Caractérisation pratique des systèmes quantiques et mémoires quantiques auto-correctrices 2D

par

Olivier Landon-Cardinal

Thèse présentée au département de physique

en vue de l'obtention du grade de docteur ès sciences (Ph.D.)

FACULTÉ des SCIENCES
UNIVERSITÉ de SHERBROOKE

Sherbrooke, Québec, Canada, 9 juillet 2013

Your file Votre référence
ISBN: 978-0-494-95092-0

Our file Notre référence
ISBN: 978-0-494-95092-0

Le 11 juillet 2013

le jury a accepté la thèse de Monsieur Olivier Landon-Cardinal dans sa version finale.

Membres du jury

Professeur David Poulin
Directeur de recherche
Département de physique

Professeur Michel Pioro-Ladrière
Membre
Département de physique

Professeur Frank Verstraete
Membre externe
Université de Vienne

Professeur David Sénéchal
Président rapporteur
Département de physique

À tous ceux qui ont toujours cru en moi... en particulier ma famille.

Sommaire

Cette thèse s'attaque à deux problèmes majeurs de l'information quantique :

- Comment caractériser efficacement un système quantique ?
- Comment stocker de l'information quantique ?

Elle se divise donc en deux parties distinctes reliées par des éléments techniques communs. Chacune est toutefois d'un intérêt propre et se suffit à elle-même.

Caractérisation pratique des systèmes quantiques Le calcul quantique exige un très grand contrôle des systèmes quantiques composés de plusieurs particules, par exemple des atomes confinés dans un piège électromagnétique ou des électrons dans un dispositif semi-conducteur. Caractériser un tel système quantique consiste à obtenir de l'information sur l'état grâce à des mesures expérimentales. Or, chaque mesure sur le système quantique le perturbe et doit donc être effectuée après avoir repréparé le système de façon identique. L'information recherchée est ensuite reconstruite numériquement à partir de l'ensemble des données expérimentales.

Les expériences effectuées jusqu'à présent visaient à reconstruire l'état quantique complet du système, en particulier pour démontrer la capacité de préparer des états intriqués, dans lesquels les particules présentent des corrélations non-locales. Or, la procédure de *tomographie* utilisée actuellement n'est envisageable que pour des systèmes composés d'un petit nombre de particules. Il est donc urgent de trouver des méthodes de caractérisation pour les systèmes de grande taille.

Dans cette thèse, nous proposons deux approches théoriques plus ciblées afin de caractériser un système quantique en n'utilisant qu'un effort expérimental et numérique raisonnable.

- La première consiste à estimer la distance entre l'état réalisé en laboratoire et l'état cible que l'expérimentateur voulait préparer. Nous présentons un protocole, dit de *certification*, demandant moins de ressources que la tomographie et très efficace pour plusieurs classes d'états importantes pour l'informatique quantique.

- La seconde approche, dite de *tomographie variationnelle,* propose de reconstruire l'état en restreignant l'espace de recherche à une classe variationnelle plutôt qu'à l'immense espace des états possibles. Un état variationnel étant décrit par un petit nombre de paramètres, un petit nombre d'expériences peut suffire à identifier les paramètres variationnels de l'état expérimental. Nous montrons que c'est le cas pour deux classes variationnelles très utilisées, les états à produits matriciels (MPS) et l'ansatz pour intrication multi-échelle (MERA)..

Mémoires quantiques auto-correctrices 2D Une mémoire quantique auto-correctrice est un système physique préservant de l'information quantique durant une durée de temps macroscopique. Il serait donc l'équivalent quantique d'un disque dur ou d'une mémoire flash équipant les ordinateurs actuels. Disposer d'un tel dispositif serait d'un grand intérêt pour l'informatique quantique.

Une mémoire quantique auto-correctrice est initialisée en préparant un état fondamental, c'est-à-dire un état stationnaire de plus basse énergie. Afin de stocker de l'information quantique, il faut plusieurs états fondamentaux distincts, chacun correspondant à une valeur différente de la mémoire. Plus précisément, l'espace fondamental doit être dégénéré. Dans cette thèse, on s'intéresse à des systèmes de particules disposées sur un réseau bidimensionnel (2D), telles les pièces sur un échiquier, qui sont plus faciles à réaliser que les systèmes 3D.

Nous identifions deux critères pour l'auto-correction :

- La mémoire quantique doit être *stable* face aux perturbations provenant de l'environnement, par exemple l'application d'un champ magnétique externe. Ceci nous amène à considérer les systèmes topologiques 2D dont les degrés de liberté sont intrinsèquement robustes aux perturbations locales de l'environnement.
- La mémoire quantique doit être *robuste* face à un environnement thermique. Il faut s'assurer que les excitations thermiques n'amènent pas deux états fondamentaux distincts vers le même état excité, sinon l'information aura été perdue.

Notre résultat principal montre qu'aucun système topologique 2D n'est auto-correcteur : l'environnement peut changer l'état fondamental en déplaçant aléatoirement de petits paquets d'énergie, un mécanisme cohérent avec l'intuition que tout système topologique admet des excitations localisées ou *quasiparticules.* L'intérêt de ce résultat est double. D'une part, il oriente la recherche d'un système auto-correcteur en montrant qu'il doit soit (i) être tridimensionnel, ce qui est difficile à réaliser expérimentalement, soit (ii) être basé sur des mécanismes de protection nouveaux, allant au-delà des considérations énergétiques. D'autre part, ce résultat constitue un premier pas vers la démonstration formelle de l'existence de quasiparticules pour tout système topologique.

Remerciements

Je tiens à remercier David Poulin, mon directeur de thèse, qui m'a guidé dans mon travail. À l'issue de ce doctorat, je suis grâce à lui un bien meilleur chercheur. Au-delà de ses qualités académiques, je suis profondément reconnaissant pour ses qualités humaines. Un doctorat va au-delà du parcours académique : c'est une expérience de vie où j'ai appris beaucoup sur moi-même et David a agi comme mentor dans ce cheminement.

Merci aux membres de mon comité de thèse, les Pr. Sénéchal et Pioro-Ladrière, qui m'ont accompagné lors de mon doctorat et dont les remarques judicieuses sur ma thèse m'ont été très précieuses. Merci au Pr. Verstraete de m'avoir fait l'honneur d'être sur mon jury, pour ses questions lors de ma soutenance de thèse. Les discussions scientifiques lors de la soutenance ont été une très belle façon de conclure mon doctorat.

Ce doctorat est une grande étape dans mon parcours académique. J'en profite donc pour remercier tous les professeurs, collègues et chercheurs dont j'ai croisé la route et qui m'ont amené à choisir le domaine dans lequel j'évolue aujourd'hui. Un merci particulier à ceux qui sont allés au-delà de leur devoir pour me donner un coup de main.

Merci à mes collègues de l'équipe Poulin pour l'ambiance de travail extraordinaire. Un merci particulier à Guillaume pour toutes les conversations scientifiques et pour avoir été un compagnon de voyage lors de multiples conférences. Merci aussi aux collègues de l'EPIQ qui m'ont permis d'en apprendre plus sur d'autres facettes de l'informatique quantique et aussi de passer de très bons moments en leur compagnie. Un merci particulier à Maxime de m'avoir pris sous son aile lors de mon arrivée à Sherbrooke. Plus généralement, le département de physique de Sherbrooke aura été un endroit vivant où il faisait bon travailler. Merci aux gens du RECSUS pour tous ces midis passés à refaire le monde !

Mon travail académique est possible grâce à l'équilibre que j'ai trouvé dans ma vie personnelle, en particulier auprès de mes amis. Merci à mes meilleurs amis, Antoine, Ian, Jean-Sébastien et Pierre pour leur amitié inconditionnelle, et d'avoir partagé mes joies et mes peines depuis si longtemps. Merci à mon grand ami Raphaël de m'avoir fait confiance en me demandant d'être le parrain de son fils. Merci à mes filleuls – Victor, Jade et Ludovic – pour leur joie de vivre et à leurs parents de me donner l'occasion de partager cette joie. Merci à Benoît et Simon, qui sont toujours aussi proches, malgré la distance entre la France et le Québec. Un énorme merci à Jean-Philippe, de m'avoir épaulé au travers de ce doctorat, en particulier grâce à nos conversations autour d'une bière le mardi soir, et d'être devenu un véritable ami. Bon courage à Sabrina pour son doctorat et merci d'avoir été là, en particulier à Paris, depuis qu'on a trois ans. Merci à tous mes amis : vous avez tous une place au chaud dans mon coeur et votre amitié compte énormément à mes yeux.

Merci à Laureen, pour les beaux moments passés ensemble et de m'avoir donné confiance au début de mon doctorat. Merci à celles qui m'ont redonné confiance depuis...

Merci à mes chats, qui ont été une source fiable de réconfort dans mes moments de solitude. Merci à Louise de m'avoir aidé à mieux me comprendre. Merci à la famille Landon, en particulier Grand-Papa qui a permis à un petit Québécois de 18 ans de partir étudier à Paris et Marie-Françoise pour sa tendresse à mon égard. Merci à la famille Cardinal, en particulier Mamie Ginette, pour ses remarques parfois acérées, mais toujours affectueuses. Merci à Sylvie pour son écoute, en particulier dans les moments difficiles.

Trois personnes ont toujours été présentes pour moi et leur soutien indéfectible m'a donné le courage de faire face à toutes les épreuves. Merci à ma mère pour son amour maternel, sa présence et son soutien inconditionnel. Merci à mon père, pour ses conseils judicieux de professeur et d'avoir été ma lumière dans les moments les plus sombres. Merci enfin à ma petite soeur, Océane, de m'avoir soutenu, parfois à bouts de bras. Elle est la seule qui ose m'affronter lorsque j'ai besoin d'être ramené sur terre et je lui en suis éternellement reconnaissant. Ma thèse vous est dédiée car vous êtes le roc sur lequel je me suis appuyé pour l'écrire.

Table des matières

Table des figures

Notations

Notations simplifiées $|\psi\rangle$ État pur

$\psi \equiv |\psi\rangle\langle\psi|$ Matrice densité correspondant à l'état pur $|\psi\rangle$

États particuliers $|\pm\rangle = \frac{1}{\sqrt{2}}|0\rangle \pm \frac{1}{\sqrt{2}}|1\rangle$

$|\pm i\rangle = \frac{1}{\sqrt{2}}|0\rangle \pm i\frac{1}{\sqrt{2}}|1\rangle$

$|GHZ_n\rangle = \frac{1}{\sqrt{2}}|0\rangle^{\otimes n} + \frac{1}{\sqrt{2}}|1\rangle^{\otimes n}$

$|W_n\rangle = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}|0\rangle^{\otimes k-1}|1\rangle \otimes |0\rangle^{\otimes n-k}$

Notations ensemblistes $[\![a, b]\!] = [a, b] \cap \mathbb{N}$

$\langle A_1, A_2, \ldots A_n\rangle$ groupe ou algèbre généré par les opérateurs $\{A_i\}_{i=1}^{n}$

Notations probabilistes $\mathbb{E}[X]$ Espérance d'une variable aléatoire X

$\mathbb{V}[X]$ Variance d'une variable aléatoire X

Notations informatiques poly(n) Ensembles des fonctions polynomiales de n

$\mathcal{O}(f(n))$ Ensemble des fonctions dominées asympotiquement par $f(n)$

Chapitre 1

Introduction

1.1 Objectifs de la thèse

L'objectif de cette thèse est de démontrer l'apport de mes travaux de doctorat au domaine de l'informatique quantique. Elle s'articule autour des quatre articles suivants :

1. « Practical characterization of quantum devices without tomography », Marcus P. da Silva, Olivier Landon-Cardinal, David Poulin. *Physical Review Letters,* **107**, 210404 (2011) [1]
2. « Efficient quantum state tomography », M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y-K. Liu, D. Poulin *Nature Communications*, **1** (9) 149, (2010) [2]
3. « Practical learning method for multi-scale entangled states », Olivier Landon-Cardinal et David Poulin. *New Journal of Physics,* **14,** 085004 (2012) [3]
4. « Local topological order inhibits thermal stability in 2D », Olivier Landon-Cardinal et David Poulin. *Physical Review Letters,* **110**, 090502 (2013) [4]

Dans cette thèse, chaque article est replacé dans le contexte de la littérature afin de mesurer sa contribution. Ainsi, des références à la littérature scientifique sont faites au fil du texte, au moment où elles sont pertinentes. De plus, j'ai rédigé cette thèse afin de fournir les outils nécessaires à la compréhension de ces articles. Les notions essentielles sont rappelées avant chaque article afin

de préparer le lecteur et de lui permettre de bien comprendre l'enjeu et le résultat de l'article. Certaines notions techniques qui permettent une compréhension plus profonde sont expliquées en annexe et le texte principal y fait référence. L'annexe E comprend aussi un article [5] dont je suis co-auteur et rédigé lors de mon doctorat mais dont le sujet est distinct du reste de cette thèse.

Cette thèse est donc un document cohérent qui se suffit à lui-même. Toutefois, elle suppose une certaine familiarité avec les concepts de base de l'information quantique, un domaine pour lequel plusieurs ouvrages de référence existent [6, 7, 8]. En fait, les idées mises de l'avant par l'information quantique commencent à faire leur chemin même dans les manuels d'introduction à la physique quantique [9]. Pour une introduction à l'information quantique pour le lecteur francophone, je suggère les premiers chapitres de [10]. Toutes ces références concernent le côté théorique du domaine alors que d'énormes progrès ont été réalisés expérimentalement. En particulier, plusieurs technologies sont candidates à la mise au point d'un futur ordinateur quantique. Pour un panorama de ces différentes technologies, on pourra lire avec intérêt l'article de revue [11].

1.2 Problématique et structure de la thèse

Cette thèse s'attaque à deux problèmes majeurs de l'information quantique :

1. Comment caractériser efficacement un système quantique ?
2. Comment stocker de l'information quantique ?

La structure de la thèse reflète ces deux grandes questions : elle se sépare en deux grandes parties qui sont indépendantes. Les chapitres 2, 3 et 4 forment la première partie qui propose des techniques pratiques de caractérisation des systèmes quantiques. Les chapitres 5, 6 et 7 forment la seconde partie qui s'intéresse aux mémoires auto-correctrices et démontre que les systèmes topologiques 2D n'en sont pas.

Le chapitres 2 sert d'introduction aux chapitres 3 et 4 qui s'articulent autour des articles [1], [2] et [3]. De même, les chapitres 6 et 7 préparent le terrain pour les résultats de l'article [4] qui seront l'aboutissement du chapitre 7. Ma contribution personnelle est détaillée avant chaque article.

1.2.1 Structure détaillée

Le chapitre 2 s'intéresse au problème général d'extraire de l'information au sujet d'un système quantique à partir de mesures expérimentales. En particulier, je montrerai que l'approche actuelle qui consiste à reconstruire la description complète d'un système n'est plus envisageable dès que le systèmes comporte plus d'une dizaine de particules. Nous proposons alors de se concentrer sur des tâches plus ciblées qui visent à extraire une information partielle, mais particulièrement intéressante. Ce chapitre sert donc d'introduction aux deux chapitres suivants qui s'articulent autour des articles [1], [2] et [3].

Le chapitre 3 propose d'estimer directement la distance entre un état théorique et un état expérimental. Cette procédure de *certification* est particulièrement intéressante lorsque l'expérimentateur cherche à préparer un état cible. Notre article [1] propose un protocole qui estime la fidélité entre l'état cible et l'état expérimental en utilisant une approche Monte Carlo où un échantillonnage judicieux détermine les observables à mesurer expérimentalement. Ce protocole est efficace pour plusieurs classes d'états particulièrement importantes en informatique quantique.

Le chapitre 4 propose la notion inédite de *tomographie variationnelle*. Ainsi, plutôt que de reconstituer la matrice densité, qui est la description la plus générale en mécanique quantique, cette approche suggère de trouver l'état variationnel qui correspond le mieux à l'état expérimental. L'espoir est qu'un petit nombre de mesures expérimentales suffit à reconstituer le petit nombre de paramètres qui caractérisent un état variationnel. Nous démontrons que cette approche est efficace pour deux classes variationnelles : les matrix product states ou MPS qui décrivent bien les systèmes 1D non-critiques et les états MERA (multi-scale renormalization ansatz) qui décrivent les systèmes critiques.

Le chapitre 5 introduit la notion d'ordre topologique, un type d'ordre qui n'est pas caractérisé par une brisure de symétrie, ni par un paramètre d'ordre local. En particulier, nous présenterons en grand détail le modèle canonique de système topologique : le code torique. Il servira d'illustration pour toutes les notions qui seront présentées par la suite. Les systèmes topologiques présentent un espace fondamental dégénéré où il est possible d'encoder de l'information quantique dans des degrés de liberté topologiques, intrinsèquement robustes à l'action locale de l'environnement.

Le chapitre 6 définit la notion de mémoire auto-correctrice et met l'emphase sur deux desiderata de tels dispositifs. D'une part, ils doivent être robustes à une perturbation afin de préserver la structure de leur spectre d'énergie. Nous verrons que l'ordre topologique, en plus

d'une condition technique dite de cohérence locale, permet de garantir cette stabilité. D'autre part, une mémoire auto-correctrice doit protéger l'information lorsqu'elle est mise en contact avec un environnement thermique à température non-nulle. Nous verrons qu'avec un modèle très général de l'environnement, ce desiderata revient à demander l'existence d'une barrière d'énergie entre les différents états fondamentaux.

Le chapitre 7, aboutissement des deux chapitres précédents, s'intéresse plus spécifiquement aux systèmes topologiques 2D et à leur capacité d'être une mémoire auto-correctrice. En particulier, nous nous intéresserons à la structure des opérateurs *logiques* qui effectuent des transformations à l'intérieur de l'espace fondamental. Nous verrons que même si le système est 2D, ces opérateurs logiques ont une structure 1D, ce qui est une source d'instabilité thermique. Cela nous amènera au résultat de l'article [4] qui démontre que les systèmes topologiques 2D dont on sait prouver qu'ils sont robustes aux perturbations ne sont pas stables thermiquement.

1.2.2 Liens entre les deux parties

Même si cette thèse se divise en deux parties, des liens existent entre elles. Ces connexions apparaissent moins au niveau conceptuel qu'au niveau des outils techniques utilisés pour démontrer des résultats. Globalement, ces liens proviennent de l'intersection entre les domaines de l'information quantique et de la physique à n corps. En effet, l'étude des systèmes à n corps du point de vue de l'information amène un éclairage différent et fournit de nouveaux outils.

La première partie de la thèse propose des techniques de caractérisation efficaces car elles ciblent non pas des états choisis au hasard dans l'espace de Hilbert, mais les états physiques qui ne forment qu'une partie infime des états possibles [12]. En effet, des états aléatoires ont des propriétés très étranges et sont définis par un nombre de coefficients qui grandit exponentiellement avec la taille du système. Ainsi, toute tentative de caractérisation pour ces états est exponentiellement coûteuse. Au contraire, les états physiques ont souvent une structure qui permet d'en donner une description efficace. En particulier, on s'intéresse particulièrement aux états fondamentaux de hamiltoniens locaux qui présentent des propriétés particulières. Par exemple, en 1D, les fondamentaux de hamiltoniens locaux gappés obéissent à une loi d'échelle. Cette loi d'échelle est capturée par une des familles que nous considérons en tomographie variationnelle, les *matrix product states* (MPS). Or, l'extension 2D de cette famille variationnelle, les *projected entangled pairs states* (PEPS), sont très utilisés analytiquement afin de mieux comprendre les hamiltoniens des systèmes topologiques

que nous étudions dans la seconde partie de la thèse.

L'autre classe variationnelle dont nous démontrons qu'elle peut être utilisée pour la caractérisation, les états *multi-scale entanglement renormalization ansatz* (MERA), est construite pour décrire les fondamentaux de système critique. En particulier, elle est très efficaces pour décrire les systèmes qui sont des points fixes de méthode de renormalisation. Un exemple de tels points fixes sont les hamiltoniens constitués d'une somme de projecteurs qui commutent. Or, les modèles jouets de systèmes topologiques sont décrits par de tels hamiltoniens. On pense donc que leurs états fondamentaux appartiennent à cette classe variationnelle. Par exemple, le modèle canonique d'ordre topologique, le code torique admet une représentation MERA [13].

Par ailleurs, bien que les hamiltoniens topologiques soient définis sur un réseau 2D, les opérateurs *logiques* qui effectuent des transformations à l'intérieur de l'espace fondamental ont une structure 1D et crée peu d'intrication. Ils peuvent donc être décrits par des *matrix product operators* (MPO), l'équivalent opérateur des états MPS. Ainsi, la compréhension de la structure des MPS dans le but de faire de la caractérisation permet de mieux comprendre la structure des codes topologiques 2D.

Première partie

Caractérisation pratique des systèmes quantiques

Chapitre 2

Caractérisation d'un système quantique

Caractériser un système consiste à en extraire expérimentalement de l'information. Dans le cas d'un système quantique, cette tâche est particulièrement difficile car la mesure perturbe en général le système [14]. Ainsi, la caractérisation d'un système est réalisée en préparant plusieurs fois le système quantique dans le même état et une mesure est effectuée sur chaque copie [1]. Le nombre minimal de copies nécessaire variera en fonction de la tâche à accomplir, c'est-à-dire de l'information que l'on cherche à obtenir.

L'information la plus complète qui puisse être extraite est la description de l'état du système (sous forme de matrice densité) et la tâche correspondante est appelée *estimation d'état* ou *tomographie*. Jusqu'à récemment, les systèmes quantiques étaient caractérisés en deux étapes : une description du système était d'abord reconstruite par tomographie, puis d'autres quantités intéressantes (comme la distance à un état cible) étaient calculées dans un second temps. Or, cette approche n'est possible que pour des systèmes de très petite taille. En effet, nous montrerons dans la section 2.2 que les ressources expérimentales et numériques nécessaires à la tomographie grandissent exponentiellement avec le nombre de particules et deviennent rédhibitoires pour des systèmes d'une dizaine de particules. Notre objectif sera alors de proposer des méthodes de caractérisation plus pratiques qui ne demandent qu'une quantité de ressources qui n'explose pas avec la taille du système. Ceci nous amènera à considérer des tâches réduites dans la section 2.3. Ces tâches ont des objectifs limités, soit car elles n'estiment que certaines quantités particulièrement

1. On se place dans le cas où chaque copie est un système quantique unique, et non pas un ensemble d'états.

intéressantes plutôt que la description complète de l'état expérimental, soit car elles ne s'appliquent qu'à certains ensembles d'états.

Ce chapitre servira donc d'introduction et de revue de littérature avant de se concentrer sur les tâche réduites de certification au chapitre 3 et de tomographie variationelle au chapitre 4. La certification estime la proximité entre l'état expérimental et un état cible que l'on cherchait à préparer. La tomographie variationnelle estime les paramètres variationnels de l'état expérimental, en supposant qu'il soit bien approximé par un état d'une classe variationnelle.

2.1 Motivation

2.1.1 Informatique quantique

Caractériser un système quantique est une tâche essentielle pour l'informatique quantique. Elles est étroitement liée à plusieurs des critères énoncées par DiVicenzo en 2000 [15] pour arriver à une implémentation physique du traitement quantique de l'information. En effet, le deuxième critère exige « la capacité d'initialiser l'état des qubits dans un état de référence simple, par exemple $|000\ldots\rangle$ ». Rappelons qu'un qubit ou *quantum bit* est l'unité de base de l'information quantique. Physiquement, il est est encodé dans l'état d'un système quantique à deux niveaux, par exemple une particule de spin-1/2. Par extension, on parle de *qudit* pour l'état d'un système à d niveaux. Vérifier que l'état de référence a bien été préparé renvoie au problème plus général de valider la préparation d'un état. En effet, une fois l'état de référence préparé, la première démonstration du contrôle quantique du système consiste à préparer des états plus complexes, p.ex. des états intriqués. Le protocole de certification, qui sera présenté dans le chapitre 3, est précisément une vérification de la qualité de la préparation expérimentale.

Deux autres critères de DiVincenzo concernent la caractérisation : le troisième demande « des temps de décohérence très longs devant le temps d'application d'une porte quantique » alors que la quatrième demande de disposer d'une « famille universelle de portes quantiques ». Afin d'allonger les temps de cohérence, une stratégie est de mieux comprendre la décohérence, en caractérisant le modèle de bruit qui agit sur les qubits physiques. Ceci relève de l'estimation dite *de transformation* puisqu'il s'agit d'estimer expérimentalement non pas un état, mais une transformation quantique. Bien que différents conceptuellement, les problèmes d'estimation d'état

et d'estimation de transformation sont équivalents mathématiquement (via l'isomorphisme de Choi-Jamiolkowski, cf. section 3.2.4.2). De même, disposer d'une famille universelle de portes quantiques demande de caractériser précisément les transformations expérimentales, voir p.ex. la caractérisation des portes de Toffoli réalisées à l'ETH Zurich dans le groupe d'Andreas Wallraff [16].

Une des grandes avancées de l'informatique quantique théorique, qui a consolidé l'espoir de mettre au point un ordinateur quantique, est la correction d'erreur quantique et le calcul tolérant aux fautes [6]. Pour simplifier, la tolérance aux fautes montre qu'il n'est pas nécessaire de disposer de transformations parfaites pour pouvoir faire un calcul quantique, aussi long soit-il. Si les portes quantiques présentent un niveau de bruit inférieur à une valeur seuil, il est possible d'effectuer un calcul quantique avec un niveau de bruit aussi faible que désiré. Vérifier les hypothèses de tolérance aux fautes demande de pouvoir caractériser les composants d'un processeur quantique.

2.1.2 Physique à n corps

Au-delà du calcul quantique, la caractérisation d'un système quantique permettrait de mieux comprendre certains phénomènes physiques. Un des problèmes majeurs de la physique à n corps est de comprendre la structure des états fondamentaux de hamiltoniens modèles. Plusieurs états variationnels ont été proposés dans la littérature comme approximation de ces états fondamentaux. Or, nous disposons à l'heure actuelle de systèmes quantiques contrôlés dont on peut ajuster les paramètres afin qu'ils obéissent aux modèles théoriques [17]. De plus, il est possible de préparer de tels simulateurs quantiques dans un état fondamental d'un hamiltonien modèle, p.ex. de façon adiabatique. Caractériser expérimentalement ces états permettrait d'en apprendre plus sur des problèmes physiques fondamentaux. La tomographie variationnelle, présentée au chapitre 4, fournit un outil permettant d'identifier quel état variationnel décrit le mieux l'état expérimental. Ainsi, elle fournit un lien précieux entre la théorie et les expériences.

2.1.3 Problématique

La tâche de reconstituer la description d'un état grâce à des mesures expérimentales sur des systèmes identiquement préparés et un post-traitement numérique classique est appelée estimation d'état ou plus généralement tomographie. Comme nous le verrons dans la section 2.2, les méthodes utilisées pour l'estimation d'état sont très coûteuses expérimentalement et numériquement. De plus,

elles ne sont envisageables que pour des systèmes de petite taille, c.-à-d., pour un petit nombre n de qubits, disons $n \leq 10$ pour fixer les idées. Plus précisément, elles demandent des ressources qui croissent exponentiellement le nombre de qubits. Ainsi, elles ne sont pas *extensibles* (de l'anglais *scalable*) à des systèmes de plus grande taille.

À prime abord, le coût exponentiel de la tomographie n'est pas surprenant. La dimension d de l'espace de Hilbert d'un système à n qubits croît exponentiellement $d = 2^n$. Afin de reconstituer la matrice densité d'un état, il faudra donc reconstruite d^2 coefficients[2], typiquement en estimant les valeurs moyennes de d^2 observables. De plus, ces valeurs moyennes seront généralement exponentiellement petites et demanderont donc un nombre exponentiel de mesures répétées afin de les estimer adéquatement. Finalement, un fois cette quantité exponentielle de données accumulée, il faudra résoudre le problème d'inférence statistique de déterminer une matrice densité compatible avec ces données, ce qui demande un traitement numérique exponentiellement long.

Le coût exponentiel de la tomographie n'est pas seulement un problème théorique, mais un problème d'actualité pour les expérimentateurs. En effet, plusieurs équipes peuvent en ce moment contrôler des systèmes de grandes tailles, pour lesquels la tomographie n'est pas envisageable. Par exemple, l'équipe de Rainer Blatt à Innsbruck peut préparer des états intriqués à $n = 14$ qubits sur des ions piégés [18]. Or, appliquer un protocole de tomographie sur un tel système demande des ressources numériques et expérimentales rédhibitoires. Il faut alors trouver une façon de contourner ce problème. Dans le cas particulier de cette expérience sur 14 qubits, les expérimentateurs ont estimé la fidélité de l'état à leur état cible (un état GHZ) en exploitant la structure simple de l'état cible. Il s'agit donc d'un exemple de tâche réduite puisqu'elle ne vise qu'à obtenir une information limitée : la fidélité à l'état GHZ.

Ce problème de caractériser des systèmes quantiques est vrai pour l'ensemble des technologies quantiques proposées afin de réaliser un ordinateur quantique. En effet, les estimations d'état et de transformations ont été appliquées aux technologies suivantes : les ions piégés (état [19], transformation [20]), l'optique quantique (état [21, 22], transformation [23]), la résonance magnétique nucléaire (état [24], transformation [25]), les qubits supraconducteurs (état [26], transformation [16]), les qubits de spin dans les points quantiques (état [27], transformation [28]), centres radiatifs dans le diamant (état [29], transformation [30]). Cette liste n'est pas exhaustive, mais montre l'ubiquité de la tomographie.

2. Pour un état pur, ou plus généralement un état de rang r, il suffit de reconstruire $\mathcal{O}(rd)$ coefficients et des techniques adaptées existent (voir 2.3.1.3).

Il est donc urgent de trouver des méthodes de caractérisation extensibles à des systèmes de grande taille[3]. Dans la section 2.2, nous expliquerons le fonctionnement de la tomographie quantique afin de montrer que son coût est rédhibitoire pour des systèmes même modérément grands. Cela nous amènera à considérer des tâches moins ambitieuses que l'estimation d'état en 2.3. Après avoir fait l'inventaire de ce qui a été proposé dans la littérature, nous proposerons des protocoles originaux : la certification au chapitre 3 et la tomographie variationnelle au chapitre 4.

2.2 Estimation d'état ou tomographie

2.2.1 Exemple introductif : le *qubyte*

Afin de fixer les idées, nous allons nous intéresser à une expérience [19] réalisée en 2005 dans le groupe de Rainer Blatt à Innsbruck. L'objectif de l'expérience était de préparer un état W

$$|W_n\rangle = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |0\rangle^{\otimes k-1} \otimes |1\rangle \otimes |0\rangle^{\otimes n-k} \tag{2.1}$$

pour $n = 8$ qubits, d'où le nom de *qubyte*. Physiquement, ces qubits correspondaient à 8 ions $^{40}Ca^{+}$ piégés dans un piège de Paul linéaire.

La préparation d'un tel état intriqué sur 8 qubits demandait un grand contrôle expérimental. Afin de transformer l'état initial $|0\rangle^{\otimes n}$ en l'état $|W_8\rangle$, une séquence d'impulsions laser utilise le degré de liberté de mouvement des ions afin de créer de l'intrication. Toutefois, une fois le protocole expérimental mis au point et reproductible, il fallait démontrer que l'état expérimental préparé correspondait bien à l'état W.

Pour ce faire, l'équipe d'Innsbruck a utilisé un protocole de tomographie. Ils ont effectué des mesures dans $3^8 = 6561$ bases différentes en répétant au moins 100 mesures dans chaque base. L'acquisition de l'ensemble des données expérimentales a duré 10 heures. Une fois les mesures expérimentales obtenues, il restait à déterminer une matrice densité compatible avec l'ensemble des mesures expérimentales. Ce problème d'inférence statistique est très exigeant numériquement : il a

3. On pourrait objecter que des portes à un et deux qubits suffisent pour réaliser n'importe quel calcul quantique [31, 32] et qu'il n'est donc pas nécessaire de caractériser des transformations à n qubits. Or, les sources d'erreur sur n qubits sont parfois différentes de celles sur 1 et 2 qubits, comme démontré dans des expériences récentes [18].

demandé une semaine sur une grappe de calcul [33]. Les ressources nécessaires à la tomographie de 8 qubits étaient donc très grandes, particulièrement au niveau numérique.

Afin de réaliser l'ampleur du problème, considérons la même procédure appliquée à non pas 8, mais 16 qubits. La capacité de préparer un état intriqué à 16 qubits est réaliste expérimentalement puisqu'une expérience menée par le même groupe montre des signes de cohérence dans un état à 14 qubits [18]. Quels seraient les ressources nécessaires à la tomographie d'un état à 16 qubits? Expérimentalement, il faudrait mesurer dans 3^{16} bases différentes. En supposant que les mesures prennent le même temps et que seulement 100 mesures soient effectuées dans chaque base[4], on obtient une durée de calcul de 7,5 années, c'est-à-dire que plusieurs doctorants travailleraient sur la même expérience avant d'obtenir des résultats! Numériquement, en supposant que la complexité du problème d'inférence grandisse seulement avec la dimension de l'espace d'opérateurs $\mathcal{O}(4^n)$, on obtient un temps de calcul de l'ordre du millénaire ! Pire, en supposant que le traitement numérique varie comme $\mathcal{O}(4^{3n})$, *i.e.* qu'il varie comme le cube des données à traiter (nous argumenterons cette dépendance en 2.2.4.2), le temps de traitement dépasse l'âge de l'Univers ! Ainsi, il est urgent de proposer des méthodes de caractérisation adaptées au système de grande taille.

2.2.2 Mesures expérimentales

Nous allons maintenant donner une description formelle de la tâche de tomographie. Plus précisément, nous nous intéressons à la tomographie dite d'état où un appareil (dont nous ne chercherons pas à décrire le fonctionnement interne) fournit sur demande un système quantique $\mathcal{S}$ préparé dans le même état, a priori inconnu, décrit par une matrice densité ρ.

