 - quantique, 198
- limite quantique, 202
- localité, 15, 450, 465
- longueur de Planck, 109
- machine de Turing, 91–97, 111, 112
 - déterministe, 113, 337
 - non déterministe, 103–105, 113, 118
 - probabiliste, 127–133
 - quantique, 293, 295, 330–335, 337
 - universelle, 95–97, 99, 100, 102
 - variantes, 102, 103
- markoviannité, 375–377, 383, 387–389, 392
- matrice de Pauli, 136
- mélange statistique d'états purs, 173, 183
- mesure, 42
 - destructive, 44
 - faible, 193–194
 - généralisée, 189–193, 204, 219, 399, 402
 - non destructive, 44
 - postulat de, 42, 47
 - projective, 43, 189, 193, 399
- moment conjugué, 40
- monogamie de l'intrication, 273
- mot, 86, 92, 93
- multiplexeur, 65
- observable, 44
- opérateur
 - densité réduit, 174
- opérateur
 - d'évolution, 37
 - d'annihilation, 152
 - déplacement, 154
 - de création, 152
 - de Lindblad, 378
 - densité, 173
 - densité réduit, 174, 175, 389
 - unitaire, 38, 210
- oracle, 122, 294–297, 299–305, 307, 308, 312, 333
- oscillateur harmonique, 135, 151–158, 396, 419
- oscillation de Rabi, 19
- paire de Cooper, 420
- particule
 - indiscernable, 27
 - libre, 20
- photodétection, 192
- photon, 11, 14, 33, 44, 155–158
 - fuite de, 409
- population, 173
- porte logique
 - classique, 87–90
 - quantique, 294, 295, 338–339
 - tolérante aux erreurs, 352–355
- positivité complète, 381, 391
- POVM, 190
- principe
 - d'indétermination, 21, 45, 155
 - de superposition, 14
 - variationnel, 16
- probabilité, 10, 11

- causale, 69
- conditionnelle, 12, 74
- de transition, 13, 69
- inductive, 69
- problème, 87
 - 3SAT, 121, 308
 - calculable, *voir* langage décidable
 - de décision, 87, 97
 - de l'arrêt, 98–99, 102, 106–110
 - de l'isomorphisme de graphes, 329
 - de la factorisation, 121, 122, 126, 322
 - du sac-à-dos, 119, 120, 124
 - du sous-groupe caché, 326–328
 - du voyageur de commerce, 120, 124
 - indécidable, 98–101
 - instance, 87
 - SAT, 121
 - satisfaisabilité d'un circuit, 121, 124, 308
- produit tensoriel, 170
- protocole
 - BB84, 232–235, 246
 - d'échange d'intrication, 242–244
 - de codage superdense, 238–239, 246, 281
 - de codage téléportation quantique, 281
 - de distillation de paires maximale-ment intriquées, 279
 - de distribution d'intrication, 246
 - de fusion d'états, 277
 - de renversement du temps, 282
 - de Slepian-Wolf complètement quantique, 279
 - de téléportation quantique, 240–241, 246
 - de transfert d'état, 274–277, 282
 - E91, 236
 - père, 284–285
- purification, 184
- quantification canonique, 41, 151, 418
- quantum de flux, 422
- qubit, 18, 135–150
 - de charge, 427
 - de flux, 429
 - supraconducteur, 426
- raisonnement EPR, 450–451
- réalisme, 450
- réduction, 123–124
- régime
 - dispersif, 398
 - résonant, 398
- règle de Born, 13, 14, 47, 171
- ressources, 244–247
- saut quantique, 411
- simulation quantique, 336, 355
- solveur variationnel quantique, 358
- sous-additivité, 65
- sphère de Bloch, 137–139, 158, 400
- SQUID, 426
- statistique
 - anyonique, 29–30
 - bosonique, 28
 - de Bose-Einstein, 28
 - de Fermi-Dirac, 28
 - fermion, 28
- super-opérateur, 381, 382, 390–392
- symétrie, 27
- système
 - à deux niveaux, 18, 41, 135–150, 396
 - composé, 168
 - ouvert, 365
- théorème
 - d'inconvertibilité, 214
 - d'indiscernabilité, 204, 208, 217
 - de Bayes, 69, 218
 - de codage de Shannon, 56, 59, 81, 83
 - de Kraus, 382
 - de non-clonage, 205–210, 214, 273, 351
 - de non-communication, 214, 245
 - de non-dissimulation, 215
 - de non-effacement, 210–212, 216
 - de non-téléportation, 245
 - de Schumacher, 256–257, 277
 - de Tsirelson, 463
 - du seuil, 354–355
 - GHJW, 184, 187, 216
 - HSW, 287
- thèse
 - de Church-Turing, 111

- de Church-Turing étendue, 133–134
- de Church-Turing étendue quantique, 334–335
- de Church-Turing-Deutsch, 112–113
- de Landauer, 111–112
- tomographie, 158, 165, 208, 406
- trajectoire quantique, 365, 370, 383, 391, 399, 402, 409–410
- transfert d'état quantique, 275
- transformée de Fourier quantique, 313–317
- transformation de jauge, 24
- transmon, 429
- tresse, 28
- trous d'Young, 366

- variables incompatibles, 45
- vecteur de Bloch, 137–139, 158