2.2.2.1 Description de la mesure

Sur chacune des copies du système, une mesure est effectuée. Cette mesure peut être décrite par une observable O, c.à.d. un opérateur hermitien dont les valeurs propres λ_i sont réelles et dont les espaces propres, caractérisés par leur projecteur Π_i, sont orthogonaux. Dans le jargon de l'informatique quantique, on parle dans ce cas de *mesure projective*. Formellement, l'observable se

4. N'effectuer que 100 mesures est discutable car 1) le choix de 100 mesures pour l'expérience à 8 qubits est arbitraire, 2) les probabilités à estimer diminuent avec la taille du système donc une estimation correcte demanderait plus de mesures répétées si le système est plus grand.

décompose en blocs orthogonaux,

$$O = \bigoplus_i \lambda_i \Pi_i \tag{2.2}$$

et la probabilité de mesurer la valeur propre λ_i est donnée par la règle de Born

$$p_i = \mathrm{Tr}\rho\Pi_i. \tag{2.3}$$

Observables vs POVM Plus généralement, une mesure peut être donnée par un POVM (positive operator valued measure), décrit par une famille d'opérateurs hermitiens positifs semidéfinis qui somment à l'unité

$$\sum_i F_i = \mathbb{I}. \tag{2.4}$$

La probabilité d'obtenir le résultat i à la mesure est donnée par

$$p_i = \mathrm{Tr}\rho F_i. \tag{2.5}$$

Ainsi, les observables correspondent au cas particulier où $F_i = \Pi_i$, *i.e.*, les éléments sont des projecteurs orthogonaux. En fait, un POVM correspond toujours à une mesure projective dans un espace de Hilbert de plus grande dimension que celui où vit l'état ρ. Physiquement, mesurer un POVM revient à adjoindre au système expérimental $\mathcal{S}$ un autre système quantique $\mathcal{A}$, dit *ancillaire*, afin d'effectuer une mesure projective sur le système $\{\mathcal{S} + \mathcal{A}\}$.

La règle de Born, éq. (2.3), définit une variable aléatoire (v.a.) qui prend la valeur λ_i avec probabilité p_i. La valeur moyenne de l'observable O sur l'état ρ est l'espérance de cette v.a.

$$\langle O \rangle_\rho = \sum_i p_i \lambda_i = \mathrm{Tr}\rho O. \tag{2.6}$$

Les observables faciles à mesurer expérimentalement varient fortement selon la technologie. Or, il est important qu'une proposition théorique tiennent compte des contraintes expérimentales. Nous nous concentrerons sur un type particulier d'observables : les opérateurs de Pauli.

2.2.2.2 Opérateurs de Pauli

Opérateur de Pauli à 1 qubit Les matrices de Pauli exprimées dans la base $\{|0\rangle, |1\rangle\}$ sont

$$\sigma_0 = \mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \sigma_1 = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \sigma_2 = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \sigma_3 = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.7}$$

Elles sont hermitiennes, unitaires et forment une base des observables sur un qubit. De plus, elles obéissent à l'algèbre

$$\sigma_a \sigma_b = \delta_{ab}\mathbb{I} + i \sum_c \varepsilon^{abc} \sigma_c \tag{2.8}$$

et sont orthogonales pour le produit scalaire de Hilbert-Schmidt $\langle A, B \rangle \equiv \mathrm{Tr} A^\dagger B$.

Les opérateurs de Pauli à un qubit sont les éléments du groupe $\mathcal{P}_1$ généré par les matrices de Pauli. Pour simplifier, on oublie souvent la phase devant un opérateur de Pauli P, c.à.d. qu'on considère que les quatre opérateurs $\pm P, \pm iP$ appartiennent à la même classe d'équivalence. Pour chaque classe d'équivalence, on choisit un représentant canonique. Pour le groupe $\mathcal{P}_1$, on choisit les quatre représentants suivant

$$\mathbb{I} \quad X \quad XZ(=iY) \quad Z \tag{2.9}$$

qui ont tous la forme $X^a Z^b$ où $a, b \in \{0, 1\}$ sont des bits classiques. Ceci permet de distinguer la « composante X » et la « composante Z » d'un opérateur de Pauli.

Opérateur de Pauli à n qubits Pour n qubits, les opérateurs de Pauli P_i sont simplement les 4^n produits tensoriels possibles d'opérateurs de Pauli sur un qubit. Par exemple, l'opérateur

$$X_1 \otimes Y_2 \otimes \mathbb{I}_3 \otimes Z_4 \otimes Z_5 \tag{2.10}$$

est un opérateur de Pauli sur 5 qubits où l'opérateur X agit sur le premier qubit, l'opérateur Y sur le second et des opérateurs Z sur les deux derniers qubits. Le poids d'un opérateur de Pauli est le nombre de qubits pour lequel il agit non-trivialement : l'opérateur de Pauli donné en (2.10) est de poids 4 car il n'agit trivialement que sur le qubit 3.

Les opérateurs de Pauli sur n qubits forment un groupe noté $\mathcal{P}_n$. Comme pour le cas à un qubit, on oublie souvent la phase devant un opérateur de Pauli P afin d'obtenir des classes d'équivalence dont le représentant canonique est de la forme $\bigotimes_i X_i^{a_i} Z_i^{b_i}$ où $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$ sont des chaînes de n bits. On représente alors l'opérateur de Pauli par sa notation symplectique $\langle \mathbf{a}|\mathbf{b}\rangle$. Ainsi, l'opérateur de Pauli (2.10) aura pour notation symplectique

$$\langle 11000|01011\rangle. \tag{2.11}$$

2.2.2.3 Mesure expérimentale de la valeur moyenne d'une observable

Afin de mesurer expérimentalement la valeur moyenne d'une observable, c.f. éq. (2.6), des mesures répétées de la même observable seront effectuées sur des copies préparés dans le même état ρ. En terme de probabilité, chaque mesure correspond à un tirage d'une valeur propre selon la probabilité donnée par la règle de Born, éq. (2.3).

Les opérateurs de Pauli sont des observables particulièrement simples puisqu'elles n'ont que deux valeurs propres possibles ± 1. En effet, la décomposition en bloc orthogonal, cf. éq. (2.2), d'un opérateur de Pauli P est

$$P = \frac{\mathbb{I} + P}{2} - \frac{\mathbb{I} - P}{2} = \Pi_+ - \Pi_- \tag{2.12}$$

où $\Pi_\pm$ sont les projecteurs sur l'espace propre ± 1.

Du point de vue probabilité, une mesure correspond donc à une épreuve de Bernoulli de paramètre $p = \mathrm{Tr}\rho\Pi_+$. Formellement, la mesure d'un opérateur de Pauli est une v.a. $X = 2\tilde{X} - 1$ où $\tilde{X}$ est une v.a. de Bernoulli de paramètre p. La répétition de N mesures se traduit par N épreuves de Bernoulli indépendantes (en supposant que les N copies du système soient préparées identiquement et indépendantes) et correspond donc à une v.a. $Y = \frac{1}{N}\sum_i X_i = \frac{2}{N}\tilde{Y} - 1$ où $\tilde{Y}$ obéit à une loi binomiale à N épreuves de probabilité de succès p. Pour N grand, la distribution binomiale converge vers la distribution normale (théorème de Moivre-Laplace, un cas particulier du théorème de la limite centrale). Afin de déterminer le nombre de mesures N nécessaire afin d'obtenir une estimation fiable de l'espérance, calculons la variance $\mathbb{V}(Y) = \frac{4}{N^2}\mathbb{V}(\tilde{Y}) = \frac{4p(1-p)}{N}$. Ainsi, la précision varie comme $\frac{1}{\sqrt{N}}$.

Estimation de la valeur moyenne d'un produit tensoriel Les opérateurs de Pauli ont la propriété d'être le produit tensoriel d'observables à un qubit, ce qui est particulièrement pratique

expérimentalement. En effet, afin d'estimer la valeur moyenne de l'opérateur de Pauli sur n qubits, il suffit d'effectuer une séquence de mesures sur un qubit. En effet, il suffit de remarquer que

$$\mathrm{Tr}\left[(P_1 \otimes P_2)\,\rho\right] = \mathrm{Tr}\left[P_2 \mathrm{Tr}_1\left[(P_1 \otimes \mathbb{I})\,\rho\right]\right] \tag{2.13}$$

pour se convaincre que la mesure de $P_1 \otimes P_2$ peut se faire en mesurant d'abord P_1 suivi de la mesure de P_2 sur la même copie du système [5].

Ainsi, une mesure d'un opérateur de Pauli à n qubits $P_i = \bigotimes_{k=1}^{n} \sigma_{i_k}$ se décomposera en n mesures successives. Il peut alors être utile de garder toutes ces valeurs intermédiaires, ce qui fournit plus d'information. Formellement, on obtient n bits d'information classique plutôt que un. Ceci peut se révéler utile afin de minimiser le nombre de manipulations expérimentales.

Méthode « 3^n bases » Un protocole très employé, en particulier pour l'expérience du qubyte [19], est la méthode dite à « 3^n bases » détaillée dans [34]. Cette méthode permet de caractériser un système quantique en effectuant des mesures dans 3^n bases distinctes. En apparence, l'idée est de ne mesurer que des opérateurs de Pauli de poids maximal, c.à.d ceux qui agissent non-trivialement sur tous les qubits. En fait, le protocole est plus subtil : il consiste à choisir pour chaque qubit de mesurer soit dans une des trois bases $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, $\{|+\rangle\langle +|, |-\rangle\langle -|\}$ ou $\{|+i\rangle\langle +i|, |-i\rangle\langle -i|\}$. En termes d'optique, la première base correspond aux polarisations horizontales et verticales, la deuxième, où $|\pm\rangle = \frac{1}{\sqrt{2}}|0\rangle \pm \frac{1}{\sqrt{2}}|1\rangle$, aux polarisations diagonales et la troisième, où $|\pm i\rangle = \frac{1}{\sqrt{2}}|0\rangle \pm i\frac{1}{\sqrt{2}}|1\rangle$ aux polarisations circulaires. Ce protocole fournit 6^n mesures qui ne sont pas linéairement indépendantes et permettent de reconstruire les valeurs moyennes des 4^n opérateurs de Pauli.

2.2.3 Reconstruction d'état

Supposons maintenant qu'une approximation $\tilde{\rho}_i$ de la valeur moyenne de chaque opérateur de Pauli $\rho_i \equiv \mathrm{Tr}\left[\rho P_i\right]$ ait été obtenue expérimentalement. Il s'agit maintenant de reconstruire une matrice densité ρ compatible avec toutes ces mesures.

5. Pour d'autres tâches, p.ex. la correction d'erreur, la mesure d'un seul coup (*single-shot*) de $P_1 \otimes P_2$ est strictement nécessaire.

Pour ce faire, notons que les opérateurs de Pauli forment une base orthogonale donc

$$\rho = \frac{1}{2^n} \sum_{i=1}^{4^n} \mathrm{Tr}\left[\rho P_i\right] P_i. \tag{2.14}$$

2.2.3.1 Pseudo-inverse

La méthode la plus simple consiste à rapporter l'état

$$\tilde{\rho}_{inv} = \frac{1}{2^n} \sum_{i=1}^{4^n} \tilde{\rho}_i P_i. \tag{2.15}$$

Il est important de noter que cette méthode d'estimation est particulièrement simple numériquement lorsque l'on mesure des opérateurs de Pauli. Si l'expérience avait fourni les valeurs moyennes d'observables qui ne sont pas linéairement indépendantes, il aurait fallu inverser une matrice de taille exponentiellement grande pour obtenir l'estimation de l'état. Cette méthode porte donc généralement le nom de « pseudo-inverse ».

Cette estimation serait exacte en absence d'incertitude sur les valeurs moyennes, *i.e.*, dans le cas où $\tilde{\rho}_i = \rho_i$. Ce cas de figure idéal correspond à un nombre infini de mesures parfaites : il n'est pas réaliste expérimentalement. En présence d'incertitude, reconstruire la matrice densité compatible avec toutes les mesures expérimentales est un problème d'inférence statistique non-trivial. En effet, l'opérateur (2.15) ne sera pas en général une matrice densité. En particulier, il pourrait avoir des valeurs propres strictement négatives.

2.2.3.2 Maximum de vraisemblance (MV)

Afin de garantir que l'opérateur estimé à partir des données expérimentales soit bien une matrice densité, Hradil a proposé une méthode plus raffinée [35]. Celle-ci est basée sur le concept statistique de vraisemblance, c.à.d. la probabilité $\mathcal{L}(\tilde{\rho}) \equiv p(\mathcal{M}|\tilde{\rho})$ d'obtenir les mesures observées $\mathcal{M}$ connaissant la matrice densité $\tilde{\rho}$. La technique de maximum de vraisemblance retourne la matrice densité $\tilde{\rho}_{MV}$ qui maximise [6] la vraisemblance $\mathcal{L}(\tilde{\rho})$.

Conceptuellement, la technique est simple. Étant donné une description des mesures $\mathcal{M} = \{M_i\}_{i=1}^{N}$ où M_i est l'opérateur hermitien positif correspondant à la mesure i (projecteur dans le

6. En principe, il faudrait effectuer une inversion bayésienne, ce qui est fait dans l'approche discutée en 2.2.3.3.

cas d'une observable ou élément de POVM en général), la vraisemblance s'écrit

$$\mathcal{L}(\tilde{\rho}) = \prod_i \mathrm{Tr}\left[M_i\tilde{\rho}\right]. \tag{2.16}$$

En pratique, maximiser la vraisemblance revient à maximiser $\log\mathcal{L}(\tilde{\rho}) = \sum_i \log\left(\mathrm{Tr}\left[M_i\tilde{\rho}\right]\right)$ qui est une fonction convexe. Numériquement, cette maximisation peut se faire grâce à de la programmation semi-définie, p.ex. l'algorithme du simplexe.

La technique MV est la méthode de choix depuis sa première utilisation sur des données expérimentales en 2001 [36]. En particulier, elle a été utilisée pour l'expérience du qubyte [19].

Toutefois, la technique MV a été remise en question dans les dernières années, cf. [33], car elle produit des matrices densité qui n'ont pas plein rang, c.à.d qu'elles ont des valeurs propres nulles. Or, une valeur propre nulle revient à prédire avec exactitude le résultat d'une mesure particulière, ce qui ne peut être fait avec un nombre fini de mesures. De plus, les valeurs propres nulles sont incompatibles avec une analyse d'erreur rigoureuse.

2.2.3.3 Approche bayésienne

Afin de pallier ces problèmes, une approche bayésienne a été proposée dans [33] : plutôt que de maximiser $\mathcal{L}$, i) on choisit une distribution de probabilité a priori $\pi_0(\rho)d\rho$ sur l'espace des matrices densité, ii) on évalue comment les mesures obtenues modifient cette distribution afin d'obtenir une distribution a posteriori $\pi_f(\rho)d\rho \propto \mathcal{L}(\rho)\pi_0 d\rho$, iii) on rapporte la valeur moyenne de cette distribution a posteriori

$$\tilde{\rho}_{Bayes} = \int \rho\pi_f(\rho)d\rho = \frac{\int \rho\mathcal{L}(\rho)\pi_0 d\rho}{\int \mathcal{L}(\rho)\pi_0 d\rho}. \tag{2.17}$$

Cette approche comporte elle aussi des problèmes conceptuels, en particulier le choix de la distribution a priori π_0. De plus, elle est très coûteuse numériquement.

2.2.3.4 Vers une tomographie plus rigoureuse

Les différentes approches de reconstruction d'état retournent des états différents, dont les propriétés varient beaucoup. Par exemple, il est dangereux de déterminer si un état est intriqué en estimant d'abord la matrice densité : la réponse peut en effet varier selon la méthode utilisée [37].

Nous verrons qu'il est toutefois possible de définir une tâche réduite, la vérification d'intrication (cf. 2.3.2.1), afin de déterminer directement si un état est intriqué.

Jusqu'à la proposition de l'approche bayésienne, l'analyse d'erreur pour les estimés tomographiques était des méthodes heuristiques. Dans les dernières années, des approches plus rigoureuses ont été proposées [33, 38]. Toutefois, ces méthodes demeurent très coûteuses numériquement pour traiter des systèmes plus grands que quelques qubits.

Un problème majeur avec la tomographie est aussi la cohérence de ses hypothèses. En effet, afin de caractériser un système quantique, on effectue des mesures. Ces mesures proviennent souvent d'interactions avec d'autres systèmes-sondes quantiques. Pour la tomographie, on découpe artificiellement le monde en deux : le système expérimental à caractériser d'une part et les systèmes-sondes d'autre part, *que l'on suppose parfaitement caractérisés.* Formellement, cela revient à avoir une connaissance parfaite des observables (et plus généralement des POVM) décrivant la mesure. Souvent, faire cette hypothèse n'est pas satisfaisant. Ce problème est encore plus criant lorsque l'on passe de la tomographie d'état à la tomographie de transformation. Dans ce dernier cas, il faut non seulement caractériser les mesures, mais aussi la préparation des états-sondes. Une proposition intéressante afin d'obtenir une tomographie de processus auto-cohérente a été formulée dans [39] et il s'agit d'un domaine de recherche actif.

Maintenant que nous avons vu le fonctionnement de la tomographie, nous allons montrer qu'elle requiert une quantité exponentielle de ressources. Elle n'est donc pas envisageable pour des systèmes de grande taille.

2.2.4 Ressources nécessaires à la caractérisation

Un protocole de caractérisation se décompose en une étape expérimentale, où de l'information est acquise par des mesures, et une étape numérique, où l'information est traitée. À ces deux étapes d'un protocole correspondent des ressources, qui sont soit expérimentales, p.ex. le nombre total de mesures, soit numériques, p.ex. le temps de calcul. Dans la perspective d'obtenir des protocoles efficaces pour des systèmes de grande taille, il est important de savoir comment varient ces ressources avec le nombre n de particules du système à caractériser. Pour simplifier la discussion, nous supposerons qu'on s'intéresse à n qubits[7] dont l'espace de Hilbert est de dimension $d = 2^n$.

7. La même analyse tiendrait pour n qudits. Un qudit est une particule décrite par un espace de Hilbert de dimension finie d.

Pour un article complet dédié à l'analyse comparative des ressources nécessaires pour la tomographie de transformation, voir [40].

2.2.4.1 Ressources expérimentales

Avant de déterminer comment les ressources expérimentales varient, il convient de s'accorder sur les mesures permises. Évidemment, la mécanique quantique interdit certaines mesures. Toutefois, les mesures permises théoriquement sont souvent inconcevables expérimentalement : une observable correspondant à mesurer dans une base hautement intriquée de n particules n'est pas réaliste, au moins avec la technologie actuelle. Les contraintes pertinentes sont donc données par ce qui est possible de faire expérimentalement. Généralement, il n'est possible que de faire des mesures sur une particule (voire quelques-unes). En particulier, mesurer des opérateurs de Pauli est réaliste pour les technologies expérimentales actuelles car elles sont le produit (tensoriel) de mesures à une particule.

Les ressources expérimentales sont quantifiées par le nombre N_1 d'observables distinctes à mesurer et le nombre total de mesures $N = \sum_{k=1}^{N_1} N_2^{[k]}$, qui tient compte des mesures répétées $N_2^{[k]}$ pour estimer une valeur moyenne $\rho_k \equiv \mathrm{Tr}\,[\rho P_k]$. Dans le cas de la tomographie, il est nécessaire de mesurer les 4^n observables de Pauli, *i.e.*, $N_1 \in \mathcal{O}(4^n)$. Afin d'estimer la valeur moyenne de chaque observable, l'erreur due aux fluctuations statistiques varie comme $1/\sqrt{N_2^{[k]}}$. Pour garantir une erreur ϵ sur l'estimation de la valeur moyenne, il faut donc répéter la mesure $N_2^{[k]} \sim \rho_k^{-2}$. Or, les valeurs moyennes ρ_k sont elles-mêmes typiquement exponentiellement petites, $\rho_k \sim 2^{-n}$ pour de grands systèmes. Ainsi, le nombre de répétitions sera exponentiellement grand $N_2 \in \mathcal{O}(4^n)$. Notons qu'il n'est pas possible de déterminer $N_2^{[k]}$ a priori en absence d'information sur l'état.

Finalement, la méthode de reconstruction utilisée peut avoir une influence sur le nombre de mesures, en particulier afin de garantir l'erreur maximale entre l'état et son estimation. En effet, à l'issue de la reconstruction d'état, on aimerait avoir une borne sur la distance entre l'état estimé $\tilde{\rho}$ et l'état expérimental ρ par rapport à une mesure de distance, par exemple la fidélité

$$F(\rho, \tilde{\rho}) \equiv \mathrm{Tr}\,[\sqrt{\rho}\tilde{\rho}\sqrt{\rho}] \tag{2.18}$$

ou encore la norme L_1, appelée norme de trace

$$\|\rho - \tilde{\rho}\|_1 \equiv \mathrm{Tr}\left[\sqrt{(\tilde{\rho} - \rho)^2}\right]. \tag{2.19}$$

Or, l'algorithme de reconstruction d'état va déterminer comment l'erreur relative sur l'estimation des valeurs moyennes influence ces distance. Ceci relève de l'analyse d'erreur qui reste un domaine de recherche actif, cf. [38].

Analyser un protocole de caractérisation particulier permet d'évaluer le nombre total de mesures nécessaire à ce protocole précis. De façon complémentaire, il est aussi possible de donner des bornes inférieures théoriques sur le nombre total de mesures nécessaires à une tâche. Flammia et Liu ont prouvé une borne inférieure intéressante quand le protocole est restreint à des mesures de Pauli sur une copie du système (pas de mesures corrélées sur plusieurs copies) [41]. Pour ce faire, ils définissent opérationnellement la tomographie comme un protocole qui distingue n'importe quel deux états ρ et σ avec précision Δ, *i.e.* qui retourne un résultat différent si les états sont Δ différents l'un de l'autre[8]. Flammia et Liu montrent que $\Omega\left(d^2/\log d\right)$ mesures sont nécessaires. Ainsi, du point de vue du nombre de copies et pour un état mixte quelconque, les protocoles de tomographie sont quasi-optimaux.

2.2.4.2 Ressources numériques

Les ressources numériques afin de traiter les données expérimentales sont essentiellement le temps de calcul et la taille mémoire.

Notons que le simple fait de stocker le nombre exponentiel de coefficients d'une matrice densité exige beaucoup de mémoire. Par exemple, stocker la matrice densité de 14 qubits avec des réels double-précision (64 bits) demande plus de 4 Go. Heureusement, il existe des représentations plus concises des états quantiques (cf. section 4.1). On voit ainsi pondre le problème de la généralité de la tomographie : en permettant la reconstruction de n'importe quel état quantique, elle demande de manipuler des matrices dont la taille est immense.

Pour l'estimation d'état, le temps de calcul dépend de la méthode employée. La complexité de l'approche MV n'est pas connue, mais puisqu'elle correspond à trouver l'état quantique le plus proche de l'opérateur fourni par l'approche pseudo-inverse, cf. [33], on peut estimer que sa complexité ressemble à celle de diagonaliser une matrice de taille 4^n, *i.e.* $\mathcal{O}(4^{3n})$ en pratique. En extrapolant les chiffres de l'expérience du qubyte pour appliquer la méthode MV à $n = 9$ qubits, on obtiendrait alors plus d'une année pour reconstruire la matrice densité. Ainsi, les ressources numériques semblent être le facteur limitant pour la tomographie à l'heure actuelle.

8. Formellement, leur fidélité $F(\rho, \sigma)$ obéit à $F(\rho, \sigma) \leq 1 - \Delta$.

En conclusion, la tomographie telle que décrite jusqu'ici n'est pas envisageable pour plus d'une dizaine de particules. Il faut donc trouver un moyen de caractériser les systèmes quantiques autrement.

2.3 Tâches réduites

Le coût exponentiel de la tomographie est lié à la généralité de la tâche à accomplir. En effet, l'objectif est de reconstruire un état quantique quelconque sur n qubits. Le simple fait d'écrire la matrice densité d'un tel état demande $d^2 - 1$ coefficients complexe. De plus, la borne inférieure de Flammia & Liu montre que le nombre total de mesures doit grandir comme d^2. Ainsi, pour espérer arriver à une caractérisation efficace, idéalement qui ne demande qu'une quantité polynomiale (en n) de ressources, il faut considérer des tâches moins générales. Deux options ont été explorées jusqu'à présent dans la littérature : l'estimation d'états appartenant à une classe particulière et l'estimation ciblée d'un petit nombre de paramètres. Plus précisément, j'ai contribué à introduire ces idées de tâches ciblées. Nous allons donc faire une revue de la littérature de ce qui a été fait par d'autres équipes avant de présenter les contributions originales de cette thèse à la caractérisation quantique dans les chapitres 3 et 4.

2.3.1 États particuliers

Plusieurs des états auxquels s'intéressent l'informatique quantique et la physique présentent beaucoup de structure. Ainsi, plusieurs classes d'états peuvent être décrits par un petit nombre de coefficients. Ces descriptions efficaces seront abordées en détails au chapitre 4. Ceci ouvre la possibilité d'estimer ces coefficients à l'aide d'un petit nombre de mesures expérimentales. Cette tâche est généralement désignée par le terme « apprentissage ». Notons qu'apprendre l'état expérimental ne requiert pas qu'il appartienne exactement à la classe d'états, ce qui serai irréaliste : il suffit qu'il soit bien approximé par un état de cette classe.

2.3.1.1 États produits

L'exemple le plus simple est celui des états produits qui peuvent être décrits par $\mathcal{O}(n)$ paramètres complexes. En effet, il suffit de décrire l'état individuel de chaque particule pour reconstruire l'état global. Ceci demande $\mathcal{O}(n)$ mesures, mais seulement un nombre constant ($\mathcal{O}(1)$) de copies du système puisque les mesures peuvent être faites sur chaque particule indépendamment.

2.3.1.2 États stabilisateurs

Plus intéressante, la classe des états stabilisateurs décrit des états intriqués, p.ex. l'état W. Un état stabilisateur est l'état propre associé à la valeur propre $+1$ d'un sous-groupe abélien $\mathcal{S}$ des opérateurs de Pauli qui ne contient pas l'opérateur $-\mathbb{I}$. Ainsi, un état stabilisateur est de la forme

$$\forall S \in \mathcal{S} \qquad S|\psi\rangle = +|\psi\rangle.$$

Le groupe stabilisateur $\mathcal{S}$ d'un état (stabilisateur) à n qubits est décrit par n générateurs $\mathcal{S} = \langle G_1, G_2, \ldots, G_n \rangle$ qui sont des opérateurs de Pauli qui commutent deux à deux et sont linéairement indépendants[9]. Par exemple, l'état GHZ sur n qubits $|GHZ_n\rangle \propto |0\rangle^{\otimes n} + |1\rangle^{\otimes n}$ est un état stabilisateur et son groupe stabilisateur est généré par les corrélations à deux corps $Z_i Z_{i+1}$ pour $1 \leq i < n$ et l'opérateur à n corps $\bigotimes_{i=1}^{n} X_i$.

Il est possible de reconstruire la description d'un état stabilisateur efficacement avec seulement $\mathcal{O}(n)$ copies du système et un effort numérique de complexité $\mathcal{O}(n^3)$ grâce à des mesures de Bell sur deux copies du système [42].

2.3.1.3 États quasi-purs

L'informatique quantique et la physique s'intéressent tout particulièrement aux états purs. Formellement, ceux-ci correspondent à des matrices densité de rang 1, autrement dit des projecteurs sur l'état pur $\rho = |\psi\rangle\langle\psi|$. Plus généralement, on peut s'intéresser aux états dont le rang r est petit. Un protocole d'estimation a été développé pour de tels états et il n'utilise que $\mathcal{O}(rd \log d)$ copies du système [43]. Il repose sur des outils mathématiques sophistiqués issus du domaine de l'acquisition comprimée (*compressed sensing*) qui analyse comment reconstruire un signal avec un très petit nombre d'échantillons [44].

9. Il est particulièrement pratique de fournir ces générateurs en notation symplectique.

2.3.1.4 Pretty good tomography

La notion de « *pretty good tomography* » présentée par Aaronson [45] est difficile à classer. Strictement, elle ne vise pas à reconstruire la description particulière d'états. En fait, elle considère un ensemble fixe de mesures $\mathcal{M}$ qui pourraient être faites sur le système et reconstruit une description qui reproduira les résultats de presque toutes les mesures tirées de $\mathcal{M}$ à l'aide de poly(n) copies du système. Par exemple, pour un grand système $n \gg 1$, la vaste majorité des mesures expérimentales ne pourront pas distinguer le véritable état de l'état maximalement mélangé $\frac{1}{d}\mathbb{I}$.

2.3.1.5 Classes variationnelles d'états

En physique, l'utilisation de classes variationnelles d'états est cruciale pour faire avancer notre compréhension d'un phénomène et sont à la base de plusieurs approches numériques. Nous reviendrons en détails au chapitre 4 sur l'idée de tomographie variationnelle dont l'objectif est d'estimer expérimentalement les paramètres variationnels de l'état expérimental. En particulier, nous présenterons deux protocoles originaux afin d'apprendre les états à produit matriciel (*matrix product states* ou MPS) utilisés en DMRG (*density matrix renormalisation group*) et les MERA (*multi-scale entanglement renormalization ansatz*).

2.3.2 Estimation de paramètres

Plutôt que de reconstruire la description d'un état, il suffit parfois d'estimer un paramètre particulièrement intéressant (ou un petit nombre d'entre eux).

2.3.2.1 Vérification d'intrication

L'exemple-type en informatique quantique d'une telle tâche réduite est celui de la vérification d'intrication. Le problème est de vérifier s'il existe de l'intrication entre deux sous-systèmes A et B d'un système expérimental. Il s'agit d'une question binaire dont le résultat est 1 s'il y a de l'intrication et 0 sinon. Pour ce faire, on utilise un témoin d'intrication, un opérateur hermitien A tel que pour tout état séparable ρ_{sep}, on a $\mathrm{Tr}\left[A\rho_{sep}\right] \geq 0$. Il suffirait alors de mesurer A et d'obtenir une valeur moyenne strictement négative afin d'affirmer que le système expérimental est intriqué. En général, A n'est pas une mesure simple à effectuer.

2.3.2.2 Étalonnage aléatoire (*randomized benchmarking*)

L'étalonnage aléatoire [46] a été développé afin de vérifier les hypothèses du calcul tolérant aux fautes et vise à estimer la probabilité d'erreur par porte quantique pour des portes choisies dans le groupe de Clifford. Le groupe de Clifford C est le normalisateur du groupe de Pauli, *i.e.*, l'ensemble des transformations unitaires U qui envoie un opérateur de Pauli $P \in \mathcal{P}$ par conjugaison vers un opérateur de Pauli $UPU^\dagger \in \mathcal{P}$. Le groupe de Clifford joue un rôle particulier car on dispose d'un algorithme classique efficace pour simuler l'application de ces portes [47].

Pour évaluer la probabilité d'erreur par porte, une séquence aléatoire de m portes de Clifford $U_m \dots U_1$ est appliqué à un état de référence, p.ex. $|0\rangle \equiv |0\rangle^{\otimes n}$, et une mesure finale détermine si l'état expérimental est bien $U_m \dots U_1|0\rangle$, ce qui est possible grâce à la simulation efficace mentionnée plus haut. En faisant varier m, on peut estimer la probabilité d'erreur moyenne par porte. Ce protocole a la propriété intéressante d'être robuste aux erreurs de préparation et de mesure.

Des approches similaires utilise la notion de tournoiement (*twirling*) [48, 49, 50] qui donne accès à une version symétrisée du modèle de bruit. Par exemple, pour un qubit, si le vrai modèle de bruit est $\Lambda(\rho) = \sum_i p_i \sigma_i \rho \sigma_i$, sa version tournoyée est $\bar{\Lambda}(\rho) = p_0\rho + \frac{p_1}{3}(X\rho X + Y\rho Y + Z\rho Z)$. La proposition [50] utilise $\mathcal{O}(n^2)$ portes, a une complexité numérique en $\mathcal{O}(n^4)$ et ne demande qu'un nombre constant de copies. De plus, elle est aussi robuste aux erreurs de préparation et de mesure. Finalement, le protocole [51] permet d'évaluer n'importe quel élément $\chi_{mm'}$ de la décomposition $\Lambda(\rho) = \sum_{mm'} \chi_{mm'} P_m \rho P_{m'}$ à l'aide de ressources polynomiales.

2.3.2.3 Certification

La tâche de certification consiste à vérifier la qualité d'une préparation d'état expérimentale en mesurant la distance entre l'état expérimental et l'état cible. Dans le prochain chapitre 3, nous expliquerons en détails un protocole qui permet d'estimer directement ce paramètre grâce à une approche Monte-Carlo et qui est très efficace pour une vaste classe d'états cibles.

Chapitre 3

Certification

Ce chapitre s'articule autour de l'article

> Practical characteriztion of quantum devices without tomography
> Marcus P. da Silva, Olivier Landon-Cardinal, David Poulin.
> *Physical Review Letters*, **107**, 210404 (2011)

Afin de faciliter la compréhension de l'article, nous le motiverons en abordant ses idées principales dans la section 3.1. Dans la section suivante 3.2, nous reproduirons l'article dans son intégralité (article et matériel supplémentaire). Finalement, dans la section 3.3, nous éclaircirons tout d'abord quelques points techniques en 3.3.1, avant de discuter de deux questions importantes soulevées par ce travail : comment améliorer une préparation imparfaite d'un état (en 3.3.2) et quels sont les états pour lesquels notre protocole de certification est efficace (en 3.3.3)?

Notons qu'un résultat très similaire à celui-ci de cet article a été obtenu indépendamment par Flammia et Liu [41].

3.1 Motivation

3.1.1 Vérification de la procédure expérimentale

Un des objectifs majeurs de l'informatique quantique expérimentale est d'atteindre la capacité de contrôler avec grande précision un système quantique et de préserver sa cohérence. En particulier, préparer des états quantiques sur des systèmes de grande taille est un défi expérimental. Il s'agit de concevoir un protocole expérimental reproductible produisant un état expérimental aussi proche que possible d'un état cible. Par exemple, dans l'expérience du qubyte [19], l'objectif était de préparer un état $|W_n\rangle$ sur $n = 3 \ldots 8$ ions.

Une fois le protocole expérimental mis au point, il faut démontrer que l'état expérimental correspond bien à l'état cible. Pour se faire, on estime la proximité de ces deux états. Dans le cas du qubyte, la tomographie a permis d'estimer la matrice densité $\hat{\sigma}$ de l'état expérimental puis, dans un second temps, la distance à l'état cible $\hat{\rho} = |W_n\rangle\langle W_n|$ a été calculée, en utilisant la fidélité $F(|W_n\rangle, \hat{\sigma}) = \langle W_n|\hat{\sigma}|W_n\rangle$ comme mesure de qualité.

Supposons que l'expérimentateur ne soit intéressé qu'à estimer la fidélité, p. ex. car il est très confiant de la qualité de sa préparation. Dans ce cas, l'étape de tomographie paraît superflue. Bien sûr, la matrice densité fournit beaucoup plus d'information[1] que la simple distance à l'état cible, mais toute cette information n'est pas nécessaire s'il s'agit d'estimer seulement la fidélité.

L'idée de base de notre protocole est précisément d'estimer directement la fidélité d'un état expérimental $\hat{\sigma}$ à l'état cible pur $\hat{\rho} = |\psi\rangle\langle\psi|$. Nous appellerons cette tâche *certification* puisqu'elle a pour objectif de certifier que la préparation expérimentale fournit un état proche de l'état cible. Évidemment, ceci est d'autant plus intéressant que les deux états sont proches. Or, un important effort expérimental a été consacré à ce que ce soit justement le cas. Ainsi, on s'éloigne du paradigme de la boîte noire de la tomographie pour se rapprocher de la réalité expérimentale.

La fidélité entre deux états purs est définie par

$$F\left(|\psi\rangle, |\phi\rangle\right) \equiv |\langle\psi|\phi\rangle|^2 \tag{3.1}$$

qu'on peut interpréter comme la probabilité de mesurer $|\psi\rangle$ dans l'état $|\phi\rangle$.

1. En mécanique quantique, la matrice densité représente *toute* l'information possible sur l'état.

Dans le contexte de la comparaison entre un état cible et un état expérimental, seul l'état cible est pur et la fidélité s'écrit [2]

$$F(\hat{\rho}, \hat{\sigma}) = \langle\psi|\hat{\sigma}|\psi\rangle = \mathrm{Tr}\,[\hat{\rho}\hat{\sigma}] \tag{3.2}$$

qui n'est autre que le produit scalaire de Hilbert-Schmidt entre les deux matrices densité.

La certification diffère fortement de la tomographie car elle n'estime qu'un seul paramètre (la fidélité) plutôt que de reconstruire les 4^n coefficients de la matrice densité. Ainsi, il y a espoir que la certification contourne le coût exponentiel de la tomographie.

3.1.2 Exigences expérimentales

Nous cherchons à proposer un protocole pratique, c.à.d. qui soit réaliste d'un point de vue expérimental. Ainsi, il ne suffit pas que les observables obéissent aux contraintes de la mécanique quantique, il faut qu'il soit possible de les mesurer dans le laboratoire.

Si l'on ne tient pas compte de ces restrictions, le problème de la certification devient trivial. En effet, on pourrait considérer que la fidélité n'est autre que la valeur moyenne de l'observable $\hat{\rho} = |\psi\rangle\langle\psi|$ sur l'état expérimental $\hat{\sigma}$. Or, cette observable est d'autant plus difficile à mesurer que $|\psi\rangle$ est intriqué. Il n'est donc pas réaliste expérimentalement de la mesurer. Notre protocole utilisera donc exclusivement des opérateurs de Pauli (cf. 2.2.2.2) dont nous avons argumenté qu'elles sont raisonnables expérimentalement et disponibles pour la plupart des technologies quantiques. Ainsi, notre protocole n'exige qu'une boîte à outil dont dispose déjà les expérimentateurs.

Dans la suite, nous nous concentrerons sur le protocole pour les particules quantiques de dimension finie, mais le protocole peut aussi être utilisé pour les systèmes de dimension infinie, par exemple pour caractériser l'état d'un résonateur. Cette extension est discutée dans l'article.

Nous montrerons que l'efficacité de notre protocole dépend fortement de la structure de l'état cible. En particulier, il demande très peu de ressources pour plusieurs classes d'états particulièrement intéressantes pour l'informatique quantique, par exemple les états stabilisateurs (définis en 2.3.1.2). Les états que les expérimentateurs cherchent à préparer ont typiquement beaucoup de structure et appartiennent souvent aux classes pour lequel le protocole est efficace.

2. Si aucun des états n'est pur, la fidélité prend une forme plus compliquée $F(\hat{\rho}, \hat{\sigma}) = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\hat{\rho}}\hat{\sigma}\sqrt{\hat{\rho}}}\right]\right)^2$.

3.2 Article : « Practical characterization of quantum devices without tomography »

Dans cette section, nous reproduisons l'article

Practical characteriztion of quantum devices without tomography
Marcus P. da Silva, Olivier Landon-Cardinal, David Poulin.
Physical Review Letters, **107**, 210404 (2011)

3.2.1 Genèse et contribution

Ma contribution à l'article concerne surtout le travail technique sur l'échantillonnage et l'analyse d'erreur. Avec Marcus da Silva, j'ai exploré les techniques disponibles pour l'échantillonnage, à commencer par l'algorithme de Metropolis-Hastings [52] qui fonctionne pour n'importe quelle classe d'état mais ne fournit pas de garanties théoriques. J'ai ensuite développé seul un algorithme d'échantillonnage dédié aux MPS, ce qui m'a amené à concevoir l'échantillonnage séquentiel et à formuler les bornes théoriques sur l'échantillonnage. L'analyse d'erreur a été effectuée conjointement avec Marcus ainsi que la borne sur les ρ_i négligeables. Finalement, mes simulations MPS ont permis de mettre le doigt sur la dépendance du nombre de mesures répétées avec les valeurs moyennes de l'état-cible ($N_2^{[k]} \sim \rho_{i_k}^{-2}$), une contrainte importante sur le protocole que nous n'avions pas identifiée à prime abord. Durant tout le projet, la réflexion générale ainsi que l'organisation de la recherche a été une collaboration entre Marcus, David et moi.

J'ai rédigé deux des trois sections de l'annexe technique de l'article, intitulé « Practical characterization of quantum devices without tomography : Supplemental material ». Il s'agit des sections S1. « Statistical bound for Monte-Carlo estimation of the fidelity » et S2. « Sampling from the relevance distribution ».

L'article comprend une section sur l'apprentissage d'un hamiltonien (ou d'un Linbladien) qui génère l'évolution d'un système. Puisque je n'ai pas contribué à cette section, elle n'est pas mentionnée dans cette thèse.

3.2.2 Article

Practical characterization of quantum devices without tomography

Marcus P. da Silva,[1,2] Olivier Landon-Cardinal,[2] and David Poulin[2]

[1]*Raytheon BBN Technologies, Disruptive Information Processing Group, Cambridge, Massachusetts, 02138, USA*
[2]*Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada*

(Dated: 16 November 2011)

Quantum tomography is the main method used to assess the quality of quantum information processing devices. However, the number of experimental settings and the data processing time required to extract complete information about a device via tomography grows exponentially with the device size. Part of the problem is that tomography generates much more information than is usually sought. Taking a more targeted approach, we develop schemes that enable (i) estimating the fidelity of an experiment to a theoretical ideal description, (ii) learning which description within a reduced subset best matches the experimental data. Both these approaches yield a significant reduction in resources compared to tomography. In particular, we demonstrate that fidelity can be estimated from a number of simple experiments that is independent of the system size, removing an important roadblock for the experimental study of larger quantum information processing units.

The building blocks for quantum computers have been demonstrated in a number of different physical systems [1–6]. In order to quantify how closely these demonstrations come to the ideal operations, the experiments are fully characterized via either *quantum state tomography* [7] or *quantum process tomography* [8]. An important advantage of these methods is that they require only simple local measurements. The main drawbacks however are that tomography fundamentally requires both experimental and data post-processing resources that increase exponentially with the number of particles n [9].

It is important to realize that the exponential cost of tomography is not a problem restricted to a large number of qubits. For example, recent ion trap experiments characterizing an 8 qubit state required 10 hours of measurements, despite collecting only 100 samples per observable [3]. Surprisingly, the post-processing of the data obtained from these experiments took approximately a week [10]. Under similar time scales, the characterization of a 16 qubit state would take years of measurements, and over a century of data post-processing. This is clearly a major obstacle in the demonstration of working quantum computers, even at sizes moderately larger than what has been demonstrated to date.

Moreover, one of the key assumptions for the fault-tolerance theorems of quantum computation is that the noise on elementary components does not scale badly with the system size [11]. Therefore, despite the fact that universal quantum computation can be realized with one- and two-qubit elementary operations, it is not sufficient to characterize small gates—larger systems may have significant noise contributions from correlated sources as seen in recent experiments [6]. The characterization of multi-qubit states and operations provides crucial information for the verification of these assumptions, and therefore the development of large quantum information processors.

Part of the problem with the usual approach is that tomography often provides more information than what is truly sought. Given an experiment that prepares a quantum state represented by a density operator $\hat{\sigma}$, one usually extracts a complete description for $\hat{\sigma}$ via quantum tomography, and then compares this description to a theoretical state $\hat{\rho}$ by computing the fidelity $F(\hat{\rho}, \hat{\sigma})$—a single number, commonly used as similarity measure. As this example illustrates, we often have an idea of what has been realized in the laboratory, so we are interested in asking for much less information—*e.g.*, we only want to know the distance to some particular theoretical target or to learn the identity of the state or operation within a restricted set of possibilities.

In this Letter, we develop targeted approaches to directly extract the information of interest. Our main results, summarized at Table I, show that it is possible to efficiently characterize a large class of states and operations—including some that are universal resources for quantum computation—without resorting to tomography and using only local measurements and the preparation of product states. Our methods apply to discrete variable systems such as qubits, as well as continuous variable systems such as oscillators. We consider two types of characterization: *certification* and *learning*.

Learning consists of identifying the theoretical description from a restricted set of possibilities that best matches the experimental data. There exists many classes of "variational" states in physics that can be specified with a small number of parameters. We provide examples where these parameters can be extracted directly from experiments, circumventing tomography and hence drastically reducing the complexity.

Certification consists of estimating the fidelity between an experimental device and some theoretical target. We demonstrate that certification always requires drastically less resources than full tomography—in some important cases, it is an exponential reduction in resources. Even in the worst case, our scheme offers four significant advantages for the characterization of quantum states (equivalent statements hold for quantum operations): *(1)* Its computational cost is bounded by $n^2 4^n$, compared to 4^{3n} required for the simplest tomography procedure based on pseudo-inverses. *(2)* The number of distinct experimental settings it requires is constant—independent of the system size and depending only on the desired accuracy of the estimate—compared to the 4^n distinct experiments needed by tomography, or the $\mathcal{O}(n2^n)$ settings required by compressed sensing techniques [12]. *(3)* The total number of measurements (counting repeated measure-

		Certification		Learning
		Sampling (C1)	Fluctuations (C2)	
States	Stabilizer	$\mathcal{O}(n)$	$\mathcal{O}(1)$	poly(n)
	W	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$
	$\|t_n\rangle$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$
	General MPS	$\mathcal{O}(n)$	?	$\mathcal{O}(n)$ [13]
	General pure state	$\mathcal{O}(n^2 2^{2n})$	$\mathcal{O}(2^n)$	$\mathcal{O}(2^{6n})$
Processes	Clifford	$\mathcal{O}(1)$	$\mathcal{O}(1)$	poly(n)
	MPS Choi matrix	$\mathcal{O}(n)$	?	$\mathcal{O}(n)$
	General unitary	$\mathcal{O}(n^2 4^{4n})$	$\mathcal{O}(2^{2n})$	$\mathcal{O}(2^{12n})$
Evolution	Local Hamiltonian	—	—	$\mathcal{O}(n)$
	Local Lindbladian	—	—	$\mathcal{O}(n)$

TABLE I. Complexity of the characterization of various states and processes. Entries in red are efficient, *i.e.* require resources that grow at most polynomially with the number of qubits n. The **Sampling** column gives the complexity of the classical processing required to sample from the relevance distribution, *c.f.* **C1**. The **Fluctuations** column gives the number of measurements required to suppress statistical fluctuations when evaluating the fidelity, *c.f.* **C2**. The **Learning** column gives the total number of measurements (including repetitions of the same measurement setting) required to learn the state within a restricted set; the classical processing is always a polynomial of that number. When both fidelity estimate and learning are efficient, it is not necessary to assume that the state belongs to a restricted set as fidelity testifies of that assumption. Stabilizer states, Clifford gates, Local Hamiltonians and Lindbladians are discussed in the main text. The W state has often been used as an experimental benchmark, *e.g.* [3]. The $|t_n\rangle$ state plays a key role in linear optics quantum computation [14]. Matrix product states (MPS) accurately describe ground states of 1D quantum systems [15]. An important example of a process with MPS Choi matrix is the approximate quantum Fourier transform [16], key component of Shor's factoring algorithm. Question marks indicate open problems, but they can be no worst than the general states and operations.

ments used to statistically estimate expectation values) of our scheme is bounded by $\mathcal{O}(2^n)$, which is at least a quadratic improvement over what is required by full tomography. *(4)* The data post-processing of our scheme is trivial, while the correct method of processing tomography data is a matter of current debates and different methods produce significantly different results [10].

The rest of this Letter is structured as follows. In the next three sections, we describe the state certification scheme for qubits, show how it extends to continuous variable systems, and the certification of quantum processes. Then, we present concrete examples drawn from Table I.

Monte Carlo state certification—To estimate the fidelity to some theoretical pure state $\hat{\rho}$, we use the fidelity

$$F(\hat{\rho},\hat{\sigma}) = \operatorname{tr}\hat{\rho}\hat{\sigma} = \sum_i \frac{\rho_i\sigma_i}{d} = \sum_i \frac{\rho_i^2}{d}\frac{\sigma_i}{\rho_i}. \qquad (1)$$

where $\rho_i = \operatorname{tr}\hat{\rho}\hat{P}_i$, $\sigma_i = \operatorname{tr}\hat{\sigma}\hat{P}_i$, d is the dimension of the Hilbert space, and $\hat{P}_i$ is some orthonormal Hermitian operator basis satisfying $\operatorname{tr}\hat{P}_i\hat{P}_j = d\delta_{ij}$. For a system composed of n qubits, the $\hat{P}_i$ could be the 4^n Pauli operators obtained by taking tensor products of the Pauli matrices and the identity. Defining the *relevance distribution* $\Pr(i) = \frac{\rho_i^2}{d}$, we can

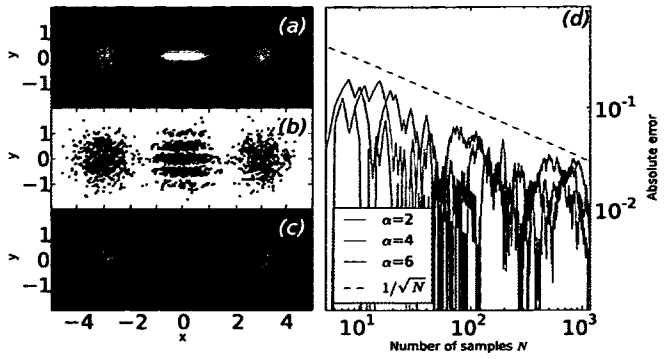


FIG. 1. (a) Wigner function representation of a harmonic oscillator in the superposition $|\psi\rangle = |\alpha\rangle + |-\alpha\rangle$ for $\alpha = 3$, (b) 10^3 samples of points in the complex plane drawn according to the relevance density of $|\psi\rangle$, (c) Wigner function representation of a harmonic oscillator in the incoherent mixture of $|\alpha\rangle$ and $|-\alpha\rangle$, corresponding to the preparation of $\hat{\sigma}$, (d) absolute error in successive estimates of the fidelity $F(\hat{\rho},\hat{\sigma})$ for 5 different runs with 10^3 samples each.

rewrite the fidelity as $F(\hat{\sigma},\hat{\rho}) = \sum_i \Pr(i)\frac{\sigma_i}{\rho_i}$, where the sum is taken over only the i with $\rho_i \neq 0$. This expression leads to an experimental procedure to estimate the fidelity based on Monte Carlo methods as follows: one generates N random indices $i_1, i_2, \ldots, i_N$ following the relevance distribution $\Pr(i)$ and estimates $\sigma_{i_k} = \langle\hat{P}_{i_k}\rangle_{\hat{\sigma}}$, the experimental expectation value of the observable $\hat{P}_{i_k}$. With high probability, the fidelity is close to $\frac{1}{N}\sum_{k=1}^{N}\frac{\sigma_{i_k}}{\rho_{i_k}}$ with an uncertainty that decreases as $\frac{1}{\sqrt{N}}$. The total number of distinct experimental settings is at most N, independent of the system size.

There are two important caveats to this technique:

C1 Generating an index i according to the relevance distribution $\Pr(i)$ can in general require an exponential amount of computational resources.

C2 Each σ_{i_k} is estimated within some finite accuracy. To estimate the fidelity with accuracy ϵ therefore requires repeating the measurement of P_{i_k} roughly $(\epsilon\rho_{i_k})^{-2}$ times, which in the worst case grows exponentially with the number of qubits.

These are important limitations, and as a consequence our method will not scale polynomially for all quantum states and operations, but nevertheless always does significantly better than tomography. In addition, there are important classes of states and operations which avoid these two problems (see Table I and the Supplemental Material for complete details).

Continuous variables systems— For infinite dimensional systems, such as a harmonic oscillator or a single optical mode in a cavity, it is more convenient to describe a state $\hat{\rho}$ by its Wigner functions $W_{\hat{\rho}}(\alpha)$ [17] (other indicator functions could also be used). Equation (1) becomes

$$F(\hat{\rho},\hat{\sigma}) = \frac{1}{\pi}\int_{\mathbb{C}} d^2\alpha\, p(\alpha)\frac{W_{\hat{\sigma}}(\alpha)}{W_{\hat{\rho}}(\alpha)} \qquad (2)$$

where the relevance density $p(\alpha) = W_{\hat{\rho}}^2(\alpha)$ is defined as the square of the Wigner function of the theoretical state, whose purity guarantees once again that $p(\alpha)$ is well defined as a probability density. The Wigner function of the experimental state $\hat{\sigma}$ can be measured by interactions with an atom and measurements of the atom's state [18]. Points in the complex plane can be selected according to $p(\alpha)$ using simple methods such as rejection sampling. As an example, we simulated this proposed method to estimate the fidelity between a quantum superposition of two harmonic oscillator states—a "cat" state $\frac{1}{\sqrt{2}}(|\alpha\rangle + |-\alpha\rangle)$—and the probabilistic mixture of those two classical states. For the given choice of parameters, this fidelity is $1/2(1 + e^{-2\alpha^2}) \approx 0.5$, and Fig. 1 clearly demonstrates a close agreement between the Monte Carlo estimate and the exact theoretical value, as the absolute error decreases like the square-root of the number of samples of the Wigner function. As expected, the error in the fidelity estimate does not depend on the state itself (e.g. average number of photons, amplitude, etc.) but only on the number of samples. We emphasize once again that no estimate of the Wigner function of the experimental state is ever made, so there is no need for maximum-likelihood fits to the data, or Radon transforms.

Monte Carlo process certification—The Choi-Jamiołkowski isomorphism [19] associates to every quantum operation $\mathcal{E}$ on a d-dimensional space a density operator $\hat{\rho}_{\mathcal{E}}$ on a d^2-dimensional space via $\hat{\rho}_{\mathcal{E}} = (\mathrm{id} \otimes \mathcal{E})(|\phi\rangle\langle\phi|)$ where $|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle \otimes |i\rangle$ and id is the identity operation. As with state certification, our goal is to compare a target unitary $\mathcal{U}$ to its experimental realization $\widetilde{\mathcal{U}}$. A good figure of merit in that case is the average output fidelity $\overline{F}(\mathcal{U}, \widetilde{\mathcal{U}})$, defined as the fidelity between the output states produced by $\mathcal{U}$ and $\widetilde{\mathcal{U}}$, averaged uniformly over all pure input states. It can be shown that $\overline{F}(\mathcal{U}, \widetilde{\mathcal{U}}) = \frac{d\, F(\hat{\rho}_{\mathcal{U}}, \hat{\rho}_{\widetilde{\mathcal{U}}}) + 1}{d+1}$ [20], reducing the problem of comparing two processes $\mathcal{U}$ and $\widetilde{\mathcal{U}}$ to the problem of comparing two states $\hat{\rho}_{\mathcal{U}}$ and $\hat{\rho}_{\widetilde{\mathcal{U}}}$. This problem is solved by the Monte Carlo state certification presented above.

While this derivation makes use of the maximally entangled state $|\phi\rangle$, the experimental realization of the protocol requires only the preparation of product states. A direct implementation of the quantum Monte Carlo state certification would prepare a maximally entangled state $|\phi\rangle$, apply $\widetilde{\mathcal{U}}$ to half of the system, and then measure random Pauli operators on all qubits. A more practical approach consists of preparing the complex conjugate of random product of eigenstates of local Pauli operators (corresponding to the resulting state after half of the entangled state is measured destructively), applying the transformation $\widetilde{\mathcal{U}}$ to the system, and finally measuring a random Pauli operator on each qubit. This simplification, based on the identity $(|\mu\rangle\langle\mu| \otimes \mathrm{id})|\phi\rangle = |\mu\rangle \otimes |\mu^*\rangle$, generates the same statistics as the direct scheme [21].

Computation via teleportation—Some of the most promising approaches to universal and scalable quantum computation are teleportation-based quantum computation [22] and measurement-based quantum computation [23]. Both these approaches rely heavily on the preparation of stabilizer states [24] and the application of quantum operations known as the Clifford group [22], which map stabilizer states to stabilizer states. Stabilizer states are also important for quantum computation in general because of their close relationship to a large class of quantum error correction codes known as *stabilizer codes*. Many of the experimental demonstrations of state preparation to date have been of stabilizer states, such as states encoded into stabilizer codes [2], cluster states [4], and the GHZ state $|00\cdots0\rangle + |11\cdots1\rangle$ [5, 6].

We first describe how to *certify* these states and operations. Stabilizer states are defined to be $+1$ eigenstates of some set of commuting Pauli operators $\hat{S}_j$ that generate the stabilizer group, *i.e.* $\hat{S}_j|\psi\rangle = |\psi\rangle$ for all $j = 1, \ldots n$. It follows that $\Pr(i) = 1/d$ if either of $\pm\hat{P}_i$ is in the stabilizer group and 0 otherwise. Sampling from $\Pr(i)$ thus amounts to generating an index i uniformly between 1 and d, avoiding the problem associated with caveat **C1**. For the same reasons, $\rho_i^2 = 1$ for all i with $\Pr(i) \neq 0$, so that the uncertainty in the estimation of σ_i is not amplified, avoiding the problem associated with caveat **C2**. It also follows that the fidelity $F(\hat{\sigma}, \hat{\rho})$ to a stabilizer state $\hat{\rho}$ can be estimated with error ϵ using $N = \mathcal{O}(\frac{1}{\epsilon^4})$ experiments involving only local projective measurements, *independently of the system size and without any prior knowledge of the experimental state* $\hat{\sigma}$. Since this result relies only on local measurements, it can immediately be generalized to states which are locally equivalent to stabilizer states.

This result carries over directly to the certification of Clifford operation because their Choi-Jamiołkowski density operators are stabilizer states. In the case of Clifford transformations similar results can be obtained using "twirling" experiments [25] or by the selective measurement of matrix elements of the Choi matrix [21], although the Monte Carlo approach described here generalizes to other cases.

While operations in the Clifford group are not sufficient to perform universal computation [22], single qubit rotations can be used to reach universality, and these can be certified efficiently thanks to local equivalence of either operations (if the rotation is applied directly) or state preparation (if the rotation is applied via "magic state" teleportation [22, 26]).

Stabilizer states can also be *learned* efficiently, as pointed out by Aaronson and Gottesman [27], although the known method for efficient stabilizer learning requires entangling measurements. Aside from the direct generalizatin of the stabilizer approach, Clifford group operations can be learned efficiently [28] if one has access to Bell measurements and the inverse of the operation being learned. The problem of performing these tasks efficiently with strictly local measurements and without the need for the inverse remains open.

Local Hamiltonians and Lindbladians—Models of universal quantum computation exist where the idea of discrete gates is not a natural fit. Instead, the system evolves in a continuous way, governed by some dynamical equation $\frac{\partial}{\partial t}\hat{\rho} = \mathcal{G}\hat{\rho}$. The most direct way to determine how accurately these dynamics can be realize is to estimate the time evolution generator $\mathcal{G}$ of the system, and explicitly check how it compares

against the ideal target generator. Important examples include local Hamiltonians and Lindbladians that are universal for adiabatic quantum computation [29] and dissipation-driven quantum computation [30] respectively.

In what follows we demonstrate how to learn such local $\mathcal{G}$ using only *(i)* the preparation of initial product states, *(ii)* the simultaneous measurement of a constant number of single-qubit operator, *(iii)* a number of experimental settings that grows linearly with the system size, *(iv)* and classical post-processing of complexity n^3 (inverting an $cn \times cn$ matrix for some constant c); improving on [31].

Consider the case of coherent evolution generated by some Hamiltonian H. For a short time t, the expectation value of any observable $\hat{A}$ evolves as

$$\langle \hat{A}(t) \rangle_{\hat{\rho}} - \operatorname{tr} \hat{A}\hat{\rho} = it\langle [\hat{H}, \hat{A}] \rangle_{\hat{\rho}} + \mathcal{O}(\|\hat{H}\|^2 t^2). \quad (3)$$

By experimentally measuring this expectation value, we obtain one linear constraint on the Hamiltonian. Varying over different observables $\hat{A}_i$ and initial states $\hat{\rho}_j$, we obtain more linear constraints that we can write as $W_{ij} = \langle \hat{A}_i(t) \rangle_{\hat{\rho}_j} - \operatorname{tr} \hat{A}_i \hat{\rho}_j = it\langle [\hat{H}, \hat{A}_i] \rangle_{\hat{\rho}_j}$ where we have dropped the higher order terms $\mathcal{O}(\|\hat{H}\|^2 t^2)$. Writing $\hat{H}$ in an operator basis $\hat{H} = \sum_l h_l \hat{P}_l$, we obtain the linear equation $W_{ij} = \sum_l T_{ij,l} h_l$ where $T_{ij,l} = it \operatorname{tr} \hat{\rho}_j [\hat{P}_l, \hat{A}_i]$. The Hamiltonian can be learned by inverting this linear equation [31].

There are in general a number important caveats to this approach, although all of these disappear when the Hamiltonian is *local*, which is nonetheless sufficient to achieve universal quantum computation [29, 30]. The Lieb-Robinson bound [32] shows that only the Hamiltonian $\hat{H}_R$ in a region R a distance $d \approx vt$ of the local observable $\hat{A}$ contributes to its evolution, *i.e.*, $e^{i\hat{H}t}\hat{A}e^{-i\hat{H}t} \approx e^{i\hat{H}_R t}\hat{A}e^{-i\hat{H}_R t}$ (for details of the proof see the Supplemental Material). This fact solves all the problems associated to the proposal of [31]:

1) The error $\mathcal{O}(\|H\|^2 t^2)$ appearing in Eq. (3) becomes $\mathcal{O}(\|H_R\|^2 t^2) = \mathcal{O}(\|\hat{A}\|^2 t^4)$, independent of the system size. Thus, it is not necessary to decrease the evolution time t as the system size increases to achieve a given accuracy.

2) Because the Hamiltonian is local, the number of non-zero terms h_l is proportional to the number of particles in any finite dimension. Thus, in the linear equation for W_{ij}, the range of the index l increases only linearly with the number of particles, as opposed to the exponential growth for generic Hamiltonians.

3) Because the dynamics is local, $T_{ij,l} = T_{ij',l}$ when $\hat{\rho}_j$ and $\hat{\rho}_{j'}$ differ only outside a region of radius k away from the local observable $\hat{A}_i$. In addition, the T become linearly dependent—and thus redundant—when the input states are linearly dependent. For each observable $\hat{A}_i$, we only need to vary the initial state locally, so the total number of observable-state pairs (ij) grows linearly with the number of particles. Thus, learning the Hamiltonian—or equivalently the h_l—amounts to inverting the linear-size linear equation $W_{ij} = \sum_l T_{ij,l} h_l$.

4) Product input states form a complete operator basis, so they are sufficient to gain all information about the Hamiltonian. Thus $\operatorname{tr} \hat{A}_i \hat{\rho}_j$ can be easily computed since $\hat{A}_i$ is local and $\hat{\rho}_j$ is a product state. The quantity $\operatorname{tr} \hat{\rho}_j [\hat{P}_l, \hat{A}_i]$ can also be evaluated efficiently because the commutator of two k-local operators is at most $2k$-local, and $\hat{\rho}_j$ is a product state.

Acknowledgments— We thank P. Sémon for many instructive discussions about Monte Carlo methods. After this work was made public, some of our findings were independently derived by Flammia and Liu [33]. This work is partially funded by FQRNT, NSERC, and numerical calculations were performed using resources from RQCHP.

[1] T. Yamamoto, Y. A. Pashkin, O. Astafiev, Y. Nakamura, and J. S. Tsai, Nature, **425**, 941 (2003)
[2] J. Chiaverini, D. Leibfried, T. Schaetz, M. D. Barrett, R. B. Blakestad, J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, R. Ozeri, and D. J. Wineland, Nature, **432**, 602 (2004)
[3] H. Haffner, W. Hansel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Korber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Guhne, W. Dur, and R. Blatt, Nature, **438**, 643 (2005)
[4] P. Walther, K. J. Resch, T. Rudolph, E. Schenck, H. Weinfurter, V. Vedral, M. Aspelmeyer, and A. Zeilinger, Nature, **434**, 169 (2005)
[5] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland, Nature, **438**, 639 (2005)
[6] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett., **106**, 130506 (2011)
[7] K. Vogel and H. Risken, Phys. Rev. A, **40**, 2847 (1989)
[8] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett., **78**, 390 (1997)
[9] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A, **77**, 032322 (2008)
[10] R. Blume-Kohout, New J. Phys., **12**, 043034 (2010)
[11] P. Aliferis, D. Gottesman, and J. Preskill, Quant. Inf. Comput., **6**, 97 (2006)
[12] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett., **105**, 150401 (2010)
[13] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nature Comm., **1**, 149 (2010)
[14] E. Knill, R. Laflamme, and G. J. Milburn, Nature, **409**, 46 (2001)
[15] F. Verstraete and J. I. Cirac, Phys. Rev. B, **73**, 094423 (2006)
[16] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä, Phys. Rev. A, **54**, 139 (1996)
[17] E. P. Wigner, Phys. Rev., **40**, 749 (1932)
[18] D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett., **77**, 4281 (1996)
[19] A. Jamiołkowski, Rep. Math. Phys., **3**, 275 (1972)
[20] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A, **60**, 1888 (1999)
[21] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett., **100**, 190403 (2008)

[22] D. Gottesman and I. L. Chuang, Nature, **402**, 390 (1999)
[23] R. Raussendorf and H. J. Briegel, Phys. Rev. Lett., **86**, 5188 (2001)
[24] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, California Institute of Technology (1997)
[25] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Science, **317**, 1893 (2007)
[26] S. Bravyi and A. Kitaev, Phys. Rev. A, **71**, 022316 (2005)
[27] D. Gottesman, "Identifying stabilizer states," (2008), http://pirsa.org/08080052/
[28] R. A. Low, Phys. Rev. A, **80**, 052314 (2009)
[29] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, and S. Lloyd, SIAM J. on Computing, **37**, 166 (2007)
[30] F. Verstrate, M. M. Wolf, and J. I. Cirac, Nature Physics, **5**, 633 (2009)
[31] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, Phys. Rev. A, **84**, 012107 (2011)
[32] E. H. Lieb and D. W. Robinson, Commun. Math. Phys., **28**, 251 (1972)
[33] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett., **106**, 230501 (2011)

Practical characterization of quantum devices without tomography : Supplemental Material

Marcus P. da Silva,[1,2] Olivier Landon-Cardinal,[2] and David Poulin[2]

[1]*Raytheon BBN Technologies, Cambridge, Massachusetts, USA*
[2]*Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, Canada*

(Dated: 16 November 2011)

S1. STATISTICAL BOUND FOR MONTE-CARLO ESTIMATION OF THE FIDELITY

We present rigorous bounds for the error of the Monte Carlo fidelity estimate in the case of an n qubit system. The result (*c.f.* Eq. (S1)) can be adapted to the case of continuous variable systems through the minor modification of replacing the expectation value ρ_{i_k} by the value $\frac{1}{2}W_{\hat{\rho}}(\alpha_{i_k})$ of the Wigner function of state $\hat{\rho}$ at point α_{i_k}.

Theorem 1. *Let $\hat{\rho} = \sum_i \frac{\rho_i}{d}\hat{P}_i$ be the decomposition of the pure state $\hat{\rho}$ over the orthogonal Hermitian operator basis $\{\hat{P}_i\}$ where $\operatorname{tr}\hat{P}_i\hat{P}_j = d\delta_{ij}$ and the operator norm $\|\hat{P}_i\| \leq 1$. One can obtain an estimate $\bar{F}$ of the fidelity $F(\hat{\rho},\hat{\sigma})$ between $\hat{\rho}$ and $\hat{\sigma}$ with error $\epsilon = \epsilon_1 + \epsilon_2$ such that*

$$\Pr(|F-\bar{F}| \geq \epsilon) \leq \frac{1}{N_1\epsilon_1^2} + 2\exp\left[-\frac{\epsilon_2^2 N_1^2}{2}\left(\sum_{k=1}^{N_1}\frac{1}{\rho_{i_k}^2 N_2^{[k]}}\right)^{-1}\right] \quad \text{(S1)}$$

where

- *$I = \{i_1 \dots i_{N_1}\}$ are N_1 indices sampled from $\Pr(i)$, corresponding to observables $\hat{P}_{i_k}$ to be measured experimentally on $\hat{\sigma}$*
- *$N_2^{[k]}$ is the number of experimental samples taken to estimate $\sigma_{i_k} = \operatorname{tr}\hat{P}_{i_k}\hat{\sigma}$*
- *ϵ_1 is the error associated to the Monte Carlo estimate*
- *ϵ_2 is the error associated to the experimental estimation of the $\{\sigma_i\}_{i\in I}$*

Proof. The fidelity $F(\hat{\rho},\hat{\sigma})$ can be rewritten as

$$F(\hat{\rho},\hat{\sigma}) = {\sum_i}' \frac{\rho_i^2}{d}\frac{\sigma_i}{\rho_i} \quad \text{(S2)}$$

where prime indicates that the summation runs only over non-zero values of ρ_i. Since $\operatorname{tr}\hat{\rho}^2 = 1$ by assumption, $\Pr(i) = \rho_i^2/d$ is a normalized probability distribution. We can thus interpret the fidelity as the expectation value of a random variable X which takes value σ_i/ρ_i with probability $\Pr(i)$. Its variance is bounded by a constant, as

$$\operatorname{Var}(X) = {\sum_i}' \frac{\sigma_i^2}{d} - F^2 \leq \operatorname{tr}\hat{\sigma}^2 - F^2 \leq 1, \quad \text{(S3)}$$

and thus, using Chebyshev's inequality, we obtain

$$\Pr(|F-\bar{F}_1| \geq \epsilon_1) \leq \frac{1}{N_1\epsilon_1^2}, \quad \text{(S4)}$$

where $\bar{F}_1 = \sum_{i\in I}\sigma_i/\rho_i$ is the estimate of the fidelity by sampling N_1 realizations of X, *i.e.*, by drawing $I = \{i_1 \dots i_{N_1}\}$ indexes from the probability distribution $\Pr(i)$ and estimating $\mathbb{E}(X)$ by the realization of $\bar{X} = N_1^{-1}\sum_{i\in I} X_i$ where all X_i are independent and distributed as X. Thus, the number of measurements settings does not depend on the dimension of the system and scales as $\mathcal{O}(1/\epsilon_1^2)$.

The expectation value σ_i of each observables with respect to the experimental state $\hat{\sigma}$ can only be estimated up to finite precision. For each $i_k \in I$, the observable $\hat{P}_{i_k}$ is measured on the experimental state and yields a number $y_{i_k}^{[m]}$ whose absolute value is bounded by the operator norm of the observables. This measurement is repeated $N_2^{[k]}$ times and the approximate realization of X_k is $\tilde{\sigma}_{i_k}/\rho_{i_k} = \left(\rho_{i_k}N_2^{[k]}\right)^{-1}\sum_{m=1}^{N_2^{[k]}} y_{i_k}^{[m]}$. This estimation proceadure is then repeated for each of the N_1 observables. Hoeffding's bound [1] states that, if the *independent* real random variables Y_i are such that $a_i \leq Y_i \leq b_i$, then for $S = Y_1 + Y_2 + \cdots + Y_n$,

$$\Pr(|S - \langle S\rangle| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right). \quad \text{(S5)}$$

In our case, for all k, we have $-1/|\rho_{i_k}| \leq y_{i_k}^{[k]}/\rho_{i_k} \leq 1/|\rho_{i_k}|$ for all $N_2^{[k]}$ experimental measurement performed to estimate σ_{i_k} and we can apply the Hoeffding's inequality to all $N = \sum_{k=1}^{N_1} N_2^{[k]}$ experimental samples to bound the distance between the sum $\bar{F}$ of $\tilde{\sigma}_{i_k}/\rho_{i_k}$ by

$$\Pr(|\bar{F}-\bar{F}_1| \geq \epsilon_2) \leq 2\exp\left[-\frac{\epsilon_2^2 N_1^2}{2}\left(\sum_{k=1}^{N_1}\frac{1}{\rho_{i_k}^2 N_2^{[k]}}\right)^{-1}\right]. \quad \text{(S6)}$$

Finally to reach eq. (S1), one applies the union bound to $|F-\bar{F}| \leq |F-\bar{F}_1| + |\bar{F}-\bar{F}_1|$. □

As can be seen in the last term of Eq. (S1), observable $\hat{P}_{i_k}$ must be sampled $N_2^{[k]} \gg \rho_{i_k}^2$ times to obtain an accurate estimate of its expectation value. While this can be large in general, there are many important cases where the ρ_i are only polynomially small, leading to a polynomial $N_2^{[k]}$. Two examples of cases of interest beyond the stabilizer states presented in main text are the W state [2] and the $|t_n\rangle$ state used in linear optics for heralded teleportation with high success probability

[3]. Both are MPS with bond dimension 2 and are uniform superpositions of a linear number of computational-basis states. For both states, the expectation value of a Pauli operator $\hat{P}$ is given by

$$\langle\psi|\hat{P}|\psi\rangle = \alpha(n)\sum_{i,j}\langle i|\hat{P}|j\rangle \tag{S7}$$

where $\alpha(n)$ is $1/n$ for the W state and $1/(n+1)$ for $|t_n\rangle$, and the sum runs over computational states that appear in the decomposition of the state. For all i, j, there exists a Pauli operator $\hat{\sigma}_{ij}$ such that $|j\rangle = \hat{\sigma}_{ij}|i\rangle$. Since the Pauli operators form a group, $\hat{P}\hat{\sigma}_{ij}$ is another Pauli operator and all terms appearing in the sums are ± 1. Thus, the smallest non-zero Pauli expectation scales as $1/n$, and the number of samples required to estimate σ_i/ρ_i to constant accuracy scales as n^2 in the worst case.

More generally, we can improve the error bound Eq. (S1) by truncating the relevance distribution. Define the set of negligible expectation values as $S \equiv \{\rho_i \text{ such that } |\rho_i| < d^{-\alpha}\}$ where α is a positive number to be determined. We split the fidelity into a significant and a negligible contribution

$$F(\hat{\rho},\hat{\sigma}) = \sum_i \frac{\rho_i\sigma_i}{d} = \sum_{\rho_i\notin S}\frac{\rho_i\sigma_i}{d} + \sum_{\rho_i\in S}\frac{\rho_i\sigma_i}{d} \tag{S8}$$

and bound the negligible contribution using

$$\left|\sum_{\rho_i\in S}\frac{\rho_i\sigma_i}{d}\right| \leq \sum_{\rho_i\in S}\frac{|\sigma_i|}{d}\max_{i\in S}|\rho_i| \leq d^{-(\alpha+1)}\sum_{\rho_i\in S}|\sigma_i|. \tag{S9}$$

The sum of a subset of $|\sigma_i|$ is bounded by the sum over *all* $|\sigma_i|$. To bound $\sum_i|\sigma_i|$, we can use the constraint on the purity of the state $\sum_i\sigma_i^2 = d\ \mathrm{tr}\,\hat{\sigma}^2 \leq d$. The sum of absolute values is maximal when all absolute values are equal, which follows from standard Lagrange multiplier techniques. The purity constraint finally leads to

$$\sum_i|\sigma_i| \leq d\sqrt{d\ \mathrm{tr}\,\hat{\sigma}^2} \leq d^{3/2}. \tag{S10}$$

Inserting this inequality that into eq. (S9) yields

$$\left|\sum_{\rho_i\in S}\frac{\rho_i\sigma_i}{d}\right| \leq d^{1/2-\alpha}. \tag{S11}$$

Hence, the sum over negligible ρ_i vanishes exponentially for $\alpha = (1+\epsilon)/2$, *i.e.*, when we drop all expectation values smaller than $d^{-\frac{1+\epsilon}{2}}$ in absolute value, for any constant $\epsilon > 0$.

We thus modify the sampling method in the following way. For each observable $\hat{P}_i$ picked from sampling the relevance distribution, compute the corresponding expectation value $\rho_i = \mathrm{tr}\,\hat{\rho}\hat{P}_i$. When ${\rho_i}^2 < d^{-1-\epsilon}$, reject this entry, otherwise you proceed as before. It is important to verify that this modification does not slow down the procedure, *i.e.* that we are not constantly rejecting samples. To see this, notice that the probability of choosing an element from the negligible set is bounded by

$$\sum_{\rho_i\in S}\frac{\rho_i^2}{d} \leq \sum_{\rho_i\in S}d^{-2-\epsilon} \leq d^{-\epsilon}. \tag{S12}$$

Since we reject all negligible ρ_i, the maximum number of repeated measurements needed for a given experimental setting scales in the worst case as $d^{1+\epsilon}$. In particular, for qubits, the maximum number of measurements is $2^{n(1+\epsilon)}$. Moreover, since the number of measurement settings does not scale with the size of the system, the total number of measurements scales as $\mathcal{O}(2^{n(1+\epsilon)})$ which is at least a *quadratic improvement over the number of measurements needed to perform brute-force tomography* on a generic state of n qubits.

Extension to continuous variables systems

The Monte Carlo method proposed here can be adapted to continuous variable systems, such as a single electromagnetic field mode in a cavity [4–6], by modifying how the state is parameterized and how the sampling is performed. The main reason for this is the obvious difficulty of measuring observables in a discrete infinite dimensional operator basis. This problem can be avoided by considering phase-space quasiprobability distribution descriptions of quantum states. If we consider the dual phase-space distributions $f_{\hat{\rho}}(\alpha)$ and $g_{\hat{P}_i}(\alpha)$ which correspond respectively to the quantum state $\hat{\rho}$ and an observable $\hat{P}_i$ [7], then $\mathrm{tr}\,\hat{\rho}\hat{P}_i = \frac{1}{\pi}\int_{\mathbb{C}} d^2\alpha\ f_{\hat{\rho}}(\alpha)g_{\hat{P}_i}(\alpha)$. It follows that the fidelity between a pure state $\hat{\rho}$ and an arbitrary state $\hat{\sigma}$ is given by $F(\hat{\rho},\hat{\sigma}) = \mathrm{tr}\,\hat{\rho}\,\hat{\sigma} = \frac{1}{\pi}\int_{\mathbb{C}}\ d^2\alpha\ f_{\hat{\rho}}(\alpha)\ g_{\hat{\sigma}}(\alpha)$, which can be re-written as $F(\hat{\rho},\hat{\sigma}) = \frac{1}{\pi}\int_{\mathbb{C}} d^2\alpha\ p(\alpha)\ \frac{g_{\hat{\sigma}}(\alpha)}{f_{\hat{\rho}}(\alpha)}$, where the integration excludes regions with $f_{\hat{\rho}}(\alpha) = 0$ and where $p(\alpha) = f_{\hat{\rho}}^2(\alpha)$ is the *relevance density function*. Sampling the relevance density can be done by standard methods, such as rejection sampling.

The choice of phase space distributions is important, as it must be possible to interpret $f_{\hat{\rho}}^2(\alpha)$ as probability distributions, and it must be possible to estimate $g_{\hat{\sigma}}(\alpha)$ at some arbitrary $\alpha \in \mathbb{C}$ easily from experimental data. One choice that fulfills both these requirements for all states is the Wigner function [7, 8]. The Wigner function is self-dual and bounded in magnitude by 2, and its value at particular α can be estimated by using simple experiments where the continuous variable system, such as an electromagnetic field mode, interacts with an atom [6, 9–11].

The same truncation technique used to evaluate the performance of this algorithm for qubits can be used for continuous variable systems. Amplification of experimental uncertainty can once again by reduced by placing a cut-off in the relevance density function. If we disregard regions in phase space where the absolute value of the relevance density is below c,

then the error E in the fidelity is bounded by

$$E = \frac{1}{\pi}\left|\int_I d^2\alpha\, W_{\hat{\rho}}(\alpha) W_{\hat{\sigma}}(\alpha)\right|, \tag{S13}$$

$$\leq \frac{1}{\pi}\sqrt{\int_I d^2\alpha\, W^2_{\hat{\rho}}(\alpha)} \tag{S14}$$

where I is the region in phase space where $|W_{\hat{\rho}}| < c$.

S2. SAMPLING FROM THE RELEVANCE DISTRIBUTION

Sampling from the relevance distribution $\Pr(i)$ is not trivial because the dimension of the operator space on n particles is exponentially large in n. Therefore, computing all $\rho_i = \operatorname{tr}\hat{\rho}\hat{P}_i$ for all observables $\hat{P}_i$ is unefficient. Furthermore, computing a given ρ_i can be a challenging task in itself. However, by choosing operators $\hat{P}_i = \hat{p}_i^{[1]} \otimes \ldots \otimes \hat{p}_i^{[n]}$ that are tensor products of single-particle operators—such as the Pauli operators for qubits—sampling can be simplified by recursively picking the observables for each particle as we now demonstrate.

A. Sampling using conditional probabilities

Consider for concreteness a system composed of n qubits, and an operator basis $\hat{P}_i$ all consisting of tensor product of single qubit operators, *e.g.* Pauli operators. The Hilbert space dimension is $d = 2^n$. For an observable $\hat{P}_i = \bigotimes_{m=1}^n \hat{p}_{i_m}^{[m]}$, denote the relevance distribution $\Pr(i) = q_{i_1,\ldots,i_n}$. Using the probability chain rule, this probability can be expressed as a product of conditional probabilities

$$q_{i_1,\ldots,i_n} = \prod_{k=1}^{n} q_{i_k|i_1,\ldots,i_{k-1}} \tag{S15}$$

where the conditional probability $q_{i_k|i_1,\ldots,i_{k-1}}$ of drawing the observable $\hat{p}_{i_k}^{[k]}$ on particle k knowing which observables have been picked on the previous particles is

$$q_{i_k|i_1,\ldots,i_{k-1}} = q^{-1}_{i_1,\ldots,i_{k-1}} \sum_{I=i_{k+1},\ldots,i_n} q_{i_1,\ldots,i_k I}. \tag{S16}$$

Using equation (S15), sampling from the probability distribution reduces to sequentially picking an observable $\hat{p}_{i_m}^{[m]}$ according to the conditional probability distribution (S16) which can be written, up to a normalization factor, as

$$q_{i_k|i_1,\ldots,i_{k-1}} \propto \sum_{\hat{P}\in\mathcal{P}_{n-k}} \operatorname{tr}\left[\left(\hat{\rho}\times\left(\bigotimes_{m=1}^{k}\hat{p}_{i_m}^{[m]}\otimes\hat{P}\right)\right)^{\otimes 2}\right],$$

$$= \operatorname{tr}\left[\hat{\rho}^{\otimes 2}\left(\bigotimes_{m=1}^{k}(\hat{p}_{i_m})^{\otimes 2}\otimes\sum_{\hat{P}\in\mathcal{P}_{n-k}}\hat{P}^{\otimes 2}\right)\right]$$

where the trace of two copies accounts for the square in the definition of $\Pr(i) = \frac{\operatorname{tr}(\hat{\rho}\hat{P}_i)^2}{d} = \frac{\operatorname{tr}(\hat{\rho}\otimes\hat{\rho}\hat{P}_i\otimes\hat{P}_i)}{d}$. The sum over all duplicated observables $\hat{P}^{\otimes 2}$ can be written as the tensor product of operators acting on each pair $[m,\, n+m]$ of particles

$$2^{-(n-k)}\sum_{\hat{P}\in\mathcal{P}_{n-k}}\hat{P}\otimes\hat{P} = \bigotimes_{m=k+1}^{n}\hat{\Omega}^{[m,n+m]} \tag{S17}$$

where $\hat{\Omega}^{[i,j]} = \frac{1}{2}\sum_m \hat{p}_m^{[i]}\otimes\hat{p}_m^{[j]}$ is an observable acting on the pair of particles (i,j). For instance, for the Pauli operator basis, $\hat{\Omega}$ is the SWAP operator. Thus, the conditional probability is proportionnal to

$$\operatorname{tr}\left[\hat{\rho}^{\otimes 2}\left(\bigotimes_{m=1}^{k}(\hat{p}_{i_m})^{\otimes 2}\bigotimes_{m=k+1}^{n}\hat{\Omega}^{[m,n+m]}\right)\right] \tag{S18}$$

which is the *expectation value of a tensor product of 2-local observables* on the state $\hat{\rho}\otimes\hat{\rho}$ on $2n$ particles.

B. Bound on the complexity of sampling

The problem of sampling reduces to, for each of the n particles, *i)* computing conditional probabilities for each of the possible observables acting on that particle *ii)* pick one of those observables by generating a random number. Conditional probabilities can be expressed as expectation values through eq. (S18). Thus, if computing expectation values on tensor product of local observables on states of n particles has complexity $q(n)$, generating an index $i = i_1 \ldots i_n$ from the relevance distribution $\Pr(i)$ has complexity at most $n \times q(2n)$.

For many states of interest, computing expectation values of local observables can be performed in polynomial time, *i.e.*, $q(n) \in \operatorname{poly}(n)$. That is the case for many families of tensor-network states such as matrix product states (MPS) [12] which are known to represent faithfully ground states of interesting many-body Hamiltonians in 1D [13]. In fact, the procedure outlined above can be simplified in the case of MPS, yielding a sampling complexity *linear* in n, see Fig. 1. Their natural extension to 2D, projected entangled pair states (PEPS) [14] also allows the efficient heuristic computation of such expectation values.

A larger class of multi-qubit states for which sampling can be done efficiently by computing conditional probabilities are computationally tractable (CT) states [15]. CT states are states in which *(a)* the overlap with any element of the computational basis can be computed efficiently, and *(b)* it is possible to sample from the distribution of outcomes from measurements in the computational basis efficiently. For such states, it is possible to efficiently compute the expectation value of tensor products of Pauli observables which only permute elements of the computational basis and thus are basis preserving.

In the generic case of a state defined as a vector of the Hilbert space, computing the expectation value of a single local observable will take time $\mathcal{O}(2^{2n})$ since we have to account for the Hilbert space of $2n$ qubits. A tensor product of local observables can be thought as the product of $\mathcal{O}(n)$ observables that act non-trivially on a few qubits. Thus, computing

the expectation value given by equation (S18) will take time $\mathcal{O}\left(n\,2^{2n}\right)$. In order to sample, such a computation has to be repeated for each particles, leading to an overall complexity of sampling from the relevance distribution of $\mathcal{O}\left(n^2\,2^{2n}\right)$ in the worst case. Learning algorithms based on compressed sensing can recover low-rank density matrices from $\mathcal{O}\left(n\,2^n\right)$ expectation values in any basis [16], which indicates that it may be possible to improve the performance of the algorithm proposed here in the case of general pure states.

S3. LIEB-ROBINSON BOUND

The characterization of local Hamiltonians and Lindbladians relies heavily on the Lieb-Robinson bound [17, 18] that shows that a local Hamiltonian generates a causal evolution, with effects propagating at a finite velocity v (note that this bound has been generalized to the setting of dissipative systems [19], so our derivation holds for local Lindbladians as well). A local Hamiltonians acting on n particles is of the form $\hat{H} = \sum_X \hat{H}_X$ where X labels subsets of n particles, each term has bounded norm $\|\hat{H}_X\| \leq E$, and acts on at most k neighboring particles, such that $H_X = 0$ when $|X| > k$. The evolution of an operator is governed by the equation $\frac{\partial}{\partial t}\hat{A}(t) = i[\hat{H}, \hat{A}]$. Break the Hamiltonian into $\hat{H} = \hat{H}_0 + \hat{H}_M$, where $\hat{H}_M$ contains all the terms $\hat{H}_X$ that intersect a membrane M surrounding the operator $\hat{A}$ (see Fig. 2). The idea of this membrane is to disconnect its interior, denoted region R, from the rest of the particles. Indeed, $e^{i\hat{H}_0 t}\hat{A}e^{-i\hat{H}_0 t} = e^{i\hat{H}_R t}\hat{A}e^{-i\hat{H}_R t}$ where $\hat{H}_R$ in the Hamiltonian acting only inside the membrane (see Fig. 2). The differential equation for $\hat{A}(t)$ is

$$\frac{\partial}{\partial t}\hat{A}(t) = i[\hat{H}_0, \hat{A}(t)] + i[\hat{H}_M, \hat{A}], \tag{S19}$$

which has solution

$$\begin{aligned}\hat{A}(t) =& e^{i\hat{H}_0 t}\hat{A}(0)e^{-i\hat{H}_0 t} \\ &+ i\int_0^t e^{i\hat{H}_M(t-s)}[\hat{H}_M, \hat{A}(s)]e^{-i\hat{H}_M(t-s)}ds \qquad \text{(S20)} \\ =& e^{i\hat{H}_R t}\hat{A}(0)e^{-i\hat{H}_R t} \\ &+ i\int_0^t e^{i\hat{H}_M(t-s)}[\hat{H}_M, \hat{A}(s)]e^{-i\hat{H}_M(t-s)}ds \qquad \text{(S21)}\end{aligned}$$

as can be verified directly by differentiation. The commutator appearing in the second term can be bounded by

$$\|[\hat{H}_M, \hat{A}(s)]\| \leq cV\|\hat{A}\|\|\hat{H}_M\|\exp\left(-\frac{d-vt}{\xi}\right) \tag{S22}$$

where V is the number of sites in the support of the observable $\hat{A}$, and c, v, and ξ are constant that depend only on the microscopic details of the system, independent of the system size. This is known as the the Lieb-Robinson bound. Integrating, we obtain

$$\begin{aligned}&\|\hat{A}(t) - e^{i\hat{H}_R t}\hat{A}(0)e^{-i\hat{H}_R t}\| \qquad \text{(S23)} \\ &\leq ctV\|\hat{A}\|\|\hat{H}_M\|\|\exp\left(-\frac{d-vt}{\xi}\right).\end{aligned}$$

Expanding the exponential to first order yields

$$\begin{aligned}&\|\hat{A}(t) - \hat{A}(0) - it[\hat{H}_R, A(0)]\| \qquad \text{(S24)} \\ &\leq ctV\|\hat{A}\|\|\hat{H}_M\|\|\exp\left(-\frac{d-vt}{\xi}\right) \\ &+ c'\|\hat{A}\|\|\hat{H}_R\|^2 t^2.\end{aligned}$$

Because $\hat{H}_R$ and $\hat{H}_M$ represent respectively the Hamiltonian of a ball of radius d and the Hamiltonian for a constant thickness membrane around that ball, they grow proportionally to d^D and d^{D-1} respectively, where D is the spatial dimension,

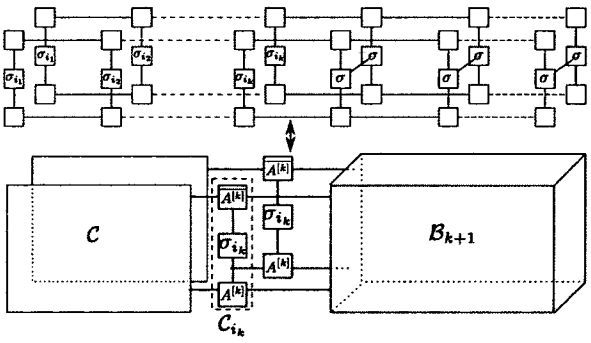


FIG. 1. Tensor network corresponding to eq. (S18) if $\hat{\rho} = |\psi\rangle\langle\psi|$ is a MPS, *i.e.*, there exist a familty of matrices $\left\{A^{[k]}_{i_k}\right\}$ such that $|\psi\rangle = A^{[1]}_{i_1}\ldots A^{[n]}_{i_n}|i_1\ldots i_n\rangle$. The upper figure represent the individual tensors in the tensor network. Each square represent a tensor and outgoing legs represent the tensor indices. Two squares connected by a line are the contraction of the corresponding indices of two tensors. Red squares correspond to the Ω operators. Orange squares correspond to the Pauli operators already chosen on the $k-1$ previous qubits. Blue squares are the MPS tensors of the two copies of $|\psi\rangle$ while the green squares are the MPS tensors of the two copies of $\langle\psi|$. The lower figure correspond to the partial contraction of the tensor network.

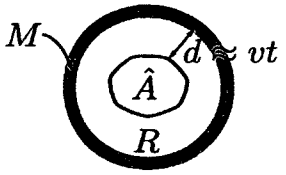


FIG. 2. When the system evolves under a *local* Hamiltonian (or Lindbladian), the operator A evolves under the full Hamiltonian H for a time t is essentially the same as the operator resulting from the evolution generated by the Hamiltonian truncated to the region R. Mathematically, $e^{i\hat{H}t}\hat{A}e^{-i\hat{H}t} \approx e^{i\hat{H}_R t}\hat{A}e^{-i\hat{H}_R t}$ with corrections that decay exponentially with d, the radius of the region R. In the figure, the region M represents a membrane of constant thickness surrounding the region R.

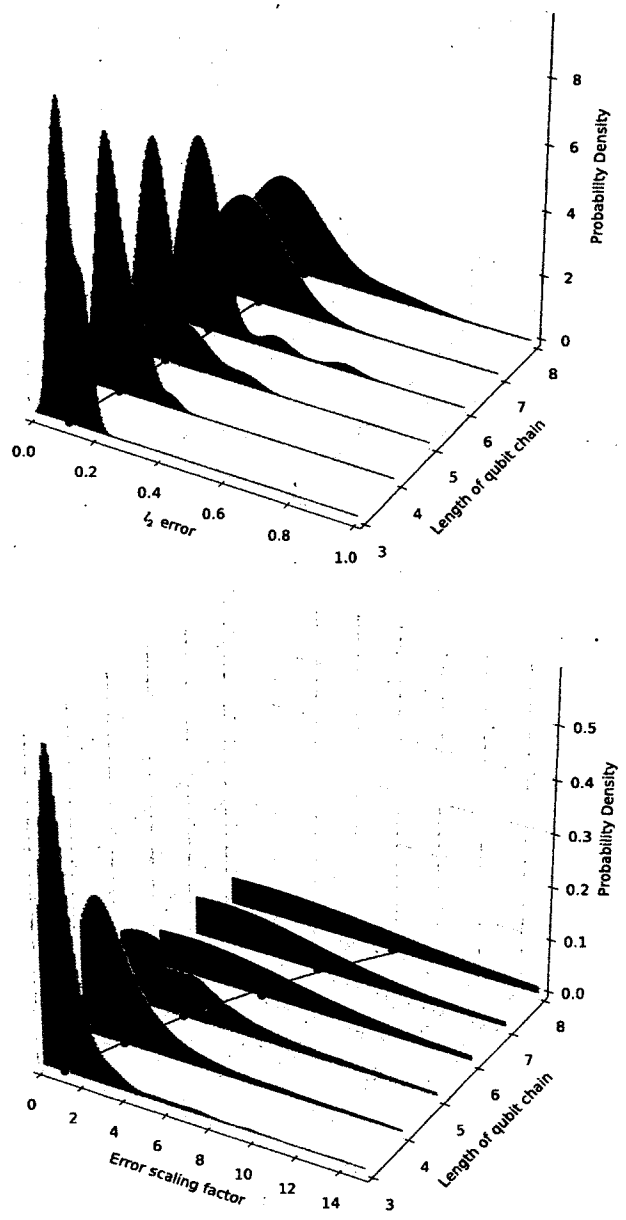


FIG. 3. Error in estimates of the parameters of local Hamiltonians. The systems consist of linear chains of qubits with randomly chosen 2-local Hamiltonians $\hat{H}$—each coefficient has norm uniformly distributed between 0.8 and 1.2. Starting in an initial product state, the system is evolved for $t = 10^{-3}$, and the expectation of randomly chosen observables is measured with precision ϵ. The resulting linear constraints Eq. (S27) are solved using Moore-Penrose pseudo-inverse to obtain an estimated Hamiltonian $\tilde{H}$. (Top) Distribution of the error $\frac{1}{d}\sqrt{\mathrm{tr}(\hat{H}-\tilde{H})^2}$ over different realization of the random Hamiltonian for $\epsilon = 10^{-4}$. The red dots correspond to the mean distance and the solid lines is a linear fit. (Bottom) Distribution of error scaling factors—i.e. the factor by which the measurement accuracy ϵ is amplified when computing the pseudo-inverse. The red dots indicate the average error scaling factor for each chain length (the red line is a quadratic fit).

i.e., $\|\hat{H}_R\| \leq \alpha d^D$ and $\|\hat{H}_M\| \leq \alpha d^{D-1}$ for some constant α. Choosing $d \approx vt + \log(cV/c't)$ such that

$$d^{D+1}\exp\left(\frac{d}{\xi}\right) \geq \frac{cV}{c't}\exp\left(\frac{vt}{\xi}\right), \tag{S25}$$

we obtain

$$\|\hat{A}(t) - \hat{A}(0) - it[\hat{H}_R, \hat{A}(0)]\| \leq \kappa\|\hat{A}\|\left[vt + \log\left(\frac{cV}{c't}\right)\right]^2 t^2 \tag{S26}$$

for some constant $\kappa = 2c'\alpha^2$.

For a short time t, the expectation value of any observable $\hat{A}$ evolves as

$$\langle\hat{A}(t)\rangle_{\hat{\rho}} - \mathrm{tr}\,\hat{A}\hat{\rho} = it\langle[\hat{H},\hat{A}]\rangle_{\hat{\rho}} + \mathcal{O}(\|\hat{H}\|^2 t^2). \tag{S27}$$

By experimentally measuring this expectation value, we obtain one linear constraint on the Hamiltonian. Varying over different observables $\hat{A}_i$ and initial states $\hat{\rho}_j$, we obtain more linear constraints that we can write as $W_{ij} = \langle\hat{A}_i(t)\rangle_{\hat{\rho}_j} - \mathrm{tr}\,\hat{A}_i\hat{\rho}_j = it\langle[\hat{H},\hat{A}_i]\rangle_{\hat{\rho}_j}$ where we have dropped the higher order terms $\mathcal{O}(\|\hat{H}\|^2t^2)$. Writing $\hat{H}$ in an operator basis $\hat{H} = \sum_l h_l\hat{P}_l$, we obtain the linear equation

$$W_{ij} = \sum_l T_{ij,l}h_l \tag{S28}$$

where $T_{ij,l} = it\,\mathrm{tr}\,\hat{\rho}_j[\hat{P}_l,\hat{A}_i]$. The Hamiltonian can be learned by inverting this linear equation [20].

There are in general four important caveats to this approach: *1)* the evolution time t must be extremely short $t \ll \|H\|^{-1}$, going to 0 as the number of particles grows; *2)* there are exponentially many h_i to learn; *3)* there are exponentially many observables $\hat{A}_k$ and initial states $\hat{\rho}_j$ to be measured and prepared experimentally; and *4)* the quantities $\mathrm{tr}\,\hat{A}\hat{\rho}$ and $\langle[\hat{H},\hat{A}]\rangle_{\hat{\rho}}$ can be exponentially difficult to compute. Based on Eq. (S27), all these problems disappear when the Hamiltonian is *local* as described in the main text.

Numerical experiments were performed for local Hamiltonians, and the results are plotted in Fig. 3. The systems we considered were small chains of qubits with random nearest neighbour interactions. The system evolution was calculated exactly for a short amount of time, and the linearized problem was inverted using the Moore-Penrose pseudoinverse. Since these Hamiltonians are drawn at random (but with maximum strength for each term independent of the system size), we calculate the average l_2 distance between the estimated Hamiltonian and the actual Hamiltonian (top of Fig. 3), as well as the quantiles for error propagation scaling factor of each of the elements of h_l, given by $\sum_{ij}|T^+_{ij,l}|^2$ (bottom of Fig. 3). The results clearly indicate well behaved error scaling for these systems, even under finite statistical error in the estimation of observable expectations.

[1] W. Hoeffding, "Probability inequalities for sums of bounded random variables," J. Am. Stat. Assoc., **58**, 13–30 (1963)

[2] H. Haffner, W. Hansel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Korber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Guhne, W. Dur, and R. Blatt, "Scalable multiparticle entanglement of trapped ions," Nature, **438**, 643–646 (2005)

[3] E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics," Nature, **409**, 46 (2001)

[4] J. M. Raimond, M. Brune, and S. Haroche, "Manipulating quantum entanglement with atoms and photons in a cavity," Rev. Mod. Phys., **73**, 565–582 (2001)

[5] Samuel L. Braunstein and H. J. Kimble, "Teleportation of continuous quantum variables," Phys. Rev. Lett., **80**, 869–872 (1998)

[6] M. Hofheinz, H. Wang, M. Ansmann, R. C. Bialczak, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, J. Wenner, J. M. Martinis, and A. N. Cleland, "Synthesizing arbitrary quantum states in a superconducting resonator," Nature, **459**, 546–549 (2009)

[7] K. E. Cahill and R. J. Glauber, "Density operators and quasiprobability distributions," Phys. Rev., **177**, 1882–1902 (1969)

[8] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," Phys. Rev., **40**, 749 (1932)

[9] D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland, "Experimental determination of the motional quantum state of a trapped atom," Phys. Rev. Lett., **77**, 4281 (1996)

[10] L. G. Lutterbach and L. Davidovich, "Method for direct measurement of the Wigner function in cavity QED and ion traps," Phys. Rev. Lett. (78)

[11] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, "Direct measurement of the Wigner function of a one-photon Fock state in a cavity," Phys. Rev. Lett., **89**, 200402 (2002)

[12] Ian Affleck, Tom Kennedy, Elliott H. Lieb, and Hal Tasaki, "Rigorous results on valence-bond ground states in antiferromagnets," Phys. Rev. Lett., **59**, 799–802 (1987)

[13] F. Verstraete and J. I. Cirac, "Matrix product states represent ground states faithfully," Phys. Rev. B, **73**, 094423 (2006)

[14] F. Verstraete and J.I. Cirac, "Renormalization algorithms for quantum-many body systems in two and higher dimensions," Arxiv preprint cond-mat/0407066 (2004)

[15] M. Van den Nest, "Simulating quantum computers with probabilistic methods," (2009), arxiv:0911.1624

[16] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," Phys. Rev. Lett., **105**, 150401 (2010)

[17] E. H. Lieb and D. W. Robinson, "The finite group velocity of quantum spin systems," Commun. Math. Phys., **28**, 251–257 (1972)

[18] M. B. Hastings, "Locality in quantum and markov dynamics on lattices and networks," Phys. Rev. Lett., **93**, 140402 (2004)

[19] D. Poulin, "Lieb-robinson bound and locality for general markovian quantum dynamics," Phys. Rev. Lett., **104**, 190401 (2010)

[20] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, "Estimation of many-body quantum hamiltonians via compressive sensing," Phys. Rev. A, **84**, 012107 (2011)

3.2.3 Marche à suivre pratique

Le protocole de certification s'accomplit donc en deux temps. Dans la phase de pré-traitement, un échantillonnage numérique est effectué à partir de la description de l'état cible. Il fournit un échantillon de N_1 observables de Pauli. Commence alors la phase expérimentale durant laquelle les valeurs moyennes de ces observables sur l'état expérimental sont estimées. Chaque observable requiert $N_2^{[k]}$ mesures répétées. Finalement, l'estimation est $F \sim \frac{1}{N_1}\sum_k \frac{\sigma_k}{\rho_k}$. Schématiquement, la marche à suivre est illustrée sur la figure 3.1.

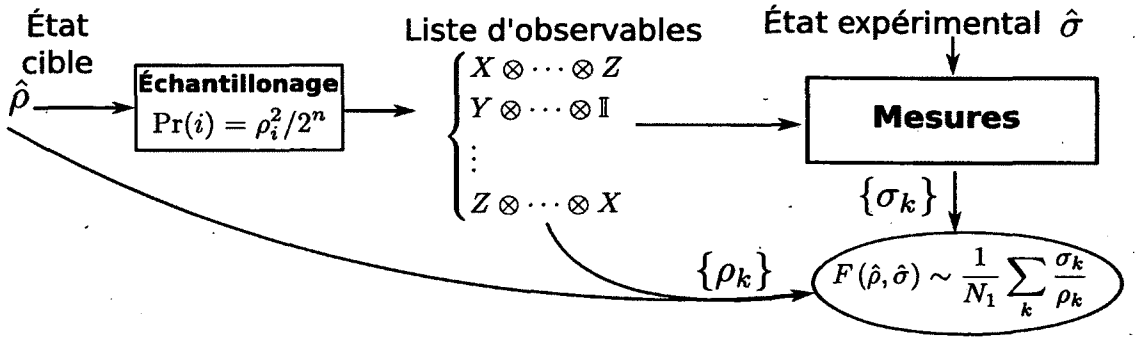


FIGURE 3.1 Marche à suivre pratique du protocole de certification

3.2.4 Certification de transformation

Notre protocole a été utilisé expérimentalement afin de certifier des portes de Toffoli sur des qubits supraconducteurs [16, 53] dans le groupe d'Andreas Wallraff à l'ETH Zurich. La porte de Toffoli est une transformation unitaire sur trois qubits. Avec les portes à un qubit [3], elle forme une famille de portes universelle pour le calcul quantique.

Profitons de l'occasion pour expliquer comment la certification d'état mène à la certification de transformations. Formellement, cette extension se fait grâce à l'isomorphisme de Choi-Jamiolkowski qui associe une matrice densité à une transformation quantique.

3. En fait, il suffit de Toffoli et de Hadamard pour obtenir une famille universelle pour le calcul quantique [54].

3.2.4.1 Transformation quantique = (super)opérateur CPTP

La transformation quantique la plus générale est un (super)opérateur[4] linéaire $\mathcal{E}$ qui transforme une matrice densité à une autre matrice densité. En particulier, elle doit préserver la trace (condition TP) mais aussi la positivité. On peut alors penser qu'il suffit que la transformation soit positive. Ceci n'est pas suffisant : en effet, il faut plus généralement que la transformation agissant sur un sous-système de l'état préserve sa positivité globale, c.à.d que $\mathbb{I} \otimes \mathcal{E}$ soit positif. Ceci aboutit à la condition de positivité complète (condition CP) sur $\mathcal{E}$. Ainsi, la transformation quantique la plus générale est un (super)opérateur CPTP.

Une autre façon d'interpréter un opérateur CPTP $\mathcal{E} : \rho_A \mapsto \mathcal{E}(\rho_A)$ est d'imaginer qu'une transformation unitaire U^{AB} agit sur l'espace de Hilbert $\mathcal{H}_A \otimes \mathcal{H}_B$ mais qu'on ne s'intéresse qu'à l'évolution du système A. Cette connexion est toujours possible grâce au théorème de dilatation de Stinespring [55].

3.2.4.2 Isomorphisme de Choi-Jamiolkowski

Il est possible d'associer à un opérateur CPTP $\mathcal{E}$ sur n qubits une matrice densité $\hat{\rho}_\mathcal{E}$ sur $2n$ qubits grâce à l'isomorphisme de Choi-Jamiolkowski. L'idée est de partir d'un état maximalement intriqué sur $2n$ qubits $|\Phi^+\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=1}^{2^n} |i\rangle \otimes |i\rangle$ et d'appliquer l'opérateur CPTP sur la moitié des qubits, *i.e.*,

$$\varphi_{CJ} : \mathcal{E} \mapsto \hat{\rho}_\mathcal{E} = (\mathbb{I} \otimes \mathcal{E}) |\Phi^+\rangle\langle\Phi^+|. \tag{3.3}$$

On peut montrer que φ_{CJ} est bien un isomorphisme, *i.e.*, il s'agit d'une bijection qui préserve la structure et dont la bijection inverse préserve aussi la structure de l'espace de Hilbert, *i.e.* son produit scalaire. L'isomorphisme est particulièrement évident si on écrit

$$\mathcal{E} = \sum_{ij} \mathcal{E}\left(|i\rangle\langle j|\right) |j\rangle\langle i| \tag{3.4}$$

$$\hat{\rho}_\mathcal{E} = \frac{1}{2^n} \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}\left(|i\rangle\langle j|\right). \tag{3.5}$$

3.2.4.3 Lien entre fidélité moyenne de sortie et fidélité d'état

Il est immédiat d'utiliser notre protocole de certification afin d'estimer la fidélité entre l'état pur $\hat{\rho}_\mathcal{U}$ correspondant à l'unitaire cible $\mathcal{U}$ et l'état mixte $\hat{\rho}_{\tilde{\mathcal{U}}}$ correspondant à la réalisation expérimentale.

4. Le terme « superopérateur » désigne un opérateur qui agit sur un espace d'opérateur.

Or, cette fidélité entre états donnent directement accès à la fidélité moyenne de sortie [56], notée $\bar{F}\left(\mathcal{U},\tilde{\mathcal{U}}\right)$, définie comme la fidélité entre $\mathcal{U}|\psi\rangle$ et $\tilde{\mathcal{U}}|\psi\rangle$ moyennée sur tous les états purs $|\psi\rangle$ puisque

$$\bar{F}\left(\mathcal{U},\tilde{\mathcal{U}}\right) = \frac{2^n F\left(\hat{\rho}_{\mathcal{U}}, \hat{\rho}_{\tilde{U}}\right) + 1}{2^n + 1}. \tag{3.6}$$

Ainsi, le protocole de certification d'état s'étend aux transformations.

3.3 Discussion

3.3.1 Commentaires techniques

Dans cette section, nous nous attarderons à plusieurs points techniques de l'article afin de les éclaircir. Le lecteur pressé peut faire l'économie de sa lecture. Ensuite, en 3.3.2 et 3.3.3, nous discuterons de deux questions intéressantes qui découlent de l'article.

3.3.1.1 Analyse d'erreur

L'analyse d'erreur aboutit à la formule suivante (voir Théorème 1 de l'article)

$$\Pr\left(\left|F - \bar{F}\right| \geq \epsilon_1 + \epsilon_2\right) \leq \frac{1}{N_1 \epsilon_1^2} + 2\exp\left[-\frac{\epsilon_2^2 N_1^2}{2}\left(\sum_{k=1}^{N_1} \frac{1}{\rho_{i_k}^2 N_2^{[k]}}\right)^{-1}\right] \tag{3.7}$$

qui fait intervenir deux termes. Le premier terme est lié à l'erreur ϵ_1 provenant des fluctuations statistiques dues à l'échantillonnage et provient d'une simple application de l'inégalité de Tchebychev à la variable aléatoire X qui prend la valeur σ_{i_k}/ρ_{i_k} avec probabilité $\Pr(i_k) \equiv \rho_{i_k}^2/d$. Le second terme s'intéresse à l'erreur ϵ_2 dues à l'estimation de la valeur moyenne des opérateur de Pauli $\hat{P}_{i_k}$ provenant de l'échantillonnage avec $N_2^{[k]}$ mesures répétées de chaque observable. Il provient de l'application du théorème de Hoeffding sur l'ensemble des $N = \sum_{k=1}^{N_1} N_2^{[k]}$ mesures. Les deux termes sont « recollés » grâce à l'inégalité de Boole.

Cette approche a le mérite de fournir une borne rigoureuse qui permet d'affirmer que le nombre de répétitions $N_2^{[k]}$ grandit polynomialement avec la taille n du système lorsque les plus petites valeurs moyennes ρ_{i_k} grandissent comme $1/\text{poly}(n)$. Toutefois, elle traite de façon

indépendante l'échantillonnage et l'estimation des valeurs moyennes alors que ces deux aspects sont étroitement liés. En effet, le nombre de répétitions $N_2^{[k]}$ nécessaire à une bonne estimation de σ_{i_k} est grand lorsque ρ_{i_k} est petit. Or, si ρ_{i_k} est petit, la probabilité $\Pr(i_k) = \rho_{i_k}^2/d$ de voir apparaître l'indice i_k lors de l'échantillonnage est faible. Ce compromis n'est pas pris en compte dans notre raisonnement.

Afin d'obtenir une borne plus serrée, il faudrait pouvoir manipuler mathématiquement une « super variable aléatoire » X dont les valeurs sont elles-mêmes des variables aléatoires. Ainsi, X deviendrait avec probabilité $\Pr(i_k)$ la valeur X_{i_k} qui prend pour valeur ± 1 avec probabilité $\frac{1}{2}\mathrm{Tr}\left[\left(\hat{P}_{i_k} \pm \mathbb{I}\right)\hat{\sigma}\right]$. La variable aléatoire X_{i_k} n'est que le reflet des mesures de Pauli sur l'état expérimental $\hat{\sigma}$. Une analyse de ce type permettrait d'arriver à une borne unifiée, sans découpler les deux problèmes puis utiliser l'inégalité de Boole pour les « recoller ». Malheureusement, je ne connais pas les outils statistiques nécessaires pour attaquer ce problème.

3.3.1.2 Adaptation de la base d'observables

Une des caractéristiques importantes de notre protocole est son efficacité. Toutefois, celle-ci dépend fortement de la structure de l'état cible. En particulier, le protocole est très efficace pour les états stabilisateurs (cf. 2.3.1.2), qui sont étroitement liés aux opérateurs de Pauli.

Or, ces opérateurs sont précisément la base d'observables utilisée pour décomposer la fidélité en une somme de termes élémentaires. On peut donc se demander s'il n'est pas possible d'adapter la base d'observables en fonction de l'état cible afin de diminuer le nombre total de mesures. Ceci serait particulièrement intéressant pour les classes d'états pour lesquelles l'échantillonnage peut être réalisé de façon efficace mais pour lesquels le nombre de mesures de Pauli nécessaire semble exponentiel, par exemple les MPS (qui seront définis en 4.2).

Formellement, le problème est le suivant : étant donné un état cible pur $|\psi\rangle$, est-il possible de trouver une base adaptée d'observables locales, c.à.d qui minimise le nombre total de mesures (en moyenne)? Le fait que les observables soient locales, au sens où elles sont des produits tensoriels d'observables agissant sur un petit nombre de particules est essentiel, puisqu'il suffirait de mesurer l'observable $|\psi\rangle\langle\psi|$ autrement. Une autre façon de voir le problème est de décomposer $|\psi\rangle\langle\psi|$ sur une base d'observables locales en s'assurant que le spectre ne soit pas uniforme, mais soit au contraire piqué autour d'une petite fraction des observables. Nous avons testé numériquement des heuristiques pour adapter les observables, mais nous avons au mieux réussi à diminuer la constante de proportionnalité devant le facteur exponentiel sans changer celui-ci, au prix de mesures sur un

nombre constant de particules (plutôt que des mesures à un corps comme pour les opérateurs de Pauli). Ceci justifie donc a posteriori de se concentrer sur les opérateurs de Pauli puisque les autres bases d'observables ne semblent pas offrir d'avantages notables.

3.3.1.3 Transformée de Fourier quantique

État de Choi-Jamiolkowski de la transformée de Fourier quantique = MPS La transformée de Fourier quantique (QFT) est un ingrédient essentiel de l'estimation de phase, une primitive importante de plusieurs algorithmes quantiques, en particulier l'algorithme de Shor [57].

Afin de donner l'expression exacte de sa transformation, il est utile d'introduire la notation binaire $0.j_\ell j_{\ell+1}\dots j_m \equiv \sum_{k=0}^{m-\ell} j_{\ell+k}/2^{k+1}$. La QFT transforme un état de la base de calcul $|j\rangle = \otimes_{k=1}^{n}|j_k\rangle$ en un état produit de la forme

$$QFT|j\rangle = \frac{1}{\sqrt{2^n}}\prod_{k=1}^{n}\left(|0\rangle + e^{2\pi i 0.j_{n-k+1}\cdots j_n}|1\rangle\right) \tag{3.8}$$

Toutefois, on montre que pour les besoins des algorithmes, il suffit de garder $m \in \mathcal{O}(\log n)$ bits de précision dans les phases. Ceci définit la transformée de Fourier approximative (AQFT)

$$AQFT|j\rangle = \frac{1}{\sqrt{2^n}}\prod_{k=1}^{n}\left(|0\rangle + e^{2\pi i 0.j_{n-k+1}\cdots j_{n-k+m}}|1\rangle\right). \tag{3.9}$$

On peut montrer que l'état de Choi-Jamiolkowski de l'AQFT est un matrix product state (MPS). Ainsi, il est possible de réaliser l'échantillonnage du protocole de certification efficacement pour cet état. Toutefois, sa distribution de valeurs moyennes sur les opérateurs de Pauli ne semble pas avoir les propriétés nécessaires afin que le nombre total de mesures demeurent polynomial.

3.3.2 Amélioration de la préparation ?

Le protocole de certification permet de déterminer la distance entre l'état expérimental et l'état cible. Ceci est très utile afin de valider une procédure expérimentale quand la préparation fournit un état très proche de l'état cible. Toutefois, en l'état actuel des technologies de l'informatique quantique, les préparations sont encore imparfaites. Dans l'expérience du qubyte, la fidélité pour l'état à 8 qubits était $F \sim 0.722$ [19]. Or, notre protocole ne fournit aucune autre information que

la distance à l'état cible. La tomographie fournit plus d'information, mais à un coût rédhibitoire pour des systèmes de grande taille, et même de taille moyenne.

On peut se demander s'il existe une tâche intermédiaire, qui n'estime pas qu'un seul paramètre (certification) ni un nombre exponentiel (tomographie), et surtout qui réponde aux besoins des expérimentateurs. En pratique, caractériser le système a souvent pour but d'améliorer la préparation. En effet, les expérimentateurs veulent savoir comment préparer un état $\hat{\sigma}$ qui soit plus proche de l'état cible $\hat{\rho}$. Formellement, ils veulent maximiser $F_{\hat{\rho}}(\hat{\sigma}) \equiv F(\hat{\rho}, \hat{\sigma})$ où on considère que l'état cible est fixé. Pour se faire, ils peuvent modifier les paramètres $\{q_i\}$ de la préparation, ce qui se traduira par un nouvel état expérimental $\hat{\sigma}(\{q_i\})$. Ainsi, le problème formel est de maximiser la fonction $F_{\hat{\rho}}(\{q_i\})$. Une première étape afin d'attaquer ce problème d'optimisation serait de pouvoir évaluer le gradient de cette fonction.

Si la certification est très efficace, on peut modifier un paramètre de préparation $q_k \mapsto q_k + dq_k$ et estimer la fidélité à cet état $\sigma + \frac{\partial \sigma}{\partial q_k} dq_k$. On obtient alors une valeur approchée de

$$\frac{\partial F}{\partial q_k} \sim F(\sigma + \frac{\partial \sigma}{\partial q_k} dq_k) - F(\sigma). \tag{3.10}$$

Une autre approche est d'utiliser une approche Monte-Carlo similaire au protocole de certification et d'échantillonner la variation $\frac{\partial \sigma_i}{\partial q_k}$ pour un petit nombre d'opérateurs de Pauli

$$\frac{\partial F}{\partial q_k} = \frac{1}{d} \sum_i \rho_i \frac{\partial \sigma_i}{\partial q_k} = \sum_i \Pr(i) \frac{1}{\rho_i} \frac{\partial \sigma_i}{\partial q_k}. \tag{3.11}$$

Une fois le gradient estimé de façon expérimentale, il est envisageable d'utiliser des approches d'optimisation comme l'algorithme du gradient ou la méthode du gradient conjugué pour atteindre une meilleure préparation.

3.3.3 Lien entre les états certifiables et les états simulables

Une question intéressante soulevée par notre protocole de certification est de caractériser les états cibles pour lequel il est efficace. Dans l'article, nous avons donné plusieurs exemples d'états (et de transformations) certifiables, mais une théorie complète fait défaut. On dira qu'un état est certifiable si l'échantillonnage peut être fait de façon efficace et le nombre total de mesures grandit

polynomialement avec la taille du système.

L'échantillonnage est un problème numérique. Il est efficace s'il peut être fait dans un temps polynomial, et donc en particulier avec un espace mémoire de taille polynomial. La restriction d'espace élimine déjà plusieurs états dans l'espace de Hilbert qui ne peuvent pas être décrit par un nombre polynomial de coefficients. Les états certifiables ont donc une représentation concise : nous reviendrons sur cette notion dans le prochain chapitre, en 4.1. Ensuite, il faut pouvoir effectuer l'échantillonnage à partir de cette représentation. Un résultat intéressant dans ce sens a été de montrer que pour une classe d'états à n qubits dont on sait calculer la valeur moyenne d'un produit tensoriel d'observables locales dans un temps $q(n)$, l'échantillonnage peut se faire en temps $n \times q(2n)$. Ainsi, ceci permet de faire une connexion aux états « manipulables numériquement » (*computationally tractable*), défini dans [58] pour lesquels $q(n) \in \text{poly}(n)$.

Le nombre total de mesures nécessaire est lié à la décomposition de l'état cible sur la base d'observables choisie. En effet, calculons l'espérance $\bar{N}$ du nombre total de mesures $N = \sum_{k=1}^{N_1} N_2^{[k]}$. Le nombre de répétitions $N_2^{[k]}$ nécessaire afin d'estimer la valeur moyenne expérimentale σ_{i_k} varie comme $(\rho_{i_k})^{-2}$. Autrement dit, $N_2^{[k]}$ est une variable aléatoire qui vaut $(\rho_{i_k})^{-2}$ avec probabilité $\Pr(i_k)$. L'espérance du nombre total de mesures est donc

$$\bar{N} = \sum_i \Pr(i)\, \rho_i^{-2} = \frac{1}{d} \left| \{ \rho_i^2 > d^{-1} \} \right| \tag{3.12}$$

où apparaît le cardinal de l'ensemble des ρ_i non-négligeables, c.à.d. ceux qui sont assez significatifs pour que l'estimation expérimentale de σ_i soit nécessaire. Dans le cas particulier des états stabilisateurs, il y a d valeurs propres non-nulles donc $\bar{N}$ est une constante indépendante de la taille du système. Dans le pire cas, il y aura d^2 indices significatifs, ce qui permet de retrouver $\bar{N} \in \mathcal{O}(d)$ annoncé dans l'article. Au-delà de l'espérance, il est important de calculer la variance $\mathbb{V}(N)$ du nombre total de mesures puisqu'il permet de calculer le nombre de mesures qu'il faudra effectuer avec grande probabilité (via l'inégalité de Tchebychev). La variance s'exprime comme suit

$$\begin{aligned} \mathbb{V}(N) &= \sum_i \Pr(i)\, \rho_i^{-4} - \bar{N}^2 && (3.13) \\ &= \frac{1}{d} \sum_i \rho_i^{-2} - \bar{N}^2 && (3.14) \\ &\leq \left| \{ \rho_i^2 > d^{-1} \} \right| && (3.15) \end{aligned}$$

La majoration donnée dans la dernière ligne est très grossière, mais laisse penser que la variance

sera exponentiellement grande pour un état avec peu de structure. Le calcul de l'espérance et de la variance du nombre total de mesures permet d'aller au-delà du critère donné dans l'article qui affirme qu'il suffit que les ρ_i non-négligeables grandissent comme $1/\text{poly}(n)$ pour que le nombre total de mesures soit polynomial.

Pour l'instant, nous n'avons que des résultats partiels sur la caractérisation des états certifiables. Il serait intéressant d'établir des connexions avec les états apprenables, c.à.d. ceux dont la description peut être reconstituée avec un nombre polynomial de mesures et un traitement numérique efficace. Ces états feront l'objet du chapitre suivant.

Chapitre 4

Tomographie variationnelle

4.1 États quantiques à description efficace

4.1.1 États physiques

La tomographie est une tâche exponentiellement coûteuse car elle vise à estimer la description d'un état quantique. Or, la description la plus générale d'un état, sous forme d'une matrice densité contient un nombre exponentiel de coefficients. Ainsi, quelles que soient les observables permises, il faudra un nombre exponentiel de mesures afin de reconstruire la matrice densité [41].

Or, les états qui nous intéressent le plus souvent ne sont pas des états quelconques, choisis au hasard dans l'espace de Hilbert, aux propriétés souvent étranges. Au contraire, les états physiques ont beaucoup de structure et ne représentent qu'une infime portion des états de l'espace de Hilbert [12]. En particulier, plusieurs d'entre eux admettent une description efficace, c.-à-d. qu'ils ne sont décrits que par un petit nombre de coefficients (polynomial dans le nombre de particules) et permettent le calcul efficace de quantités physiques. Pour ces états, il est concevable qu'un protocole ciblé d'apprentissage permette de les estimer avec un petit nombre de mesures expérimentales. C'est ce que nous démontrerons dans le cas de deux classes variationnelles d'états, les MPS (en section 4.2) et les MERA (en section 4.5).

4.1.2 Description efficace

Depuis von Neumann [59], la mécanique quantique est formulée dans le cadre mathématique d'un espace de Hilbert, *i.e.*, d'un espace vectoriel complexe muni d'un produit scalaire et complet pour la norme issue du produit scalaire. En informatique quantique, on se place habituellement dans un espace de dimension finie, *i.e.*, un espace hermitien. Les états purs sont représentés par des vecteurs unitaires dans cet espace.

Le problème fondamental de la physique à n corps est que la dimension d de l'espace de Hilbert de n particules grandit exponentiellement. Par exemple, pour n qubits, on a $d = 2^n$. Ainsi, les problèmes deviennent ingérables. Par exemple, pour trouver le fondamental d'un hamiltonien, il suffirait de diagonaliser une matrice hermitienne de taille d. Or, il n'est même pas possible d'écrire cette matrice, encore moins de la diagonaliser dès que le nombre de particules atteint quelques dizaines. Ainsi, il faut trouver des méthodes pour contourner cette catastrophe exponentielle.

4.1.2.1 Représentation concise

Pour certaines familles d'états, la description comme vecteur dans un espace de Hilbert est excessive et redondante. Par exemple, pour un état produit à n qubits, il suffit de donner la description individuelle de chacun des n qubits, ce qui ne demande que $2n$ coefficients. Il s'agit d'un gain énorme : on passe d'un nombre exponentiel de coefficients à un nombre linéaire. Plus généralement, nous définirons les *états à représentation concise* de la façon suivante

Définition 1. Une famille d'états $\{|\psi_n\rangle\}_{n\in\mathbb{N}^*}$ admet une *représentation concise* si la donnée d'un nombre polynomial de coefficients permet de reconstruire (pas nécessairement efficacement) n'importe quel coefficients de son vecteur de Hilbert.

Ainsi, les états produits admettent une représentation concise. Nous verrons que plusieurs classes d'états, intéressantes physiquement, possèdent une représentation concise, en particulier plusieurs classes variationnelles d'états très utilisées numériquement.

Bien évidemment, une représentation concise est nécessaire afin de pouvoir manipuler efficacement un état. Toutefois, elle ne suffit pas à pouvoir calculer efficacement des quantités d'intérêt. Par exemple, le fondamental d'un hamiltonien local sur réseau admet aussi une représentation concise, le hamiltonien local lui-même. En effet, ce hamiltonien est défini par un nombre polynomial de coefficients : chaque terme du hamiltonien contient un nombre constant (indépendant de la taille

du système) de coefficients et il y a un nombre linéaire de termes dans le hamiltonien. Il existe donc une procédure théorique pour reconstruire n'importe quel coefficient de son vecteur de Hilbert : écrire la matrice correspondant au hamiltonien puis la diagonaliser. Malheureusement, cette procédure n'est pas efficace et cette représentation, bien que concise, ne permet pas de calculer des propriétés physiques. Il faut donc ajouter d'autres contraintes que la représentation concise pour obtenir des états dont on peut « faire quelque chose ».

4.1.2.2 États à description efficace

Typiquement, la description d'un état intéresse le physicien car il veut l'utiliser pour calculer des quantités physiques d'intérêt, telles l'énergie, des fonctions de corrélation, la magnétisation, etc. Ces quantités physiques sont de façon générale les valeurs moyennes d'observables locales, *i.e.*, qui n'agissent que sur un nombre de particules indépendant de la taille du système. Conformément à cette intuition, nous définirons les états manipulables efficacement de la façon suivante :

Définition 2. Une famille d'états $\{|\psi_n\rangle\}_{n\in\mathbb{N}^*}$ ont une *description efficace* si elle admet une représentation concise permettant de calculer efficacement la valeur moyenne $\langle O\rangle \equiv \langle\psi_n|O|\psi_n\rangle$ de n'importe quelle observable locale O, *i.e.*, s'il existe un algorithme polynomial dont l'entrée est la représentation concise et dont la sortie est (une approximation [1]) de la valeur moyenne.

Ainsi, la représentation concise du fondamental d'un hamiltonien local donnée plus haut ne fournit pas une description efficace. Au contraire, un état produit a une description efficace. Toutefois, admettre une description efficace est une propriété mathématique : elle ne dit rien sur le caractère physique des états.

4.1.2.3 Représentation fidèle d'un hamiltonien

L'utilité d'états à description efficace réside dans sa capacité à résoudre des problèmes physiques [2]. Les états à description efficace sont très utilisés numériquement afin de déterminer les états fondamentaux de hamiltonien. Ainsi, on regroupera plusieurs familles d'états afin d'obtenir

1. La question de la précision de l'estimation de la valeur moyenne est épineuse. Numériquement, on demande typiquement un algorithme qui approxime $\langle O\rangle$ en poly(m,n) où m est le nombre de bits significatifs. Il s'agit donc d'une précision exponentiellement. Or, si l'on disposait expérimentalement de poly(n) copies de $|\psi_n\rangle$ et qu'on mesurait de façon répétitive l'observable O, on n'obtiendrait une approximation de $\langle O\rangle$ dont la précision ne varie que comme $1/\text{poly}(n)$ (approximation polynomiale) avec une probabilité exponentiellement proche de 1. Opérationnellement, la notion d'approximation polynomiale est donc mieux justifiée et c'est celle que nous privilégierons. Toutefois, dans le cadre de cette thèse, cette distinction n'interviendra pas. Pour une discussion plus complète, voir [58].

2. Comme le dit le proverbe, *the proof of the pudding is in the eating.*

une classe d'états sur laquelle on minimisera l'énergie. La question est alors de savoir si on obtient une bonne approximation de l'énergie fondamentale et des états fondamentaux.

Une première indication en ce sens est de déterminer si l'état fondamental déterminé numériquement reproduit les propriétés attendues. Par exemple, le fondamental d'un hamiltonien local gappé présente une décroissance exponentielle des corrélations [60]. Ainsi, une méthode numérique visant à approximer le fondamental d'un tel hamiltonien peut se restreindre aux états à description efficace qui présentent cette propriété. Une notion plus forte de fidélité physique serait de démontrer qu'il existe dans la classe d'états une famille d'états $|\psi_n\rangle$ qui approxime un état fondamental de H_n. Par exemple, les états MPS qui seront définis en 4.2 représentent fidèlement les états fondamentaux des hamiltoniens locaux 1D [61].

En pratique, on construit souvent une classe d'états en déterminant quelles sont les propriétés physiques attendues et en exploitant cette structure afin d'aboutir à une description efficace. Un exemple particulièrement important est la notion de loi d'aire (area law) sur laquelle nous reviendrons en 4.1.3.2.

4.1.3 Classes variationnelles d'états

Un exemple particulièrement important de classes d'états est celle des classes variationnelles qui sont très utilisées en physique. Formellement, une classe variationnelle est un ensemble de familles d'états indicés par des paramètres variationnels $\alpha \in \mathbb{C}^N$ où N est une fonction polynomiale du nombre n de particules. Les classes variationnelles sont très utilisées analytiquement et surtout numériquement afin de résoudre des problèmes à n corps.

4.1.3.1 Ansatz pour des problèmes à n corps

Un problème à n corps est un problème qui fait intervenir un grand nombre de particules qui ne peuvent pas être traités individuellement. Or, il est hors de question d'écrire l'état à n corps comme un état pur dans un espace de Hilbert de dimension exponentielle. Ainsi, plusieurs classes variationnelles ont été proposées comme ansatz pour différents problèmes. Par exemple, l'approche champ moyen correspond aux états produits et la méthode de Hartree-Fock est une recherche variationnelle sur les états produits antisymmétrisés (déterminants de Slater).

Souvent, de grands progrès ont été faits en physique quand des classes variationnelles ont été proposées, car elles cristallisent les connaissances antérieures et sont le reflet de l'intuition de leurs auteurs. Ainsi, la fonction d'onde de Laughlin [62] a permis de mieux comprendre l'effet Hall quantique fractionnaire (FQHE) et la fonction d'onde BCS résultait d'une théorie microscopique pour la supraconductivité [63, 64].

4.1.3.2 Structure de l'intrication : loi d'aire (area law)

Une propriété vérifiée par les états fondamentaux de nombreux hamiltoniens locaux est la loi d'aire pour l'entropie. Cette propriété est importante car elle a mené au développement de plusieurs classes variationnelles. Nous allons donc définir la loi d'aire avant de s'intéresser à la classe variationnelle qui lui correspond en 1D : les MPS.

L'intrication est une quantité particulièrement étudiée en informatique quantique qui mesure l'écart à un état produit et donc les corrélations quantiques entre particules. Un objectif important du domaine est de mieux comprendre la structure des états intriqués.

Considérons un état $|\psi\rangle$ de n qubits disposés sur un réseau Λ. À un sous-ensemble de particules $X \subset \Lambda$ correspond l'état $\rho_X = \mathrm{Tr}_{\Lambda\setminus X}\left[|\psi\rangle\langle\psi|\right]$. Une façon de quantifier l'intrication entre la région X et le reste du réseau est de calculer son entropie (de von Neumann)

$$S_X \equiv S(\rho_X) \equiv -\mathrm{Tr}\left[\rho_X \log \rho_X\right]. \tag{4.1}$$

Une question importante est alors de comprendre comment cette quantité varie en fonction de X. Pour un état choisi aléatoirement dans l'espace de Hilbert, l'entropie varie comme le nombre de particules dans la région X. En effet, avec grande probabilité, l'état ρ_X sera proche de l'état complètement mélangé $\mathbb{I}/d_X$ dont l'entropie est $\log d_X$ qui est proportionnel au nombre de particules. Formellement, on a $S_X \in \mathcal{O}(|X|)$ et on parle de loi de volume.

Toutefois, pour des états physiques, la situation est souvent dramatiquement différente. En particulier, certains états obéissent à une loi d'aire : leur entropie varie comme la taille de la frontière ∂X

$$S_X \leq \mathcal{O}(|\partial X|). \tag{4.2}$$

Intuitivement, cela correspond au fait que l'intrication soit concentrée sur la frontière. Notons tout de suite que le terme « loi d'aire » est malheureux lorsqu'il est utilisé en 1D et en 2D : on devrait

plutôt parler de loi de périmètre en 2D et d'entropie qui sature en 1D. Le vocable « loi de frontière » serait peut-être plus approprié. Toutefois, nous sacrifierons à la tradition en utilisant le terme « loi d'aire ».

Pour quels états cette loi d'aire est-elle respectée ? Le folklore voulait que l'état fondamental d'un hamiltonien local obéisse à une loi d'aire. Toutefois, cette intuition s'est révélée fausse. En effet, même en 1D [65] et même avec un modèle invariant sous translation [66], il existe des hamiltoniens dont l'état fondamental obéit à une loi de volume, *i.e.*, l'entropie de l'état fondamental est proportionnelle à la longueur de la chaîne. Il s'agit toutefois d'états fondamentaux de modèles critiques dont le gap se ferme. Le gap d'un hamiltonien est la différence d'énergie entre l'énergie fondamentale et l'énergie des premiers états excités. La notion de hamiltonien gappé ou non doit se comprendre dans la limite thermodynamique, quand le nombre de particules n tend vers l'infini. Un hamiltonien sera gappé si son gap est plus grand qu'une constante indépendante de la taille du système et non-gappé sinon. Ainsi, un système non-gappé aura souvent un gap qui décroit en $1/\text{poly}(n)$ ou $\exp(-n)$. En 1D, pour un hamiltonien local gappé, un fondamental aura une entropie qui sature, *i.e.*, S_X sera borné par une constante indépendante de la taille du système [67, 68, 69].

Il est possible de construire des classes variationnelles basés sur la loi d'aire. En 1D, cela mènera à la notion d'états à produit matriciel (MPS) qui fera l'objet de la section 4.2. Les MPS fournisse une bonne approximation des fondamentaux de hamiltonien 1D gappé puisque leur entropie sature [67]. Après avoir défini les MPS, nous nous intéresserons à la question suivante : est-il possible d'apprendre les paramètres variationnels d'un MPS à l'aide d'un petit nombre de mesures expérimentales ? Nous montrerons que oui en proposant un protocole basé sur la représentation d'un MPS en tant que sortie d'un circuit quantique dans la section 4.3. Ceci fournit un outil utile pour caractériser les états fondamentaux de hamiltoniens gappés en 1D.

Toutefois, les MPS ne sont pas adaptés pour les modèles critiques, en particulier ceux dont l'entropie du fondamental diverge logarithmiquement, indiquant de l'intrication à plusieurs échelles. Dans ce cas, il faudra utiliser une classe variationnelle plus grande, celle des MERA, abordée dans la section 4.5. Après avoir défini cette classe, nous montrerons qu'il est aussi possible de les apprendre avec un petit nombre de mesures simples et un traitement numérique efficace en 4.6.

4.2 États à produit matriciel (MPS)

Les états à produit matriciel (matrix product states), que nous désignerons désormais par leur acronyme MPS, ont plusieurs caractérisations équivalentes. Nous allons les définir par leur représentation matricielle, en 4.2.1.1. Nous montrerons ensuite (en 4.2.1.3) que les MPS sont un cas particulier d'états à réseaux de tenseurs, ce qui permet d'en donner une représentation graphique simple et de les manipuler facilement. Finalement, nous montrerons en 4.2.3 que les MPS peuvent être obtenus comme états de sortie d'un circuit quantique où l'arrangement des portes a une structure particulière. Cette caractérisation permettra de préparer le terrain pour l'article sur l'apprentissage des MPS, présenté dans la section 4.3.

Les états MPS forment une classe variationnelle très utile pour les systèmes 1D non-critique. En effet, par construction, l'entropie d'un bloc de L particules sature vers une constante quand L grandit, une propriété attendue pour ces systèmes. Cette propriété est détaillée en A.1. Ainsi, les MPS sont très utilisés numériquement : la méthode DMRG (density matrix renormalization group) est une méthode numérique très efficace et très employée pour l'étude des systèmes 1D. Originellement proposée par White en 1992 [70], il a été montré qu'elle était équivalente à une recherche variationnelle sur l'espace des MPS [71, 72] [3].

4.2.1 Définition des MPS

4.2.1.1 Représentation matricielle

Un MPS de dimension de lien D est un état de n particules à d niveaux, ou *qudits*, qui s'écrit sous la forme

$$|\Psi\rangle = \sum_{i_1 \dots i_n = 0}^{d-1} \mathrm{Tr}\left[A^{[1]i_1} A^{[2]i_2} \dots A^{[n]i_n}\right] |i_1 \dots i_n\rangle \tag{4.3}$$

où $\left\{A^{[k]i_k}\right\}_{i_k=0}^{d-1}$ est une famille de matrices complexes de dimension $D_k \times \tilde{D}_k$, telle que $\tilde{D}_k = D_{k+1}$ et $\forall k\ D_k \leq D$.

Notons que tout état pur est un MPS avec $D = d^{n/2}$, *i.e.*, avec une dimension de lien exponentiellement grande. Généralement, on réserve le terme MPS quand la dimension de lien D est indépendante de la taille du système ($D \in \mathcal{O}(1)$), voire polynomiale ($D \in \mathrm{poly}(n)$).

3. Pour plus de détails sur la DMRG comme méthode variationnelle sur les MPS, lire [73].

La dimension de lien D quantifie le rang des matrices densité réduites et donc l'intrication de l'état. Ainsi, le rang de tout bloc de L particules est D^2 et son entropie est bornée par

$$S(L) \leq 2 \log D. \tag{4.4}$$

Le plus simple pour comprendre cette relation est d'expliciter le lien entre les MPS et la décomposition de Schmidt donnée en annexe A.1.

Les conditions aux frontières imposées sur l'état se reflète dans la dimension des matrices $A^{[1]}$ et $A^{[n]}$. Pour des conditions aux frontières ouvertes (*open boundary conditions* ou OBC), c.à.d. des particules sur une ligne, on a $D_1 = \tilde{D}_n = 1$. Sinon, les conditions aux frontières sont périodiques (*periodic boundary conditions* ou PBC) car on peut effectuer une permutation circulaire dans la trace et la notion de 1ère particule n'a plus de sens : les particules sont sur un cercle. Un MPS dont toutes les matrices sont identiques est invariant sous translation (*translationally invariant* ou TI).

4.2.1.2 Exemples de MPS

État GHZ L'état GHZ non-normalisé est $|GHZ_n\rangle = |0\rangle^{\otimes n} + |1\rangle^{\otimes n}$

Une représentation MPS invariante sous translation est

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad A^1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.5}$$

État W Une représentation OBC[4] de l'état $|W_n\rangle$ défini par l'éq. (2.1) est

$$A^{[1]0} = \begin{pmatrix} 1 & 0 \end{pmatrix} A^{[1]1} = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad A^{[n]0} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} A^{[n]1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{4.6}$$

$$A^{[k]0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad A^{[k]1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{4.7}$$

4. Étonnamment, bien que l'état W soit invariant sous translation (et même sous permutation), une représentation PBC semble demander une dimension de lien qui grandit avec n.

4.2.1.3 États à réseaux de tenseurs

Les MPS sont un exemple d'une classe plus générale d'états appelés « états à réseau de tenseurs » (*tensor network states*). En effet, un MPS est défini par n familles de matrices $\left\{A^{[k]i_k}\right\}_{i_k=0}^{d-1}$. On peut penser à cet objet comme à un tenseur de rang 3 $\left(A^{[k]}\right)_{\alpha\beta}^{i_k}$ où i_k est l'indice physique qui décrit l'état de la particule k et α, β sont des indices correspondant à des degrés de liberté virtuels.

Un état pur, *i.e.*, un vecteur dans l'espace de Hilbert, est décrit par ses coefficients

$$|\psi\rangle = \sum_{i_1 \ldots i_n} c_{i_1 \ldots i_n} |i_1 \ldots i_n\rangle \tag{4.8}$$

dans une base de référence. On peut penser aux coefficients comme à un énorme tenseur de rang n. Pour un état quelconque, ce tenseur n'aura pas de structure et sera donc constitué de d^n coefficients indépendants. Or, pour des états plus structurés, il peut être possible de décomposer cet énorme tenseur en une collection de plus petits tenseurs. C'est le cas pour les MPS : le tenseur des coefficients se ramène à n tenseurs de rang 3. À l'aide de ces tenseurs, il est possible de reconstruire le grand tenseur des coefficients en contractant les plus petits tenseurs, cf. Fig. 4.1.

Représentation schématique des MPS Plutôt que de traîner de longues formules, il est souvent plus pratique de représenter les quantités calculées sur un MPS grâce à un schéma où chaque symbole, typiquement un carré, représente un tenseur et chaque ligne sortante du symbole (ou « patte ») représente un indice de ce tenseur. Par exemple, il est facile de représenter un état quantique par le tenseur $c_{i_1 \ldots i_n}$ de ses coefficients dans la base de calcul, voir éq. (4.8).

Pour un état quelconque, ce tenseur n'a aucune structure. Pour un état produit, ce tenseur se décompose en n tenseurs indépendants. Pour un MPS, ce tenseur se décompose en n tenseurs de rang trois dont l'indice physique prend d valeurs différentes et dont deux indices sont virtuels et prennent D valeurs. Schématiquement, ces classes d'états sont représentés sur la figure 4.1.

Contraction de tenseurs Le calcul d'une quantité physique à partir d'un état à réseau de tenseurs se fera en contractant les tenseurs. Par exemple, la valeur moyenne d'une observable est représentée sur la Fig. 4.3. Schématiquement, la contraction est représentée par la fusion de deux lignes appartenant à des symboles différents (produit matriciel), voire au même symbole (trace).

On peut penser à la contraction de tenseurs par analogie au produit matriciel. Pour le produit

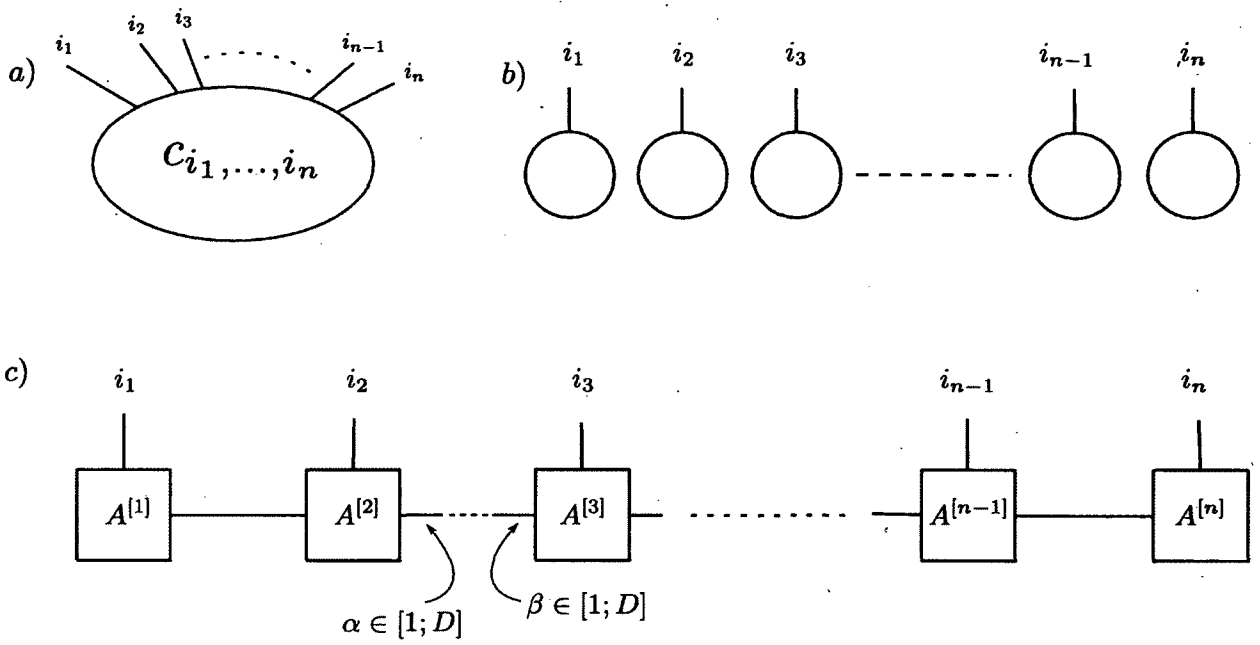


FIGURE 4.1 Exemple d'états à réseau de tenseurs.
a) État quelconque. b) État produit. c) MPS

C de deux matrices A et B, les coefficients du produit sont obtenus grâce à la formule suivante

$$\sum_k A_{ik} B_{kj} = C_{ij} \tag{4.9}$$

qu'il est pratique de réécrire en utilisant la convention d'Einstein (somme implicite sur les indices répétés) comme

$$A_{ik} B_{kj} = C_{ij}. \tag{4.10}$$

Ainsi, formellement, on identifie l'indice k qui est le 2e indice de A et le 1er de B afin d'obtenir C.

Pour l'usage que nous en ferons dans cette thèse, il suffit de voir les tenseurs comme des objets mathématiques qui généralisent la notion de matrices puisqu'ils possèdent plus de deux indices. Ainsi, un tenseur A de rang r_A possède r_A indices et une matrice correspond à un tenseur d'ordre 2. De façon similaire au produit matriciel, on définit la contraction C de deux tenseurs A (de rang r_A) et B (de rang r_B) suivant l'indice γ comme étant un tenseur C de rang $r_C = r_A + r_B - 2$ dont les entrées sont données par l'identification de l'indice contracté γ et une sommation sur cet

indice. Par exemple, pour $r_A = r_B = 3$, on a par exemple

$$A_{\alpha\beta\gamma} B_{\delta\gamma\epsilon} = C_{\alpha\beta\delta\epsilon}. \tag{4.11}$$

Ici, l'indice contracté est le troisième indice du tenseur A et le deuxième du tenseur B.. Cette contraction est représentée sur la figure 4.2 a).

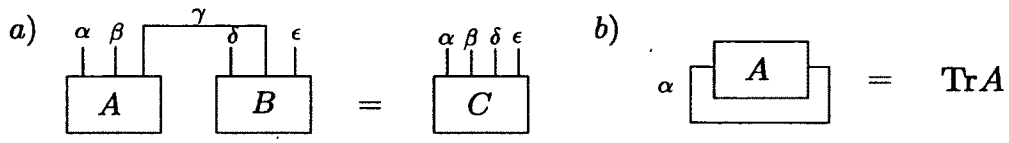


FIGURE 4.2 Exemples de contractions de tenseurs. a) Contraction de l'exemple (4.11) b) Cas particulier de la trace

Une autre contraction possible est de contracter deux indices du même tenseur afin d'obtenir un nouveau tenseur, p.ex. $A_{\alpha\beta\beta} = B_\alpha$. En particulier, la contraction $A_{\alpha\alpha}$ est un tenseur d'ordre 0, *i.e.*, un scalaire, qui n'est autre que la trace de A, cf. figure 4.2 b).

Notons qu'il est aussi possible de contracter plusieurs indices à la fois. Quitte à définir des super-indices qui regroupent plusieurs indices, toute contraction se ramène au cas d'une contraction simple sur deux tenseurs de rang 2, *i.e.* à un produit matriciel, de type $A_{\alpha\beta} B_{\beta\gamma} = C_{\alpha\gamma}$ où α, γ regroupent les indices non-contractées de A et B et β regroupent les indices contractés entre A et B. Ainsi, le coût numérique d'une contraction revient au coût numérique de multiplier des matrices. L'algorithme naïf[5], *i.e.* le calcul explicite de (4.9) pour des matrices $m \times n$ et $n \times p$, se fait en temps $\mathcal{O}(mnp)$. Autrement dit, le temps de calcul est proportionnel aux nombre de valeurs que peut prendre les indices non-sommés des tenseurs A et B et au nombre de valeur possible pour l'indice sommé. Ceci permet donc de calculer le coût numérique d'une opération incluant plusieurs contractions de tenseurs, p.ex. le calcul de la valeur moyenne d'une observable comme nous le verrons en 4.2.2.

5. La complexité du produit de deux matrices n'est pas connue. Supposons pour simplifier qu'elles soient de taille $n \times n$. L'algorithme naif est en $\mathcal{O}(n^3)$. Des algorithmes plus sophistiqués ont une complexité moindre, p.ex. l'algorithme de Strassen en $\mathcal{O}(n^{\log_2 7})$. Toutefois, ces algorithmes sont souvent instables et ne présentent un réel avantage que pour des matrices de très grandes tailles.

4.2.2 MPS comme état à description efficace

Par définition, les MPS ont une représentation concise. De plus, il est possible de calculer efficacement la valeur moyenne d'une observable locale à partir de la représentation MPS. En fait, les MPS permettent le calcul de toutes quantités de la forme

$$\langle\phi|\bigotimes_{k=1}^{n} S_k|\psi\rangle \tag{4.12}$$

où $|\psi\rangle$ et $|\phi\rangle$ sont des MPS et $\bigotimes_{k=1}^{n}$ est un produit tensoriel d'observables.

Pour ce faire, il suffit de représenter la quantité donnée par l'éq. (4.12) sous la forme d'un schéma tensoriel, cf. Fig 4.3. On se convainc facilement que le calcul peut se faire efficacement en contractant les matrices de transfert, notées $E[k]$ sur la figure puis en les contractant. Cette approche naïve se fait en $\mathcal{O}(nD^4)$ (et une approche plus astucieuse donne $\mathcal{O}(nD^3d)$) pour des MPS-OBC. Avec des conditions aux frontières périodiques, le calcul est plus coûteux : en effet, pour des MPS-PBC, on aurait $\mathcal{O}(nD^6)$ avec l'approche naïve et $\mathcal{O}(nD^5d)$ avec l'approche astucieuse.

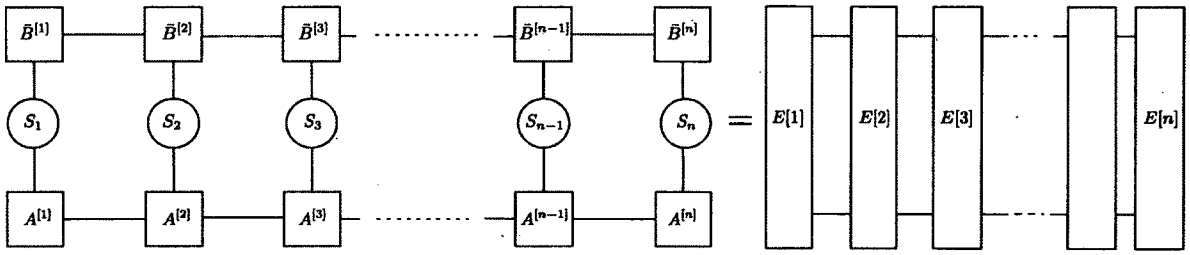


FIGURE 4.3 Calcul de la quantité $\langle\phi|\bigotimes_{k=1}^{n} S_k|\psi\rangle$ pour des MPS.
Les tenseurs $\{A^{[k]}\}$ (resp. $\{\bar{B}^{[k]}\}$) correspondent à $|\psi\rangle$ (resp. $\langle\phi|$).

4.2.3 Représentation en circuit

Tout MPS-OBC peut aussi être vu comme l'état de sortie d'un circuit quantique avec une structure très particulière, qui est le reflet de la structure de ses tenseurs. Il s'agit d'un circuit qui prend un état de référence, disons $|0\rangle^{\otimes n}$ et lui applique un circuit « en escalier », cf. Fig. 4.4. Un tel circuit est une séquence de $\mathcal{O}(n)$ portes quantiques qui agissent sur $\ell = \lceil \log_d D \rceil + 1$ qudits. La première porte agit sur les qudits numérotés 1 à ℓ. Une fois transformé, le premier qudit (qudit 1)

ne sera plus affecté par aucune porte. Par contre, les $\ell - 1$ autres qudits seront en entrée de la prochaine porte quantique, en plus d'un qubit ancillaire préparé dans l'état $|0\rangle$. Une fois la porte appliquée, le qudit 2 ne sera plus affecté, mais les $\ell - 1$ autres qudits vont dans la prochaine porte. En résumé, la porte k agit sur les qudits $[\![k, k + \ell - 1]\!]$.

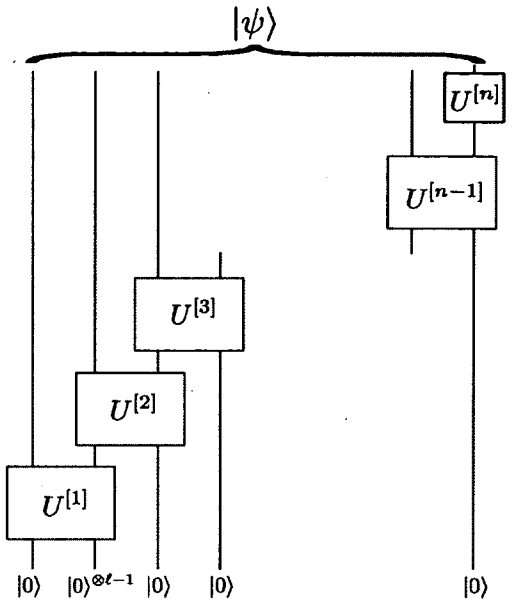


FIGURE 4.4 Circuit préparateur de MPS où les portes quantiques sont disposées « en escalier ». Contrairement à la convention habituelle, l'entrée du circuit $|0\rangle^{\otimes n}$ est située en bas alors que le MPS $|\psi\rangle$ apparaît en sortie en haut du circuit.

Relations de passage Nous allons maintenant prouver que l'état de sortie d'un circuit en escalier est un MPS-OBC en donnant les règles afin de passer d'une représentation à l'autre. Réciproquement, tout état MPS-OBC est l'état de sortie d'un circuit en escalier en inversant ces relations de passage.

Afin de déterminer les relations de passage, il est utile de déformer la représentation tensorielle d'un MPS-OBC afin qu'elle ressemble à un circuit, cf. Fig. 4.5. Ceci permet d'intuiter la relation entre les matrices $A^{[k]z_k} = \sum_{\alpha\beta} A^{[k]z_k}_{\alpha\beta} |\alpha\rangle\langle\beta|$ et les portes quantiques $U^{[k]}$. Formellement, on veut écrire l'état sous la forme $|\psi\rangle = \prod_{k=1}^{n} U^{[k]} |0\rangle^{\otimes n}$. Il suffit de remarquer que

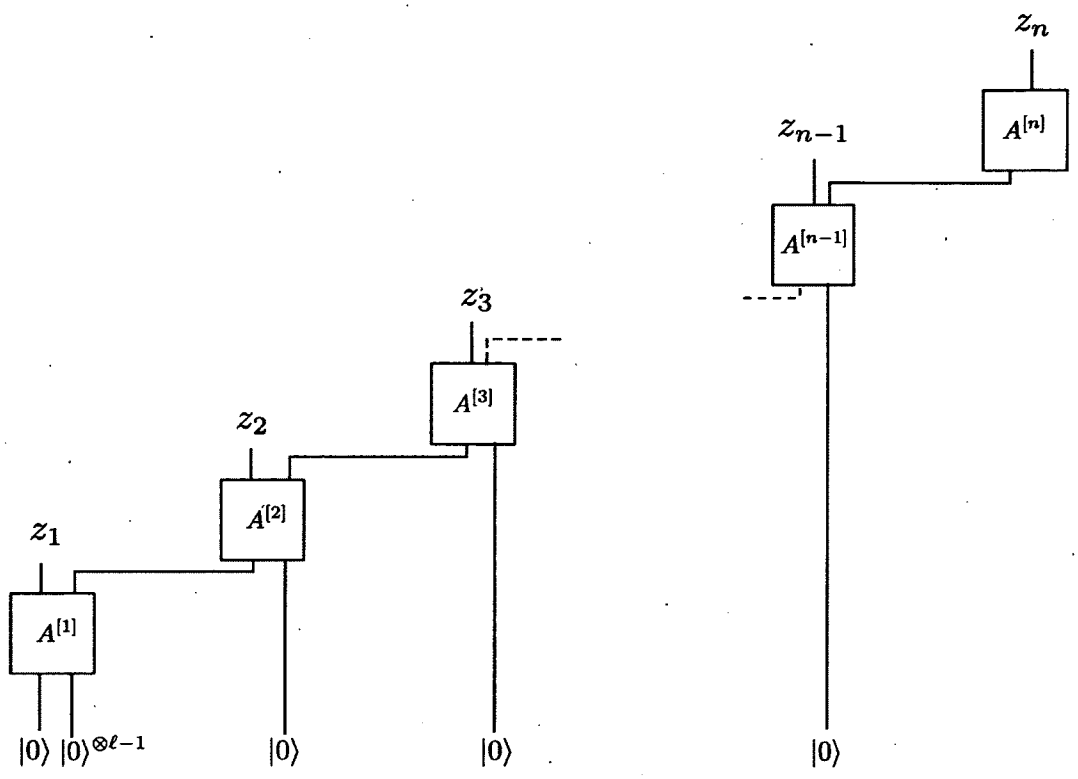


FIGURE 4.5 Transformation graphique de la représentation tensorielle MPS vers un circuit préparateur en escalier. Il s'agit d'une déformation de la représentation tensorielle habituelle des MPS donnée à la Fig. 4.1 c)

$$\langle z_1 \dots z_n|\psi\rangle = \langle z_1 \dots z_n| \prod_{k=1}^{n} U^{[k]} |0 \dots 0\rangle \tag{4.13}$$

$$= \langle z_2 \dots z_n| \prod_{k=2}^{n} U^{[k]} |0_{\ell+1} \dots 0_n\rangle \sum_{y_2} |y_2\rangle\langle y_2|\langle z_1|U^{[1]}|0_1 \dots 0_\ell\rangle \tag{4.14}$$

$$= \sum_{y_2,\dots y_n} \langle z_n|U^{[n]}|y_n\rangle \dots \langle y_{k+1}, z_k|U^{[k]}|y_k 0_{k+1}\rangle \dots \langle y_2, z_1|U^{[1]}|0_1 \dots 0_\ell\rangle \tag{4.15}$$

ce qui permet d'identifier les tenseurs MPS

$$A^{[1]z_1}_{y_2} = \langle y_2, z_1|U^{[1]}|0_1 \dots 0_\ell\rangle \tag{4.16}$$

$$\forall 2 \leq k \leq n-\ell+1 \quad A^{[k]z_k}_{y_k,y_{k+1}} = \langle y_{k+1}, z_k|U^{[k]}|y_k 0_{k+1}\rangle \tag{4.17}$$

$$\forall k > n-\ell+1 \quad A^{[k]z_k}_{y_k,y_{k+1}} = \langle y_{k+1}, z_k|U^{[k]}|y_k\rangle \tag{4.18}$$

$$A^{[n]}_{y_n} = \langle z_n|U^{[n]}|y_n\rangle \tag{4.19}$$

Graphiquement, l'éq. (4.17) est représentée sur la figure 4.6. En fait, il ne s'agit que d'un agrandisse-

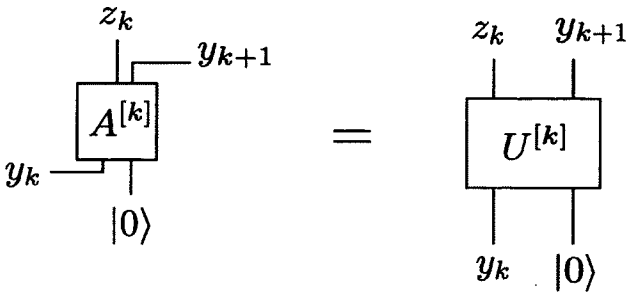


FIGURE 4.6 Représentation graphique de la relation entre tenseur MPS et porte quantique

ment d'une partie de la figure 4.5. Notons qu'on aurait aussi pu construire un circuit où les portes agissent d'abord sur les derniers qudits. On obtient alors une relation différente entre tenseur MPS et portes quantiques.

Nous venons de montrer qu'il est possible d'écrire une représentation tensorielle MPS pour un état préparé par un circuit en escalier. Réciproquement, il suffit d'inverser ces relations [6] pour obtenir un circuit quantique qui produit un état MPS.

La représentation en circuit sera au coeur de notre proposition de méthode tomographique pour les MPS. L'idée sera alors de progressivement apprendre le circuit avec des mesures sur un petit nombre de qudits et de progressivement défaire le circuit afin de récupérer des qudits dans l'état $|0\rangle$. On peut alors reconstruire les tenseurs MPS en inversant les relations (4.16) à (4.19).

6. Ces relations ne suffisent pas à contraindre complètement les transformations unitaires. Il reste donc beaucoup de liberté dans le choix des portes.

4.3 Article : « Efficient quantum state tomography »

4.3.1 Genèse et contribution

Dans cette section, nous reproduirons l'article

Efficient quantum state tomography,
M. Cramer, M. B. Plenio, S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma, O. Landon-Cardinal, Y-K. Liu, D. Poulin
Nature Communications, **1** (9) 149, (2010)

qui propose une méthode d'apprentissage des MPS.

La genèse de cet article est particulière : il résulte de la fusion de trois proto-articles provenant du travail indépendant de trois équipes. Ces proto-articles sont disponibles sur arXiv :

Reconstructing quantum states efficiently
M. Cramer, M. B. Plenio
arXiv:1002.3780

Heralded Polynomial-Time Quantum State Tomography
S. T. Flammia, D. Gross, S. D. Bartlett, R. Somma
arXiv:1002.3839

Efficient Direct Tomography for Matrix Product States
O. Landon-Cardinal, Y-K. Liu, D. Poulin
arXiv:1002.4632

L'article est donc une présentation unifiée des résultats de ces trois proto-articles.

Les deux premiers proposaient une méthode d'apprentissage des MPS basée sur les idées de compressed sensing et l'existence d'un hamiltonien parent correspondant à un MPS, voir annexe A.3. L'idée est donc d'obtenir des matrices densité sur chaque bloc de quelques particules par tomographie standard. À l'aide de ces matrices densité, il est possible de reconstruire un hamiltonien dont l'état expérimental est un fondamental.

Notre proto-article propose une approche légèrement différente. L'idée est de progressivement identifier une porte unitaire du circuit préparateur, d'appliquer son inverse physiquement sur l'état

expérimental afin de désintriquer une particule. Cette procédure est ensuite appliquée récursivement afin de désintriquer toute la chaîne de particules. Afin d'identifier une porte unitaire, il suffit d'effectuer des relevés tomographiques sur un bloc de quelques particules.

La distinction entre les deux approches est le compromis entre ressources qui est différent. L'approche « hamiltonien parent » est très exigeante numériquement, mais ne demande que des mesures locales expérimentales. Au contraire, notre approche demande un traitement numérique trivial (diagonaliser des matrices de taille constante), mais exige d'appliquer physiquement les transformations unitaires sur l'état expérimental et demande donc un contrôle unitaire. Notons toutefois qu'il est possible de contourner le contrôle unitaire grâce aux idées développées dans l'article sur l'apprentissage des MERA, cf. section 4.6.

Ma contribution concerne particulièrement la méthode basée sur les transformations unitaires, présentée dans la section II.A. de l'article. En particulier, j'ai formalisé et écrit la section IV. A. sur l'analyse d'erreur de cette méthode, en collaboration avec David. J'ai aussi écrit le second paragraphe de l'exemple des cluster states (II.C.).

4.3.2 Article

Efficient quantum state tomography

Marcus Cramer and Martin B. Plenio
Institut für Theoretische Physik, Albert-Einstein Allee 11, Universität Ulm, 89069 Ulm, Germany

Steven T. Flammia and Rolando Somma
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5 Canada

David Gross
Institute for Theoretical Physics, Leibniz University Hannover, 30167 Hannover, Germany

Stephen D. Bartlett
School of Physics, The University of Sydney, Sydney, New South Wales 2006, Australia

Olivier Landon-Cardinal and David Poulin
Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, Canada

Yi-Kai Liu
Institute for Quantum Information, California Institute of Technology, Pasadena, CA, USA

Quantum state tomography [1], the ability to deduce the state of a quantum system from measured data, is the gold standard for verification and benchmarking of quantum devices. It has been realized in systems with few components [2–7], but for larger systems it becomes infeasible because the number of quantum measurements and the amount of computation required to process them grows exponentially in the system size. Here we show that we can do exponentially better than direct state tomography for a wide range of quantum states, in particular those that are well approximated by a matrix product state ansatz. We present two schemes for tomography in 1-D quantum systems and touch on generalizations. One scheme requires unitary operations on a constant number of subsystems, while the other requires only local measurements together with more elaborate post-processing. Both schemes rely only on a linear number of experimental operations and classical postprocessing that is polynomial in the system size. A further strength of the methods is that the accuracy of the reconstructed states can be rigorously certified without any *a priori* assumptions.

I. INTRODUCTION

One of the principal features distinguishing classical from quantum many-body systems is that quantum systems require exponentially many parameters in the system size to fully specify the state, compared to only linearly many for classical systems. Put to use constructively, the exponential complexity enables the construction of information processing devices fundamentally superior to any classical device. At the same time, however, this "curse of dimensionality" makes engineering tasks—such as verifying that the quantum processing device functions as intended—a daunting challenge.

The full determination of the quantum state of a system, known as quantum state tomography [1], has already been achieved by measuring a complete set of observables whose expectation values determine the quantum state [2–7]. As it is typically formulated [8], simply to output an estimate for a generic state would take exponential time in the system size N, given that there are an exponential number of coefficients in a generic state's description. This is but one of several inefficiencies. Most quantum states have exponentially small amplitudes in almost every basis, so to distinguish any one of those amplitudes from zero, one must take an exponential number of samples. Assuming one were able to collect all the data from an informationally complete measurement, one is left with the intractable computational task of inverting the measured frequencies to find an estimate of the state.

However, the traditional representation of quantum states is in a sense too general. Indeed, states which occur in many practical situations are specified by a small number of parameters. An efficient description could be a practical preparation scheme which outputs the state; or, in the case of thermodynamical equilibrium states, a local Hamiltonian and a temperature. This insight is not new: researchers in many-body physics and quantum information theory have found many classes of states which are described by a number of parameters scaling polynomially in N [9–13] and which closely approximate the kind of states found in physical systems [14, 15]. However, the question of whether these restricted classes can be put to use in the context of tomography has remained largely open.

In this work, we address the above challenge. The physical system we have in mind is one where the constituents are arranged in a one-dimensional configuration (e.g., ions in a linear trap [3]). But the methods that we are presenting here can be generalized to higher-dimensional arrangements such as those realized in optical lattices [16]. It is highly plausible that in such a set-

ting, correlations between neighboring particles are much more pronounced (due to direct interaction) than correlations between distant systems (mediated e.g. by global fluctuations of control fields). An efficiently describable class of states anticipating exactly this behavior has long been studied under the names of finitely correlated states (FCS) or matrix product states (MPS) [9, 10]. Importantly, restricting to this class is not a limitation since *every* state may be written as a MPS with a suitable, albeit possibly large, matrix dimension. Since many states that are relevant for quantum information processing or quantum many-body physics have a small (independent of N) bond dimension, our methods are directly applicable to such states; examples include, but are not limited to, ground and thermal states, the GHZ, W, cluster, and AKLT states, the latter two being universal resources states for quantum computing.

Given that standard tomography is no longer feasible in a range of recent and upcoming experiments involving large numbers of qubits, our results represent a significant advance in the ability to verify and quantitatively and efficiently benchmark systems of experimental importance.

II. RESULTS

In the following we present two schemes for identifying systems that are well approximated by an MPS, initially focusing on pure states for simplicity. We will view each system as consisting of a linear chain of N qudits, each having dimension d. Both schemes require the measurement of linearly (in the system size N) many local observables within finite accuracy, polynomial classical post-processing of the data and can certify the accuracy of the reconstructed state without making any technical assumptions about the state in the laboratory. The first scheme requires unitary control and local measurements while the second scheme removes the need for unitary control at the cost of more elaborate post-processing.

A. Scheme based on unitary transformations

The key idea of the method consists of a sequential procedure to *disentangle* the left hand side of the chain from the right hand side, using a sequence of unitary operations with small interaction length independent of N. The result will be a product state and a sequence of local unitary operations from which to construct the original state.

Suppose the ideal state in the laboratory is $\hat{\varrho} = |\phi\rangle\langle\phi|$, which we assume for clarity is a MPS of given bond dimension. This implies that the rank of reductions to one part of a bipartite (left vs. right) split of the chain is bounded by a constant R. The protocol starts by estimating, through standard state tomography, the reduced density matrix of the first $\kappa = \lceil \log_d(R) \rceil + 1$ sites,

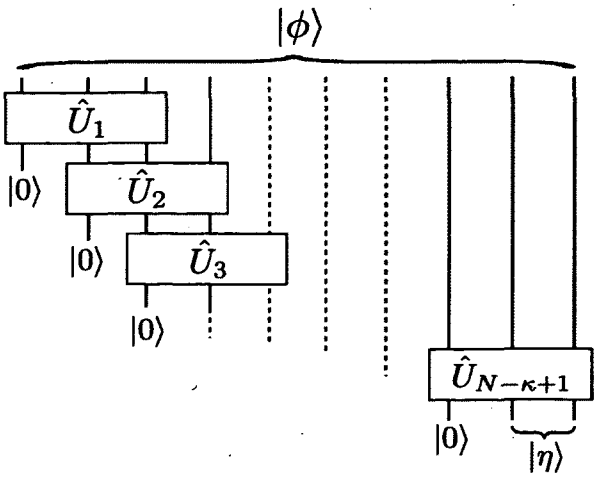


FIG. 1. Quantum circuit that transforms $|\phi\rangle$ into $|0\rangle^{\otimes N-\kappa+1} \otimes |\eta\rangle$ with $\kappa = 3$. The unitaries $\hat{U}_i$ successively disentangle the particles and the state $|\eta\rangle$ on the last sites acts as a boundary condition to determine the MPS description of $|\phi\rangle$.

$\hat{\sigma} \approx \mathrm{tr}_{\kappa+1,\ldots,N}[\hat{\varrho}]$. This reduced density matrix has the eigendecomposition $\hat{\sigma} = \sum_{r=1}^{R} \sigma_r |\phi_r\rangle\langle\phi_r|$ where the rank $R \leq d^{\kappa-1}$. Hence there exists a density matrix with one fewer qudit that has the same rank R and eigenvalues σ_r as $\hat{\sigma}$. Therefore we can disentangle the first site in $\hat{\sigma}$ with the following unitary acting on the first κ sites:

$$\hat{U} = \sum_{s=0}^{d-1} \sum_{s'=0}^{d^{\kappa-1}-1} |s\rangle_1 \otimes |s'\rangle_{2,\ldots,\kappa} \langle \phi_{sd^{\kappa-1}+s'+1}|_{1,\ldots,\kappa}, \quad (1)$$

where $|\phi_1\rangle, \ldots, |\phi_R\rangle$ have been extended in some arbitrary way to get a complete basis for sites $1, \ldots, \kappa$. Applying $\hat{U}$ produces the state $\hat{U}|\phi\rangle = |0\rangle_1 \otimes |v\rangle_{2,\ldots,N}$, where $|v\rangle$ is some pure state on sites $2, \ldots, N$. Hence $\hat{U}$ disentangles the first qudit from all the others. Now, set aside this first qudit, look at sites $2, \ldots, \kappa+1$, and repeat the above process as indicated on Fig. 1. In this way, one obtains a sequence of unitaries $\hat{U}_1, \ldots, \hat{U}_{N-\kappa+1}$, where each $\hat{U}_i$ acts on sites $i, \ldots, i+\kappa-1$. This sequence transforms $|\phi\rangle$ into $\hat{U}_{N-\kappa+1} \cdots \hat{U}_1 |\phi\rangle = |0\rangle^{\otimes N-\kappa+1} \otimes |\eta\rangle$, where $|\eta\rangle$ is some pure state on the last $\kappa - 1$ sites.

In summary, this scheme infers the quantum circuit used to prepare an MPS [17]. The MPS decomposition of $|\phi\rangle$ can then be obtained readily from the U_i and $|\eta\rangle$ [18]. If the state in the laboratory $\hat{\varrho}$ is arbitrary, then the reduced density matrices $\hat{\sigma}$ will generally have full rank. Hence in each step we will need to truncate the κ qudit state $\hat{\sigma}$ to a rank R subspace with $R < d^{\kappa-1}$. Then the above method will produce an MPS approximation to $\hat{\varrho}$. The accuracy of this estimate can be certified, without any assumptions on the state, by keeping track of the effects of truncating each of the reduced states $\hat{\sigma}$. As shown in the Methods, errors of magnitude ϵ due to finite measurement precision or truncation error (as measured

by the weight of the truncated space) accumulate at most linearly with the number of sites, and can be evaluated directly from the data, leading to an estimate of accuracy $N\epsilon$.

The present scheme requires unitary control of κ neighboring qudits, which is challenging to implement in many current experimental settings. We now present a second scheme that avoids unitary control and requires only local measurements.

B. Scheme based on local measurements: Certification of estimated state

Consider again the state $\hat{\varrho}$ in the laboratory and suppose that a tomographically-complete set of local measurements on all groups of k neighbouring qubits has been performed. Further suppose that sufficient data has been taken to yield estimates $\hat{\sigma}_i$ of the local reductions $\hat{\varrho}_i = \mathrm{tr}_{1,\ldots,i;i+k+1,\ldots,N}[\hat{\varrho}]$ such that

$$\|\hat{\varrho}_i - \hat{\sigma}_i\|_{\mathrm{tr}} \leq \epsilon_i \,. \tag{2}$$

Determining these approximate reduced density operators $\hat{\sigma}_i$ completes the experimental work. In the remainder we describe the classical post-processing that will result in a MPS estimate $|\psi\rangle$ to $\hat{\varrho}$ and a lower bound to the fidelity $\langle\psi|\hat{\varrho}|\psi\rangle$ [19] that does *not* require assumptions on the nature of the state $\hat{\varrho}$.

We start with the latter, since the following easy calculation yields the fidelity bound we have in mind and hints at the MPS estimate we are after. Suppose, for the sake of the argument, that $|\psi\rangle$ is the unique ground state (with energy zero) of a local Hamiltonian $\hat{H} = \sum_i \hat{h}_i$, where the $\hat{h}_i$ is a projection operator acting only on sites $i+1,\ldots,i+k$ (as it turns out, generic MPS are of this type). Then, expanding in the eigenbasis $\hat{H} = \sum_{n=0}^{2^N-1} E_n |E_n\rangle\langle E_n|$,

$$\mathrm{tr}[\hat{H}\hat{\varrho}] \geq \Delta E \sum_{n>0} \langle E_n|\hat{\varrho}|E_n\rangle = \Delta E(1 - \langle\psi|\hat{\varrho}|\psi\rangle), \tag{3}$$

where we denoted by ΔE the energy gap above the ground state $|\psi\rangle$. Hence, we have the fidelity bound

$$\langle\psi|\hat{\varrho}|\psi\rangle \geq 1 - \frac{\sum_i \mathrm{tr}[h_i\hat{\rho}]}{\Delta E} \geq 1 - \frac{1}{\Delta E}\Big(\sum_i \mathrm{tr}[\hat{h}_i\hat{\sigma}_i] + \epsilon_i\Big). \tag{4}$$

In other words, the Hamiltonian acts as a *witness* for its ground state $|\psi\rangle$. This bound is tight: suppose the experimental estimates $\hat{\sigma}_i$ and the reductions of the state in the laboratory $\hat{\varrho}_i$ coincide, that is $\epsilon_i = 0$. If, in addition, the reductions of $|\psi\rangle$ match the $\hat{\sigma}_i$ then, as $|\psi\rangle$ was assumed to be the unique ground state with energy zero of $\sum_i \hat{h}_i$, we have $\sum_i \mathrm{tr}[\hat{h}_i\hat{\sigma}_i] = 0$, i.e., $\langle\psi|\hat{\varrho}|\psi\rangle = 1$.

The goal is now clear: find a local gapped Hamiltonian $\hat{H}$ such that the reductions of its ground state are close to the $\hat{\sigma}_i$. *A priori* it is unclear whether such a convenient witness always exists and how it could be found efficiently. However, using formal methods, one can show that if the true state $\hat{\varrho}$ is close to a generic MPS, then such a witness Hamiltonian always exists [10]. What is more, it can be constructed from the estimate of the algorithm sketched below. Its properties, chief among them the gap, are efficiently computable. In the Methods section, we detail the efficient computation of these quantities, and we also consider how to treat states such as the GHZ for which local marginals alone are not quite sufficient for complete characterization (they violate our "generic" condition).

C. An illustrative example

We illustrate how our certification procedure operates if our estimate for the state $\hat{\varrho}$ is a linear cluster state [20]. The cluster state is defined as the unique eigenstate (with eigenvalue $+1$) of stabilizers $\hat{K}_i = \hat{Z}_{i-1}\hat{X}_i\hat{Z}_{i+1}$ for all $i = 2,\ldots,N-1$ (together with boundary terms $\hat{X}_1\hat{Z}_2$ and $\hat{Z}_{N-1}\hat{X}_N$, which we do not treat separately for simplicity). Assume that we have performed standard quantum state tomography on sets of three neighbouring spins, $k = 3$, which is the smallest useful set because in a cluster state the rank of the reduced density matrices of contiguous blocks is upper bounded by $R = 4$. Let us now assume that on the basis of these data, our procedure suggests that the linear cluster state is indeed our estimate. The local Hamiltonian in this case is given by $\hat{H} = \sum_i(\mathbb{1} - \hat{K}_i)/2$, where the $(\mathbb{1} - \hat{K}_i)/2 = \hat{h}_i$ are projectors, $\hat{H}$ has the cluster state as its unique ground state (with energy zero) and an energy gap $\Delta E = 1$. The fidelity between the cluster state $|\psi_{CS}\rangle$ and the state $\hat{\varrho}$ is bounded by $\langle\psi_{CS}|\hat{\varrho}|\psi_{CS}\rangle = 1 - \sum_i(\mathrm{tr}[\hat{h}_i\hat{\sigma}_i] + \epsilon_i)$, where ϵ_i from Eq. (3) quantifies the statistical and experimental error in the local experimental estimates, and $\mathrm{tr}[\hat{h}_i\hat{\sigma}_i]$ quantifies how much the laboratory state $\hat{\varrho}$ deviates from an exact cluster state.

Tomography on a cluster state can also be performed with the scheme based on unitary transformations. A cluster state is the output of a quantum circuit where each qubit is initially prepared in the state $|+\rangle = 1/\sqrt{2}\,(|0\rangle + |1\rangle)$ and a controlled phase transformation CZ acts successfully on each pair of neighboring qubits. The CZ gate changes the sign of state $|11\rangle$ and acts trivially on the other states of the computational basis. Thus, a cluster state is the output of a circuit whose structure corresponds to the one indicated on Fig. 1 with $\kappa = 2$ and the scheme based on unitary transformations directly applies. Note that the unitary scheme takes advantage of the decreased rank of a reduced density matrix on the *boundary* to save one qubit worth of local tomographic effort ($\kappa = 2$ vs. $k = 3$).

D. Scheme based on local measurements: Efficient determination of an MPS estimate

With the experimentally obtained $\hat{\sigma}_i$, we now turn to the task of finding an MPS $|\psi\rangle$ such that its reductions $\mathrm{tr}_{1,\ldots,i;i+k+1,\ldots,N}[|\psi\rangle\langle\psi|]$ closely match the $\hat{\sigma}_i$. In other words: Let $\hat{P}_m^{i,k}$ be all possible products of Pauli operators (enumerated by m) that act non-trivially only on sites $i+1,\ldots,i+k$. Then, as the $\hat{P}_m^i$ are an orthogonal basis, the $\hat{\sigma}_i$ may be expanded as

$$\hat{\sigma}_i = \frac{1}{2^N}\sum_m \mathrm{tr}[\hat{\sigma}_i \hat{P}_m^i]\hat{P}_m^i, \tag{5}$$

where the expectations $\mathrm{tr}[\hat{\sigma}\hat{P}_m^i] \in \mathbb{R}$ are obtained as results of tomographic measurements. Then we need to find a matrix $|\psi\rangle\langle\psi|$ such that for all m and i the expectations $\mathrm{tr}[|\psi\rangle\langle\psi|\hat{P}_m^i]$ coincide with those of the tomographic estimates, $\mathrm{tr}[|\psi\rangle\langle\psi|\hat{P}_m^i] = \mathrm{tr}[\hat{\sigma}_i\hat{P}_m^i]$.

The method of choice for such a problem is singular value thresholding (SVT) [21, 22], which has been developed very recently in the context of classical *compressive sampling* or *matrix completion* [23] and may also be applied to the quantum setting [24–28]. SVT is a recursive algorithm that provably converges to a low-rank matrix satisfying constraints of the type $\mathrm{tr}[|\psi\rangle\langle\psi|\hat{P}_m^i] = \mathrm{tr}[\hat{\sigma}_i\hat{P}_m^i]$. Unfortunately, a straightforward application of SVT requires time and memory that scale exponentially with the number of particles. However, a modification of the algorithm allows us to overcome this problem.

Given the measured values $\mathrm{tr}[\hat{\sigma}_i\hat{P}_m^i]$ the algorithm may then be described as follows. First set up the operator $\hat{R} = \sum_{m,i}\mathrm{tr}[\hat{\sigma}_i\hat{P}_m^i]\hat{P}_m^i/2^N$ and initialize $\hat{Y}_0$ to some arbitrary matrix (e.g., the zero matrix). Then proceed recursively by finding the eigenstate $|y_n\rangle$ with largest eigenvalue, y_n, of $\hat{Y}_n$ and set

$$\begin{aligned}\hat{X}_n &= y_n \sum_{m,i}\frac{\langle y_n|\hat{P}_m^i|y_n\rangle}{2^N}\hat{P}_m^i,\\ \hat{Y}_{n+1} &= \hat{Y}_n + \delta_n(\hat{R}-\hat{X}_n).\end{aligned} \tag{6}$$

So far, this algorithm still suffers from the fact that in every step the $2^N \times 2^N$ matrix $\hat{Y}_n$ needs to be diagonalized. However, the $\hat{Y}_n$ are of the form $\sum_{m,i}^M a_{m,i}\hat{P}_m^i$, $a_{m,i} \in \mathbb{R}$, where the $\hat{P}_m^i$ act non-trivially only on sites $i+1,\ldots,i+k$, i.e., they have the form of a local "Hamiltonian". Hence, $|y_n\rangle$ can be determined as the highest energy state of this Hamiltonian. For this task standard methods have been developed in condensed matter physics [13, 29] for which the number of parameters scale polynomially in the system size and converge rapidly [30, 31] to the optimal MPS approximation. The standard but exponentially inefficient SVT algorithm possesses a convergence proof while our efficient modification does not. Hence, we now present numerical examples for different target states $|\phi\rangle$ to demonstrate the feasibility and efficiency of the proposed algorithm.

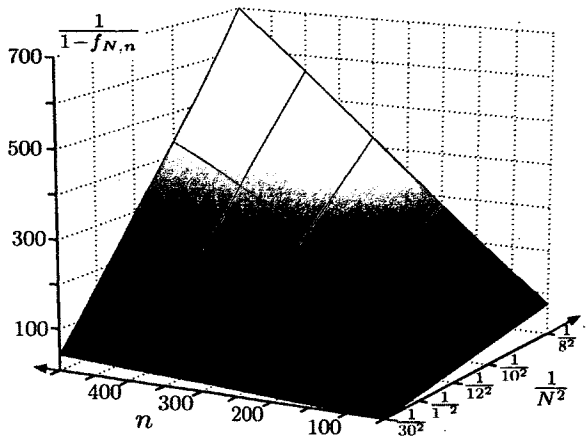


FIG. 2. Performance of the MPS-SVT algorithm for the ground state (the target state $|\phi\rangle$) of the critical Ising model. We chose $k = 2$, i.e., only nearest neighbour reductions are used to reconstruct the state. Plot illustrates the scaling of the error in the fidelity $1 - f_{N,n} = 1 - |\langle\phi|y_n\rangle|^2 \sim N^2/n$ with the number of spins N and the iterations n of the algorithm.

These numerical examples suggest convergence of our algorithm to a MPS that closely matches the experimentally obtained reductions $\hat{\sigma}_i$. To arrive at the fidelity bound we follow the steps of Section II B: (i) Obtain estimates $\hat{\sigma}_i$ of the reductions to k adjacent spins $\hat{\varrho}_i$ of the state in the laboratory such that $\|\hat{\sigma}_i - \hat{\varrho}_i\|_{\mathrm{tr}} \leq \epsilon_i$, (ii) compute the expectations $p_{m_i} = \mathrm{tr}[\hat{\sigma}_i\hat{P}_m^i]$, which are the input to the MPS-SVT algorithm, (iii) obtain a MPS estimate $|y_n\rangle$, the reductions of which closely match the $\hat{\sigma}_i$ by utilizing the MPS-SVT algorithm. As $|y_n\rangle$ is an MPS, one can then efficiently obtain a *parent Hamiltonian* [10] and a lower bound, Δ, on the energy gap above the ground state (see Methods for details). Putting all this together, the above programme returns a state $|y_n\rangle$, its parent Hamiltonian $\hat{H} = \sum_i \hat{h}_i$, and a number Δ such that

$$\langle y_n|\hat{\varrho}|y_n\rangle \geq 1 - \frac{\sum_i(\epsilon_i + \mathrm{tr}[\hat{h}_i\hat{\sigma}_i])}{\Delta}. \tag{7}$$

E. Example: Strongly interacting quantum systems

We start with ground states of nearest-neighbor Hamiltonians on a chain, i.e., the $|\phi\rangle = |gs\rangle$ are completely determined by all the reductions to two adjacent spins and we expect the above algorithm not only to produce states that match the reduced density matrices of the ground states but, in fact, states that are themselves close to the ground states. Among ground states of one-dimensional nearest-neighbor Hamiltonians, those at a critical point are the most challenging to approximate by MPS as they violate the entanglement area law [32]. We demonstrate

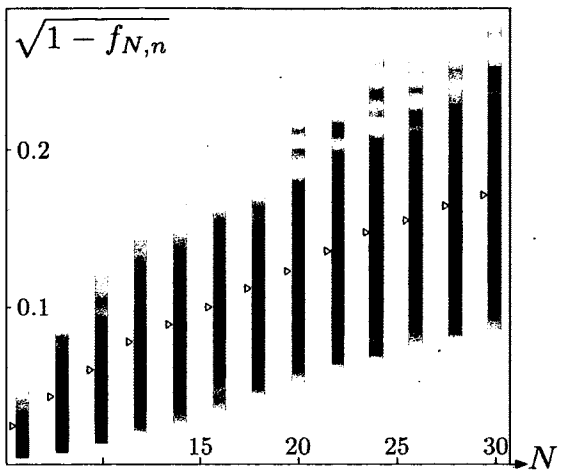


FIG. 3. Fidelity $f_{N,n} = |\langle\phi|y_n\rangle|^2$, for a fixed value $n = 5$, as a function of the number of spins N for the ground state (the target state $|\phi\rangle$) of the random nearest-neighbour Hamiltonians of Eq. 8. As in Fig. 2, we chose $k = 2$ to reconstruct the state. The plot shows densities (arrows indicate the mean) obtained from 1000 random realizations. Similar to the Ising model, the scaling for fixed n of $1 - f_{N,n}$ is better than $\sim N^2$.

the effectiveness of our algorithm for such an example: the critical Ising model in a transverse field on a chain of length N with open boundary conditions. The ground state of this Hamiltonian is unique and hence it is completely characterized by its reductions to $k = 2$ neighbouring sites. In other words, if we find an MPS that has the same reductions, the fidelity will be one. We proceed as follows. (i) We obtain the reductions $\hat{\varrho}_i$ to two neighbouring sites of the true ground state, (ii) from these reductions we obtain the expectations $p_{m_i} = \mathrm{tr}[\hat{\varrho}_i \hat{P}_{m_i}]$, which are the input to the MPS-SVT algorithm. In Fig. 2 we show the fidelity $f_{N,n} = |\langle y_n|\phi\rangle|^2$ of the true ground state $|\phi\rangle$ and the n'th iterate of the above algorithm as a function of n and the length N of the chain. It shows that for fixed system size $1 - f_{N,n}$ decreases as $\sim 1/n$ while for fixed n it increases slower than N^2. This provides an indication that our algorithm is polynomial in the system size.

The Ising model is solvable and, in order to show that we are not considering a special case that is particularly favourable, we also consider one-dimensional random Hamiltonians of the form

$$\hat{H} = \sum_{i=1}^{N-1} \hat{r}_i^{(i)} \hat{r}_{i+1}^{(i)}, \tag{8}$$

where the $\hat{r}_i^{(i)}$, $\hat{r}_{i+1}^{(i)}$ act on spin i and $i+1$, respectively, and are hermitian matrices with entries that have real and imaginary part picked from a uniform distribution over $[-1, 1]$. For each Hamiltonian, as before, we first determine the ground state $|gs\rangle$ (our target state $|\phi\rangle$) and its reductions and then computed the fidelity $f_{N,n}$ after n iterations of the MPS-SVT algorithm, see Fig. 3.

F. Example: W-state preparation in ion traps

Our method is of interest for many situations in which standard tomography will not be feasible. This is the case for the verification of state preparation in experiments with too many particles. An example is the recent ion trap experiment [3] for the preparation of W-states, $|\phi\rangle = (|10\cdots0\rangle + |010\cdots0\rangle + \cdots + |0\cdots01\rangle)/\sqrt{N}$, that were limited to 8 qubits principally because the classical postprocessing of data became prohibitive for longer chains. Here we demonstrate the efficiency of our approach (we are not limited to few ions and demonstrate convergence for up to 20 ions – even higher number of ions are easily accessible due to the MPS alteration of the SVT method) by illustrating how one would postprocess experimentally obtained reduced density matrices to guarantee the generation of $|\phi\rangle$ or a state very close to it. Note that, as in the above spin chain examples, one need only take tomographic data on pairs ($k = 2$) of neighbouring qubits. We mimic experimental noise by adding Gaussian distributed random numbers with zero mean to the p_{m_i}. After initializing the algorithm with $\hat{Y}_0 = \hat{R}$, where we obtain $\hat{R}$ from the MPS representation of $|\phi\rangle$, we use $x_n := \sum_{m_i} |p_{m,i} - \langle y_n|\hat{P}_m^{i,k}|y_n\rangle|$ as a figure of merit for convergence, i.e., after a given number of iterations, we pick the $|y_n\rangle$ with minimum x_n. The result of such a procedure is shown in Fig. 4.

III. DISCUSSION

We have presented two schemes that efficiently produce an MPS description as a tomographic estimate of a quantum state, along with a tight fidelity bound. We emphasize that no assumptions are necessary for our scheme; if no such MPS description exists, this will be evident from the local tomographic data and our schemes will herald a failure. However, the enormous successes of MPS for describing both one-dimensional quantum systems found in nature as well as a host of states relevant to quantum information ensures that our methods will be very useful in practice.

So far we presented the method for pure states and one-dimensional systems. Various generalizations are possible: Our first scheme using unitary control can also handle mixed states corresponding to small ensembles of pure MPS. Suppose we are presented with a state $\hat{\varrho}$ that is a mixture of M pure states, each of which is an MPS having bond dimension D. Then the reduction of $\hat{\varrho}$ to any subsystem has rank at most MD. We can proceed as before, performing unitary operations on blocks of $\kappa = \lceil \log_d(MD) \rceil + 1$ sites, in order to disentangle individual sites from the rest of the chain. At the end of

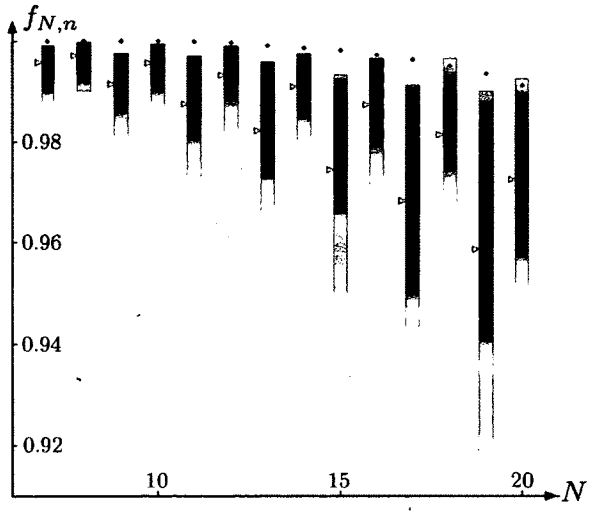


FIG. 4. Performance of the MPS-SVT algorithm for W-states, $|\phi\rangle = (|10\cdots0\rangle + |010\cdots0\rangle + \cdots + |0\cdots01\rangle)/\sqrt{N}$. We are not limited to few ions and demonstrate convergence for up to 20 ions – even higher number of ions are easily accessible due to the MPS alteration of the SVT method, demonstrating the efficiency of our approach. We mimic experimental noise by adding Gaussian distributed random numbers with zero mean to the local expectations $p_{m_i} = \mathrm{tr}[\hat{\sigma}_i \hat{P}_{m_i}]$ and show results for $n = 4000$ MPS-SVT iterations. Plot shows $f_{N,n} = |\langle\phi|y_n\rangle|^2$ as a function of the number of ions, N, for no noise (dots) and Gaussian noise (densities obtained from 100 realizations for each N, arrows indicate mean) with a standard deviation of 0.005 (blue; only even values of N plotted for clarity) and 0.01 (black, odd N).

the chain, we will find a mixed state η of rank M on the last $\kappa - 1$ sites. We decompose η as a mixture of M pure states. This yields a representation of $\hat{\varrho}$ as an ensemble of M pure MPS, each with bond dimension at most dMD.

Our second scheme using local measurements can also be extended to handle mixed states. While the k-particle reduced density matrices do not uniquely determine the mixed state $\hat{\varrho}$, reconstructions of better and better quality can be obtained by increasing k. As an example, suppose $\hat{\varrho}$ is the Gibbs state corresponding to a k-local Hamiltonian $\hat{H}$, i.e. the state minimizing the free energy

$$\mathrm{tr}[\hat{\varrho}\hat{H}] - TS(\hat{\varrho}). \tag{9}$$

The first term is, as before, determined by the reduced density matrices. The entropy of the total state however can only be learned exactly from the complete density matrix. Yet, for essentially all reasonable physical systems, the entropy density $\lim_{k\to\infty} S(\mathrm{tr}_{k+1,\ldots}(\hat{\varrho}))/k$ in the thermal state of a Hamiltonian exists [33] (In particular, finitely correlated states are precisely those states whose entropy density is exactly equal to $S(\mathrm{tr}_{k+1,\ldots}(\hat{\varrho})) - S(\mathrm{tr}_{k,\ldots}(\hat{\varrho}))$ for some finite value of k). As a consequence, the total entropy of the state can be estimated efficiently from knowledge of the reduced density matrices. Hence, our second scheme using local measurements may be extended to mixed states by considering purifications and can then, e.g., also handle thermal states of local Hamiltonians, under the physically reasonable assumption that the entropy density exists [33]. In addition, it can be generalized to all mixed finitely correlated states (FCS) [9], though it is not always possible to certify the resulting estimates.

Higher-dimensional systems are more challenging, because the most straightforward generalization of MPS, known as projected entangled-pair states (PEPS) [34], cannot be computed as efficiently. However, the certification method using a frustration-free parent Hamiltonian remains efficient in the case of qubits with nearest-neighbor couplings [35]. Combinations of our techniques can be used to reconstruct other classes of states, such as tree tensor networks [36] and multiscale entanglement renormalization ansatz (MERA) states [37], for which efficient heuristics for minimizing local Hamiltonians are available.

ACKNOWLEDGMENTS

The authors acknowledge discussion with F.G.S.L. Brandão at early stages of this project. The work at Ulm University has been supported by the EU Integrated Project QAP, the EU STREP's CORNER and HIP and an Alexander von Humboldt Professorship. SDB acknowledges the support of the Australian Research Council and the Perimeter Institute. STF and RS were supported by the Perimeter Institute for Theoretical Physics. Research at Perimeter is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. DG is glad to acknowledge support from the EU (CORNER). OLC and DP are partially funded by NSERC and FQRNT. YKL is funded by the US ARO/NSA.

IV. METHODS

We start by recalling the MPS representation of a state $|\psi\rangle$ with open boundary conditions (generalisations to periodic boundary conditions are entirely straightforward).

$$|\psi\rangle = \sum_{s_1=0}^{d_1-1} \cdots \sum_{s_N=0}^{d_N-1} M_1[s_1]\cdots M_N[s_N]|s_1\cdots s_N\rangle, \tag{10}$$

where the $M_i[s]$ are $D_i \times D_{i+1}$ matrices with $D_1 = D_{N+1} = 1$. We denote the *bond dimension* by $D = \max D_i$.

A. Direct tomography

This method proceeds by disentangling all the qudits of the chain sequentially. Thus, it yields a valid MPS

description if every unitary exactly disentangles one qudit. Put another way, while it is crucial to obtain a good estimate of the $d^{\kappa-1}$-dimensional subspace on which $\hat{\sigma}$ is supported, it is *not necessary* to identify the eigenvectors of $\hat{\sigma}$ exactly : *any* set of orthonormal vectors generating the subspace is sufficient for our tomography procedure and leads to an MPS description in another gauge [10]. This property will be central to our error analysis.

To understand the effect of errors and imperfections in our tomography procedure, consider the very first step of the recursive procedure. Tomography is performed on the first κ sites to ideally find a state with non-maximal support, and unitary $\hat{U}_1$ is applied to rotate that state into the subspace $\mathcal{H}_1^{\text{cutoff}} = |0\rangle \otimes (\mathbb{C}^d)^{\otimes n-1}$. In any experimental setting, the resulting state $\hat{U}_1|\phi\rangle$ would surely not lie entirely in that subspace. This can be either because the state of the system is not exactly an MPS with bond dimension D, but merely close to one or because our estimate of the density matrix $\hat{\sigma}$ on sites 1 to κ is slightly wrong due to measurements that are, in practice, noisy and restricted to finite precision. Indeed, we can expect that the reduced density matrix on the first κ sites will actually be full-rank, though most of its probability mass will lie on a subspace of dimension at most D. So, each time we apply a disentangling operation $\hat{U}_i$, we also want to *truncate* the reduced state to the desired subspace. Similarly, a faulty estimate of $\hat{\sigma}$ will result in a small probability mass that lies outside the estimated support.

Given an estimated disentangling unitary $\hat{U}_1$, any state $|\phi\rangle$ can be expressed as

$$\hat{U}_1|\phi\rangle = \frac{|0\rangle \otimes |\eta_1\rangle + |e_1\rangle}{\sqrt{1 + \langle e_1|e_1\rangle}} \tag{11}$$

where $|e_1\rangle$ is some sub-normalized error vector supported on the subspace orthogonal to $\mathcal{H}_1^{\text{cutoff}}$. The unitary $\hat{U}_1$ is chosen to minimize the norm $\epsilon_1 = \langle e_1|e_1\rangle$ of this error vector. It is possible to directly estimate the magnitude of the error ϵ_1 by measuring the first qudit in the standard basis; the error is equal to the population of the non-zero states.

In subsequent steps of the recursion, we are given a state of the form

$$\hat{U}_i \dots \hat{U}_1|\phi\rangle = \frac{|0\rangle^{\otimes i} \otimes |\eta_i\rangle + |e_i^{\text{cm}}\rangle}{\sqrt{1 + \langle e_i^{\text{cm}}|e_i^{\text{cm}}\rangle}} \tag{12}$$

where $|e_i^{\text{cm}}\rangle$ is the accumulated error vector that lies in the subspace orthogonal to $\mathcal{H}_i^{\text{cutoff}} = |0\rangle^{\otimes i} \otimes (\mathbb{C}^d)^{\otimes N-i}$. As a first step, we can truncate this error vector by measuring the first i particles in the standard basis and post-select on the all-zero outcome. This occurs with a probability roughly $1 - \epsilon_i^{\text{cm}}$, and leaves the system in the state $|0\rangle^{\otimes i} \otimes |\eta_i\rangle$. We then repeat the steps leading to Eq. 11 with the post-selected state $|\eta_i\rangle$. The resulting state will be

$$\hat{U}_{i+1}\hat{U}_i \dots \hat{U}_1|\phi\rangle = \frac{|0\rangle^{\otimes i} \otimes \hat{U}_{i+1}|\eta_i\rangle + \hat{U}_{i+1}|e_i^{\text{cm}}\rangle}{\sqrt{1 + \langle e_i^{\text{cm}}|e_i^{\text{cm}}\rangle}} \tag{13}$$

$$= \frac{|0\rangle^{\otimes i+1} \otimes |\eta_{i+1}\rangle + |e_{i+1}^{\text{cm}}\rangle}{\sqrt{1 + \langle e_{i+1}^{\text{cm}}|e_{i+1}^{\text{cm}}\rangle}} \tag{14}$$

where

$$|e_{i+1}^{\text{cm}}\rangle = \frac{|e_i\rangle + U_{i+1}|e_i^{\text{cm}}\rangle}{\sqrt{1 + \langle e_i^{\text{cm}}|e_i^{\text{cm}}\rangle}} \tag{15}$$

and therefore $\epsilon_{i+1}^{\text{cm}} \leq \epsilon_i^{\text{cm}} + \epsilon_{i+1}$.

Thus, we see that errors accumulate *linearly* with the number of particles; if we denote $|\psi\rangle = \hat{U}_1^\dagger \dots \hat{U}_{N-\kappa+1}^\dagger |0\rangle^{N-\kappa+1}|\eta_{N-\kappa+1}\rangle$ the estimated MPS, we have

$$\big\||\phi\rangle - |\psi\rangle\big\| = \big\||e_{N-\kappa+1}^{\text{cm}}\rangle\big\| \leq \sum_{i=1}^{N-\kappa+1} \big\||e_i\rangle\big\| \leq N\,\epsilon \tag{16}$$

where $\epsilon = \max_i \big\||e_i\rangle\big\|$. The overall error is at most the sum of the individual errors on each step. In addition, each of the ϵ_i is revealed during the tomographic procedure because they correspond to the post-selection success probability. This provides a direct method to *certify* the inferred state.

B. Parent Hamiltonians

Let $|\psi\rangle$ be as in (10) and such that $\sum_s M_i[s]M_i[s]^\dagger = \mathbb{1}$ for all $i = 1, \dots, N$. This can always be achieved by subsuming qudits at the beginning and end of the chain into qudits with higher dimension and successive singular value decompositions [9, 10]. Now let $N, k \in \mathbb{N}$ such that $N/k \in \mathbb{N}$, and assume that the MPS is injective [10] such that for all $j = k, \dots, N - 2k$ the set

$$\Big\{ M_{j+1}[s_1] \cdots M_{j+k}[s_k] \,\Big|\, s_i = 1, \dots, d_i \Big\} \tag{17}$$

spans $\mathbb{C}^{D_{j+1} \times D_{j+k+1}}$. Then $|\psi\rangle$ is the unique ground state of

$$\hat{H} = \sum_{n=0}^{N/k-2} \hat{P}_n, \tag{18}$$

where $\hat{P}_n$ is the projector onto the subspace orthogonal to the range of the mapping $\Gamma_n : \mathbb{C}^{D_{nk+2k+1} \times D_{nk+1}} \to \mathbb{C}^{d_{nk+1} \cdots d_{nk+2k}}$,

$$\Gamma_n(X) = \sum_{\substack{s_{nk+1}, \\ \dots, \\ s_{nk+2k}}} \text{tr}\Big[X M_{nk+1}[s_{nk+1}] \cdots M_{nk+2k}[s_{nk+2k}] \Big] \times |s_{nk+1} \cdots s_{nk+2k}\rangle. \tag{19}$$

To get an efficiently computable lower bound the energy gap, we use

$$\Delta E = \max\left\{\lambda \,\middle|\, \hat{H}(\hat{H}-\lambda) \geq 0\right\}. \tag{20}$$

We find

$$\hat{H}^2 = \hat{H} + \sum_{\substack{n,m \\ n\neq m}} \hat{P}_n\hat{P}_m \geq \hat{H} + \sum_{\substack{n,m \\ 1\leq|n-m|k\leq 2k}} \frac{\hat{P}_n\hat{P}_m + \hat{P}_m\hat{P}_n}{2}, \tag{21}$$

where we omitted non-negative summands. Now consider the following quantity, which will bound the individual terms in the previous equation,

$$\begin{aligned}\gamma_{n,m} &= \min\left\{\lambda \,\middle|\, \hat{P}_n\hat{P}_m + \hat{P}_m\hat{P}_n + \lambda(\hat{P}_n+\hat{P}_m) \geq 0\right\} \\ &= \min\left\{\lambda \,\middle|\, (\hat{P}_n+\hat{P}_m)^2 \geq (1-\lambda)(\hat{P}_n+\hat{P}_m)\right\} \\ &= 1 - \max\left\{\lambda \,\middle|\, (\hat{P}_n+\hat{P}_m)^2 \geq \lambda(\hat{P}_n+\hat{P}_m)\right\},\end{aligned} \tag{22}$$

where the maximum is given by the smallest non-zero eigenvalue of $\hat{P}_n + \hat{P}_m$. Hence

$$\hat{H}^2 \geq \hat{H} - \sum_{\substack{n,m \\ 1\leq|n-m|k\leq 2k}} \gamma_{n,m}\hat{P}_n \geq (1-\gamma)\hat{H}, \tag{23}$$

and therefore we have the lower bound $\Delta E \geq 1-\gamma$, where

$$\gamma = \max_{0\leq n\leq N} \sum_{\substack{m \\ 1\leq|n-m|k\leq 2k}} \gamma_{n,m}. \tag{24}$$

Injectivity may fail to hold in certain singular cases. The simplest example given by the family of GHZ-type states $\frac{1}{\sqrt{2}}(|0,\ldots,0\rangle + e^{i\phi}|1,\ldots,1\rangle)$. Since any local reduced density matrix is independent of ϕ, it is *impossible* to distinguish the members of that family based on local information alone. Indeed, in this example, the ground state space of $\hat{H}$ will be two-dimensional, spanned by $|0,\ldots,0\rangle$ and $|1,\ldots,1\rangle$. One may check that the gap analysis above continues to hold in the degenerate case (unlike the original in [9]), now certifying the overlap between $\hat{\varrho}$ and the ground-state space. This fact alone implies an exponential reduction in the number of unknown parameters. It is easy to see that the small remaining uncertainty about $\hat{\varrho}$ can be lifted in our example by measuring the "string operator" $\sigma_x \otimes \ldots \otimes \sigma_x$, which has expectation value $\cos\phi$. Using e.g. the results of [38], the method of this particular example immediately generalizes to any MPS with a two-fold degeneracy, such as the W state. Higher degeneracies may be treated by straight-forward, but more tedious, methods.

[1] Vogel, K. and Risken, H. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. Phys. Rev. A **40**, 2847–2849 (1989).

[2] Smithey, D.T., Beck, M., Raymer, M.G. and Faridani, A. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. Phys. Rev. Lett. **70**, 1244–1247 (1993).

[3] Häffner, H., Hänsel, W., Roos, C.F., Benhelm, J., Chekal-kar, D., Chwalla, M., Körber, T., Rapol, U.D., Riebe, M., Schmidt, P.O., Becher, C., Gühne, O., Dür, W. and Blatt, R. Scalable multiparticle entanglement of trapped ions. Nature **438**, 643–646 (2005).

[4] Leibfried, D., Knill, E., Seidelin, S., Britton, J., Blakestad, R.B., Chiaverini, J., Hume, D.B., Itano, W.M., Jost, J.D., Langer, C., Ozeri, R., Reichle, R. and Wineland, D.J. Creation of a six-atom 'Schrödinger cat' state. Nature **438**, 639–643 (2005).

[5] James, D.F.V., Kwiat, P.G., Munro, W.J. and White, A.G. Measurement of qubits. Phys. Rev. A **64**, 052312 (2001).

[6] Dunn, T. J. and Walmsley, I. A. and Mukamel, S. Experimental Determination of the Quantum-Mechanical State of a Molecular Vibrational Mode Using Fluorescence Tomography, Phys. Rev. Lett. 74, 884 (1995).

[7] Lvovsky, A.I. and Raymer, M.G. Continuous-variable optical quantum-state tomography. Rev. Mod. Phys. **81**, 299–332 (2009).

[8] Paris, M. and Řeháček, J., eds., *Quantum State Estimation*, no. 649 in Lect. Notes Phys. (Springer, Heidelberg, 2004).

[9] Fannes, M., Nachtergaele, B. and Werner, R.F. Finitely correlated states on quantum spin chains. Comm. Math. Phys. **144**, 443–490 (1992).

[10] Perez-Garcia, D., Verstraete, F., Wolf, M.M and Cirac, J.I. Matrix product state representations. Quant. Inf. Comp. **7**, 401–430 (2007).

[11] Vidal, G. Entanglement Renormalization. Phys. Rev. Lett. **99** 220405 (2007).

[12] Rizzi, M., Montangero, S. and Vidal, G. Simulation of time evolution with multiscale entanglement renormalization ansatz. Phys. Rev. A **77**, 052328 (2008).

[13] Östlund, S. and Rommer, S. Thermodynamic Limit of Density Matrix Renormalization. Phys. Rev. Lett. **75**, 3537–3540 (1995).

[14] Hastings, M.B. Solving gapped Hamiltonians locally. Phys. Rev. B **73**, 085115 (2006).

[15] Hastings, M.B. An area law for one-dimensional quantum systems. J. Stat. Mech. P08024 (2007).

[16] Bloch, I., Dalibard, J. and Zwerger, W., Many-Body Physics with Ultracold Gases, Rev. Mod. Phys. **80**, 885 (2008).

[17] Schoen, C., Hammerer, K., Wolf, M.M., Cirac, J. I. and Solano, E. Sequential Generation of Entangled Multiqubit States. Phys. Rev. A **75**, 032311 (2007).

[18] Vidal, G. Efficient Classical Simulation of Slightly Entangled Quantum Computations. Phys. Rev. Lett. **91**, 147902 (2003).

[19] Gilchrist, A., Langford, A. K., Nielsen, M. A. Distance measures to compare real and ideal quantum processes, Phys. Rev. A 71, 062310 (2005).

[20] Briegel, H. J. and Raussendorf, R. Persistent Entanglement in arrays of Interacting Particles. Phys. Rev. Lett. **86**, 910–913 (2001).

[21] Candes, E.J. and Recht, B. Exact Matrix Completion via Convex Optimization. arXiv:0805.4471 [cs.IT].

[22] Cai, J-F., Candes, E.J. and Shen, Z. A Singular Value Thresholding Algorithm for Matrix Completion. arXiv:0810.3286 [math.OC].

[23] Candes, E.J. and Wakin, M.B. An introduction to compressive sampling. IEEE Sig. Proc. Mag. **25**, 21–30 (2008).

[24] Kosut, R.L. Quantum Process Tomography via l_1-norm Minimization. arXiv:0812.4323 [quant-ph].

[25] Gross, D., Liu, Y-K., Flammia, S.T., Becker, S. and Eisert, J. Quantum state tomography via compressed sensing. arXiv:0909.3304 [quant-ph]

[26] Gross, D. Recovering low-rank matrices from few coefficients in any basis. arXiv:0910.1879 [cs.IT].

[27] Shabani, A., Kosut, R.L. and Rabitz, H. Compressed Quantum Process Tomography. arXiv:0910.5498 [quant-ph].

[28] Shabani, A., Mohseni, M., Lloyd, S., Kosut, R.L. and Rabitz, H. Efficient estimation of nearly sparse many-body quantum Hamiltonians. arXiv:1002.1330 [quant-ph].

[29] Schollwöck, U. The density-matrix renormalization group. Rev. Mod. Phys. **77**, 259–315 (2005).

[30] Schuch, N. and Cirac, J.I. Matrix Product State and mean field solutions for one-dimensional systems can be found efficiently. Phys. Rev. A **82**, 012314 (2010).

[31] Aharonov, D., Arad, I. and Irani, S. An Efficient Algorithm for approximating 1D Ground States. Phys. Rev. A **82**, 012315 (2010).

[32] Eisert, J., Cramer, M., and Plenio, M.B. Area laws for the entanglement entropy - a review, Rev. Mod. Phys. **82**, 277 (2010).

[33] Bratteli, O. and Robinson, D.W. *Operator Algebras and Quantum Statistical Mechanics*, Texts and Monographs in Physics Vol. 2 (Springer, New York) 1979.

[34] Verstraete, F. and Cirac, J.I. Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions. arXiv:cond-mat/0407066.

[35] de Beaudrap, N., Ohliger, M., Osborne, T.J. and Eisert, J. Solving frustration-free spin systems. arXiv:1005.3781 (2010).

[36] Tagliacozzo, L., Evenbly, G. and Vidal, G. Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law. Phys. Rev. B 80, 235127 (2009).

[37] Evenbly, G. and Vidal, G. Algorithms for entanglement renormalization. Phys. Rev. B 79, 144108 (2009).

[38] Walgate, J., Short, A., Hardy, L. and Vedral, V. Local Distinguishability of Multipartite Orthogonal Quantum States. Phys. Rev. Lett **85**, 4972–4975 (2000).

4.4 Discussion

Dans cette discussion, nous nous concentrerons sur le protocole avec contrôle unitaire puisqu'il s'agit de celui qui découle directement de notre travail (cf. 4.3).

4.4.1 Adaptation du protocole pour les MPS-PBC

Le protocole présenté dans l'article fonctionne bien sur des MPS-OBC. En effet, dans ce cas, il suffit de démarrer la procédure à une extrémité de la chaîne et de progressivement désintriquer une particule afin d'appliquer la même méthode sur une chaîne de $n-1$ particules.

Dans le cas de particules disposées sur un cercle, *i.e.*, pour un MPS avec des conditions aux limites périodiques (MPS-PBC), il n'est pas clair comment procéder. Une option est de transformer le MPS-PBC en MPS-OBC en écrasant le cercle en une chaîne double, mais cette approche demande de prendre le carré de la dimension de lien $D \mapsto D^2$. Une meilleure alternative serait d'ouvrir la chaîne, mais se pose alors le problème d'identifier si l'intrication des particules considérées est liée aux particules « à gauche » ou « à droite ».

4.4.2 Application de l'apprentissage MPS à la discrimination d'états

Dans cette section, nous montrons que le protocole de discrimination d'état proposé dans [74] peut être reformulé comme une application de notre protocole d'apprentissage MPS.

Considérons le problème suivant, celui de la discrimination d'état : un oracle choisit $k \in [\![1; K]\!]$ et prépare $|\psi_k\rangle \in \mathcal{H}_d^{\otimes N}$ qui est un MPS de dimension de lien D sur N qudits. Autrement dit, on commence avec un mélange statistique

$$\rho = \frac{1}{K} \sum_{k=1}^{K} |\psi_k\rangle\langle\psi_k| \tag{4.20}$$

L'objectif est de déterminer k avec grande probabilité. La solution optimale afin de distinguer des états est d'utiliser la mesure de Helstrom [75]. Or, il s'agit d'une mesure collective sur N qudits qu'il est difficile d'effectuer en pratique. Est-il possible de réduire la complexité de la mesure en

manipulant l'état sur un ordinateur quantique et en tirant avantage de la structure des MPS ? Il s'agit de la question centrale de [74].

L'état 4.20 reçu de l'oracle présente la propriété cruciale que chacune de ses matrices densité réduites a un rang d'au plus DK. Appliquons maintenant la méthode d'apprentissage des MPS sur l'état

$$|\Psi\rangle = |0\rangle_{KD} \otimes \rho \tag{4.21}$$

i.e., l'état obtenu en ajoutant un système ancillaire $\mathcal{A}$ de dimension KD préparé dans l'état $|0\rangle$ à l'état reçu de l'oracle. Ce système ancillaire correspond au petit ordinateur quantique de [74].

Considérons la matrice densité réduite $\sigma^{[1]}$ on $\mathcal{A}\mathcal{S}_1$. Son rang est au plus KD et puisque nous avons la description MPS de chaque $|\psi_k\rangle$, nous pouvons calculer cette matrice densité réduite et trouver une base de son support[7], par exemple les vecteurs propres de $\sigma^{[1]}$, notés $\{|\phi_j^{[1]}\rangle_{\mathcal{A}\mathcal{S}_1} = |0\rangle_{\mathcal{A}}|\nu_j^{[1]}\rangle_{\mathcal{S}_1}\}$. Dans la méthode d'apprentissage, cette matrice densité réduite est estimée par tomographie, mais elle peut être calculée numériquement dans le scénario de discrimination. Une fois cette base calculée, on peut désintriquer une particule grâce à la transformation unitaire

$$\hat{U}_1 = \sum_{j=1}^{KD} |j\rangle_{\mathcal{A}}|0\rangle_{\mathcal{S}_1}\langle\phi_j^{[1]}|_{\mathcal{A}\mathcal{S}_1} \tag{4.22}$$

et les fils correspondant au sous-systèmes $\mathcal{A}$ et $\mathcal{S}_1$ sont échangés dans le circuit (cf. Fig. 4.7) . Le sous-système $\mathcal{S}_2$ est ensuite ajouté au sous-système ancillaire et on calcule de nouveau la matrice densité réduite $\sigma^{[2]}$ sur $\mathcal{A}\mathcal{S}_2$ dont le rang est au plus KD et on détermine ses vecteurs propres $\{|\phi_j^{[2]}\rangle_{\mathcal{A}\mathcal{S}_2}\}$. Tout ceci peut être fait efficacement à partir de la description MPS de $\{|\psi_k\rangle\}$. Il suffit alors de tourner le support de $\sigma^{[2]}$ vers le sous-espace $\mathbb{C} \otimes \mathcal{H}_{KD}$ en appliquant

$$\hat{U}_2 = \sum_{j=1}^{KD} |j\rangle_{\mathcal{A}}|0\rangle_{\mathcal{S}_2}\langle\phi_j^{[2]}|_{\mathcal{A}\mathcal{S}_2} \tag{4.23}$$

puis à échanger les fils correspondant aux sous-systèmes $\mathcal{A}$ et $\mathcal{S}_2$.

En appliquant cette procédure récursivement, tous les sous-systèmes $\mathcal{S}_1 \ldots \mathcal{S}_N$ sont progressivement désintriqués, *i.e.*, sont envoyés dans l'état $|0\rangle$ et le système ancilllaire $\mathcal{A}$ contient toute

7. Le support d'une matrice densité $\rho \in \mathcal{B}(\mathcal{H})$ est le sous-espace vectoriel des états qui ne sont pas annihilés. Formellement, il s'agit de $\mathcal{H}\backslash\text{Ker}\rho$ qui est isomorphe à $\text{Im}\rho$.

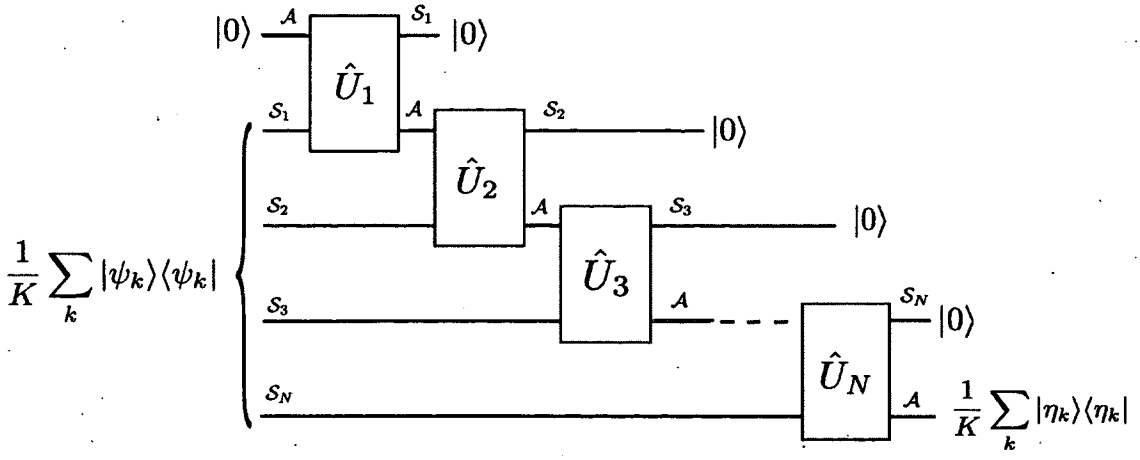


FIGURE 4.7 Circuit de tomographie pour la discrimination d'état. Toute l'information sur les systèmes $\mathcal{S}_1 \ldots \mathcal{S}_N$ est comprimée dans le système ancillaire $\mathcal{A}$ qui correspond aux fils rouges.

l'information. Plus précisément, $\mathcal{A}$ est maintenant dans l'état

$$\frac{1}{K}\sum_{k=1}^{K} |\eta_k\rangle\langle\eta_k| \tag{4.24}$$

où $|0\rangle^{\otimes N}|\eta_k\rangle = V|0\rangle_{KD}|\psi_k\rangle$ et $V = \hat{U}_N \ldots \hat{U}_1$ est le circuit d'apprentissage. Ainsi, toute l'information contenue initialement dans ρ a été comprimée dans le système ancillaire $\mathcal{A}$ et la mesure de Helstrom peut être effectuée sur ce système de petite taille.

Le raisonnement présenté jusqu'ici montre que le protocole de discrimination d'état de [74] peut être reformulé comme un cas particulier de l'apprentissage MPS où les matrices densité réduites ne sont pas estimées expérimentalement, mais plutôt calculées analytiquement grâce à la description MPS des $\{|\psi_k\rangle\}$ qui est connue a priori.

4.5 Ansatz pour intrication multi-échelle (MERA)

Les MPS sont une classe variationnelle très utile pour les systèmes 1D non-critique. En particulier, par construction, l'entropie d'un bloc de L particules sature vers une constante indépendante de L, une propriété attendue pour ces systèmes. Par contre, ils ne sont donc pas adaptés aux états pour lesquels l'entropie diverge logarithmiquement, *i.e.* comme $\log L$ avec la taille du bloc, une propriété qui se manifeste dans les états fondamentaux de systèmes dits « critiques ». La classe MERA (multiscale entanglement renormalization ansatz), proposée par Guifré Vidal en 2008 [76], est adaptée à l'étude de tels systèmes. Par abus de langage, nous appellerons un MERA un état appartenant à la classe MERA.

Le MERA est l'ansatz correspondant à une procédure de renormalisation en espace réel. Cette notion n'est pas strictement nécessaire à la compréhension initiale de notre article, même si elle permet une compréhension plus profonde. Nous avons donc fait le choix de présenter cette procédure de renormalisation dans l'annexe B.

Nous nous contenterons de voir les états MERA comme étant les états obtenus en sortie de circuits quantiques à la structure particulière (p.ex. celui de la Fig. 4.8) et de dire en deux mots en quoi ils correspondent à une procédure de renormalisation en renvoyant le lecteur à l'annexe B pour les détails. Il s'agit donc d'une classe variationnelle d'états obtenus en faisant varier les portes quantiques dans le circuit mais en respectant leur position dans le circuit, comme les MPS.

La structure particulière du circuit nous permettra d'obtenir un protocole d'apprentissage similaire à celui des MPS, décrit dans notre article reproduit en 4.6.

4.5.1 Représentation en circuit

La définition d'un MERA en terme de renormalisation explique comment transformer un état sur n qudits en un état sur un espace de Hilbert plus petit. Ainsi, on passe d'un état intriqué $|\psi\rangle$ à un état très simple à l'aide d'un circuit quantique, représenté sur la Fig. 4.5.1, qui se composent de trois types de transformations unitaires, les désintricateurs, les isométries et le tenseur sommital.

Nous allons décrire la structure de l'état MERA représenté sur la figure 4.5.1 en considérant que chaque fil correspondant à un qubit et que toutes les portes quantiques agissent sur deux qubits (MERA binaire). Plus généralement, les fils quantiques pourraient représenter des qudits où d peut

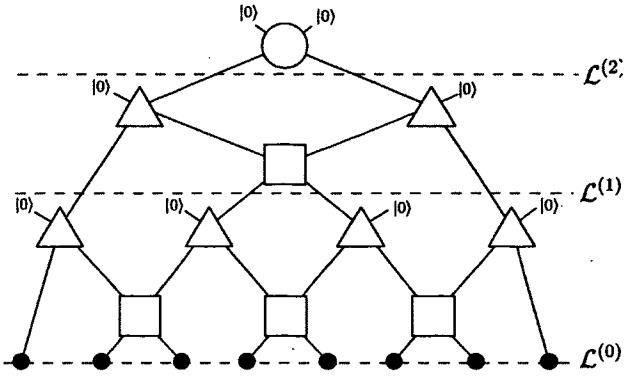


FIGURE 4.8 Circuit quantique correspondant à un MERA binaire.

varier d'un fil à l'autre et les portes pourraient agir sur un plus grand nombre de particules. Nos résultats s'étendent directement à ces cas plus généraux. Par ailleurs, nous nous concentrerons sur les MERA 1D qui agissent sur une chaîne de particules. La notion de MERA peut être étendu en dimension supérieure, mais notre travail se limite au cas 1D.

- Les désintricateurs (représentés sur la Fig. 4.5.1 par des carrés □) sont des transformations unitaires quelconques sur deux qubits.
- Les isométries (représentés sur la Fig. 4.5.1 par des triangles △) sont des transformation unitaires sur deux qubits dont l'un des qubits d'entrée est dans l'état $|0\rangle$. Ainsi, ce sont des portes qui envoie un sous-espace de dimension 2 vers un espace de Hilbert de dimension 4.
- Le tenseur sommital (représenté sur la Fig. 4.5.1 par un cercle ○) est un type particulier d'isométries dont les deux qubits d'entrée sont dans l'état $|00\rangle$.

Ainsi, un MERA peut être vu comme un circuit quantique qui prend en entrée un état $|0\rangle^{\otimes n}$ et applique des portes quantiques, arrangées en plusieurs couches d'isométries et de désintricateurs, afin de préparer l'état $|\psi\rangle$. En faisant varier les portes quantiques, on obtient une classe variationnelle d'état, le MERA.

4.5.2 Renormalisation

Afin de comprendre comment un MERA correspond à une procédure de renormalisation, il faut échanger l'entrée et la sortie du circuit. Le circuit transforme alors un état intriqué $|\psi\rangle$ vers un état produit $|0\rangle^{\otimes n}$. Chaque couche d'isométries et de désintricateurs correspond à une étape de renormalisation. Par exemple, sur la figure 4.5.1, l'état $|\psi\rangle$ sur 8 qubits au niveau physique $\mathcal{L}^{(0)}$ est renormalisé vers un état à 4 qubits au niveau $\mathcal{L}^{(1)}$. Les isométries ont pour rôle de transformer deux

qubits vers un qubit effectif. Or, si $|\psi\rangle$ est l'état fondamental d'un système critique, les isométries seules ne suffisent pas à transformer $|\psi\rangle$ en un état produit. En effet, il resterait trop d'intrication entre deux qubits renormalisés et la procédure de renormalisation échouerait à une étape ultérieure, voir B.2 pour plus de détails.

Afin de gérer cette accumulation d'intrication, le MERA propose d'introduire des transformations unitaires qui minimisent courte portée entre deux blocs adjacents de deux qubits afin que chaque bloc ne soit renormalisé vers un qubit effectif, cf. section B.3. La figure 4.9 donne une intuition schématique du rôle du désintricateur.

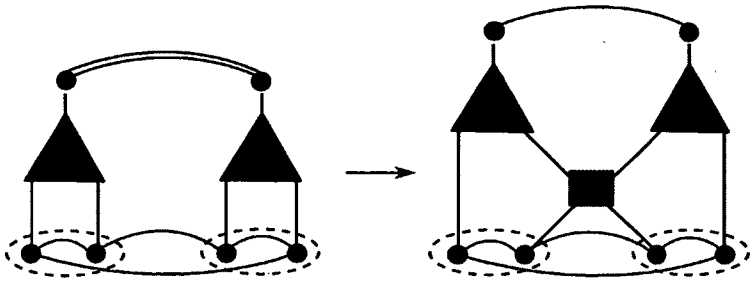


Figure 4.9 La procédure de renormalisation représentée à gauche prend deux blocs de 2 qubits et les transforme vers deux qubits effectifs. Dans ce cas, l'intrication courte-portée entre les deux blocs de particules, symbolisée par le lien vert, s'ajoute à l'intrication longue portée, symbolisée en rouge. Les deux qubits renormalisés sont alors très intriqués. L'ajout d'un désintricateur (□) avant les isométries (△) (figure de droite) permet d'éliminer l'intrication courte-portée inter-bloc. L'intrication longue portée, en rouge, sera éliminée lors de la prochaine étape de renormalisation.

4.5.3 Efficacité de la description MERA

Les états MERA possèdent donc une représentation concise qui n'est autre que la donnée des $\mathcal{O}(n)$ portes quantiques qui le constitue. De plus, les états MERA ont une description efficace : elle permet le calcul de valeurs moyennes de produit tensoriel d'observables locales. Pour comprendre cette propriété, il est utile de voir les MERA comme un cas particulier d'état à réseau de tenseurs. En effet, les MERA, comme les MPS, peuvent être vus soit comme un circuit quantique soit comme un réseau de tenseurs. Il suffit de considérer les portes unitaires du circuit comme des tenseurs : chaque fil quantique représente la contraction de tenseurs. La condition d'unitarité prend alors une représentation graphique simple : la contraction d'une isométrie ou d'un désintricateur v avec son transconjugué $v^\dagger$ donne le tenseur trivial (cf. Fig 4.10).

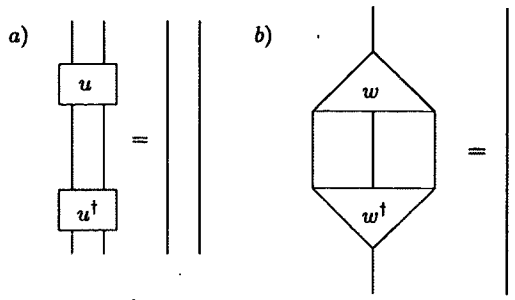


FIGURE 4.10 Simplification des tenseurs conjugués en raison de la relation d'isométrie dans un MERA. a) désintricateur b) Isométrie

Le calcul de la valeur moyenne d'une observable locale est donc une contraction de tenseurs. Or, une grande partie du réseau à contracter se simplifie grâce à la condition d'unitarité. En fait, seuls les tenseurs situés dans le cône causal des particules sur lesquelles agit l'observable ne se simplifie pas. Le cône causal d'une particule est l'ensemble des tenseurs dont la modification affecte l'état de la particule. Pour le MERA, pour toute particule au physique $\mathcal{L}^{(0)}$, le cône causal ne contient qu'un nombre constant de tenseurs, indépendant du nombre n de particules. On dit alors que la largeur du cône est bornée (par une constante). Cette propriété permet de montrer que le calcul de la valeur moyenne d'une observable peut se faire en temps $\mathcal{O}(\log n)$. Pour expliquer ce point, le plus simple est de le représenter comme la contraction d'un réseau de tenseurs, cf. Fig. 4.11. Or, dans cette contraction, la vaste majorité des tenseurs se simplifient en raison de la propriété d'isométrie. Le calcul peut donc être fait de façon efficace.

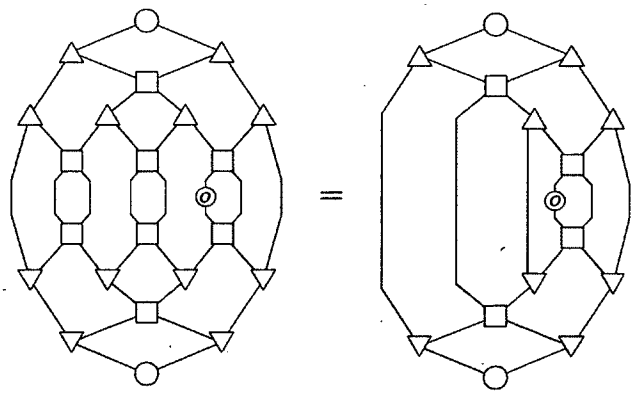


FIGURE 4.11 Calcul de la valeur moyenne d'une observable sur un qudit physique. L'observable O est représentée par un cercle rouge. La contraction de tenseur est simplifiée par les relations d'isométrie illustrées sur la Fig. 4.10.

4.6 Article : « Practical learning method for multi-scale entangled states »

Dans cette section, nous reproduisons l'article

Practical learning method for multi-scale entangled states
Olivier Landon-Cardinal et David Poulin.
New Journal of Physics, **14**, 085004 (2012)

4.6.1 Genèse et contribution

L'idée essentielle de la procédure de tomographie pour les MERA est très similaire à celle des MPS : il s'agit de progressivement reconstruire le circuit préparateur de l'état expérimental à l'aide d'estimations tomographiques de matrice densité sur un nombre constant de particules. Dans sa version naïve, le protocole de tomographie procède étage par étage et demande donc d'appliquer physiquement les portes quantiques avant d'utiliser le protocole sur l'étage suivant. Or, ce contrôle unitaire n'est pas nécessaire : il peut être éliminé au prix d'un effort numérique plus important en considérant qu'une observable au niveau physique $\mathcal{L}^{(0)}$ correspond à une observable sur les niveaux subséquents $\mathcal{L}^{(\tau)}$. Ainsi, il est possible d'effectuer de la tomographie sur les niveaux $\mathcal{L}^{(\tau)}$ en n'utilisant que des mesures au niveau $\mathcal{L}^{(0)}$.

L'article sur les MERA est le reflet d'un travail que j'ai mené seul, suivant les conseils de mon directeur, David. Les idées sont le fruit de l'interaction entre David et moi. J'ai rédigé l'article, écrit le code Matlab pour simuler le protocole et réalisé toutes les simulations numériques.

4.6.2 Article

Practical learning method for multi-scale entangled states

Olivier Landon-Cardinal* and David Poulin†
Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada
(Dated: 1 Agugust 2012)

We describe two related methods for reconstructing multi-scale entangled states from a small number of efficiently-implementable measurements and fast post-processing. Both methods only require single-particle measurements and the total number of measurements is polynomial in the number of particles. Data post-processing for state reconstruction uses standard tools, namely matrix diagonalization and conjugate gradient method, and scales polynomially with the number of particles. Both methods prevent the build-up of errors from both numerical and experimental imperfections. The first method is conceptually simpler but requires unitary control. The second method circumvents the need for unitary control but requires more measurements and produces an estimated state of lower fidelity.

I. INTRODUCTION

Quantum state tomography [1] is a method to learn a quantum state from measurements performed on many identically prepared systems. This task is crucial not only to assess the degree of control exhibited during the preparation and transformation of quantum states, but also in comparing theoretical predictions to experimental realizations. For instance, numerical methods are used to compute the ground states or thermal states of model quantum systems. Quantum state tomography could be used to check that the experimental state corresponds to the predicted one, thus providing an essential link between theory and experiments. For example, one could in principle use tomography to settle the question [2] of which states correctly describe the quantum Hall fluid at various filling parameters.

In practice however, the state of n particles is described by a number of parameters that scales exponentially with n. Therefore, tomography requires an exponential number of identically prepared systems on which to perform exponentially many measurements needed to span a basis of observables that completely characterizes the state. Furthermore, solving the inference problem to determine the quantum state that is compatible with all these measurement outcomes requires an exponential amount of classical post-processing. These factors limit tomography to at most a few tens of particles. Thus, the ability to efficiently check the state of experimental systems is a current roadblock to demonstrate quantum control over increasingly large quantum systems.

While this exponential blowup in resources is unavoidable for a generic state due to the exponentially large dimension of Hilbert space, many states encountered in nature have special properties that could be exploited to simplify the task of tomography. Indeed, many states of interest can be described by only a polynomial number of parameters. In fact, the overwhelming majority of tomographic experiments performed to date [3–9] were used to learn states described with only a few parameters. Such variational states—belonging to a family of states specified with only a few parameters—are omnipresent in many-body physics because they are tailored for numerical calculations and can predict many phenomena observed in nature (Kondo effect, superconductivity, fractional statistics, etc). One example, familiar to both the quantum information and computational many-body communities, is matrix product states (MPS) [10–13] that are at the heart of the density matrix renormalization group (DMRG) numerical method [14, 15], suitable for the description of one-dimensional quantum systems with finite correlation length [16].

Variational tomography consists in identifying a state within a variational family of states. Since those states are described by a few parameters, variational tomography amounts to learning those parameters and might require fewer resources than required by general tomography.

Recently, we and others have demonstrated [17] that tomography can be performed efficiently on MPS, *i.e.*, such states can be learned from a small number of simple measurements and efficient classical post-processing. Here, we take this result one step further and demonstrate that it is possible to efficiently learn the states associated to the multi-scale entanglement renormalization ansatz (MERA) introduced by Vidal [18], for which efficient numerical algorithms to minimize the energy of local Hamiltonians exist [19]. As opposed to MPS, these MERA states are not restricted to one dimension and can describe systems with algebraic decay of correlations. This last distinction is important because one of the most interesting phenomena in physics, quantum phase transitions, leads to a diverging correlation length and are therefore not suitably described by MPS. In contrast, MERA have been successfully used to study numerous many-body models, such as the critical Ising model in 1D [19–22] and 2D [23], and can also accurately describe systems with topological order [24–26].

In this article, we present two related methods to learn the one-dimensional MERA description of a state using tomographic data obtained from local measurements

*Electronic address: olivier.landon-cardinal@usherbrooke.ca
†Electronic address: david.poulin@usherbrooke.ca

performed on several copies of the states. Our learning methods for MERA are based on the identification of the unitary gates in the quantum circuit that outputs the MERA state. In that regard, this article is a continuation of our work on MPS and is reminiscent of early methods to numerically optimize MERA tensors [27]. However, going from MPS to MERA is non-trivial because MERA exhibits a spatial arrangement of gates that is more elaborate. Since MERA is a powerful numerical tool, our learning method bridges the gap between numerical simulations and experiments by allowing the direct comparison of numerical predictions to experimental states.

The first method we present requires unitary control of the system and the ability to perform tomography on blocks of a few particles, which can be realized using the correlations between single-particle measurements. Crucially, the size of those blocks does not depend on the total size of the system, making it a *scalable* method. In an experiment, one cannot know beforehand if the state in the lab is a MERA. However, our method contains a built-in certification procedure from which one can assess the proper functioning of the method as the experiments are performed and conclusively determine if the state is well described by the MERA. This method is experimentally challenging since it requires the ability to apply generic quantum gates on a few particles. We therefore present a second method that builds on the first one, but completely circumvents the need for unitary control. Thus, this second MERA learning method can be implemented with existing technologies. While we consider the second method to be the most interesting from a practical point of view, we chose to present the first approach for pedagogical reasons as it is conceptually simpler.

The rest of the paper is organized as follows. We begin with a more comprehensive exposition of variational tomography in the next section. For pedagogical reasons, we then present the method for MERA learning which uses unitary control in Sec. III. Section III A explains how to identify the disentanglers. We start by deriving a necessary condition for the existence of suitable disentangler and then turn this criterion into a heuristic objective function that we minimize numerically. In Sec. III B, we carefully analyze the buildup of errors in our procedure and show that errors only accumulate linearly with the size of the system. In Sec. III C, we present numerical benchmarks of our tomography method. In Sec. IV, we present the simplified method that does not require unitary control. We demonstrate in Sec. IV A that it is not necessary to apply the disentanglers to the experimental state since we can simulate the effect of those disentanglers numerically, albeit at the cost of more repeated measurements, as analyzed in Sec. IV B. We emphasize the practical aspect of the method with an example in Sec. IV C. Finally, we discuss the error scaling and certification in Sec. IV D. In Sec. V, we discuss the relationship between the numerical tractability of a variational family of states and the ability to learn efficiently the variational parameters. Finally, we present in appendix B a tool to contract two different MERA states, which allows for the efficient comparison of a MERA whose parameters have been identified experimentally using our method to a predicted theoretical MERA state.

II. VARIATIONAL TOMOGRAPHY

Consider a variational family of states $\mathcal{C} \equiv \{|\phi(\alpha)\rangle\}_\alpha$, *i.e.*, a family of states described by the possible values of a collection of variational parameters collectively denoted α. To be useful and interesting, this family of states has to be numerically tractable and physically faithful. Numerical tractability requires that states are specified by a few (at most polynomial) parameters, but also that this efficient representation allow for the efficient numerical computation of quantities of interest, such as the energy of the system, correlation functions, or more generally expectation values of local observables. Physical faithfulness means that states of the variational family exhibit the properties of the system of interest.

Given access to multiple copies of the system in a state σ, one is interested in i) verifying that the experimental state σ belongs to the variational family, *i.e.*, check that σ is close to a state $|\phi(\alpha_0)\rangle \in \mathcal{C}$; and ii) learn those variational parameters α_0 directly from experimental measurements. In our methods, those two questions will be answered simultaneously. One will perform *variational tomography* to extract the values α_0 and use the same experimental data to *certify* that the experimental state is indeed close to the state $|\phi(\alpha_0)\rangle$.

The main motivation for variational tomography is that, from a logical standpoint, it might be possible to entirely learn the state from a small number of measurements since it is described by a small number of parameters. This is to be contrasted with quantum state tomography whose goal is to learn a generic state in the full Hilbert space, and thus has to identify an exponential number of parameters. However, in variational tomography, it is not necessarily the case that the variational parameters can be accessed directly by measurements; one could imagine that the only way to identify a quantum state inside a variational family would be to perform quantum state tomography in the full Hilbert space and then compute the variational description of the state. The relevant questions thus become i) for which variational families is variational tomography efficient? and ii) for such classes, what experimental toolbox is required?

A natural variational family of states are those produced by a polynomial-size quantum circuit starting from, *e.g.*, the all $|0\rangle$ state. In that case, the variational parameters α correspond to the description of the circuit namely a description of its gates and their locations. More refined variational families are obtained by further restricting these quantum circuits, *e.g.*, by fixing the layout of gates but leaving the value of each gate arbitrary. In that case, the parameter space of α scales precisely

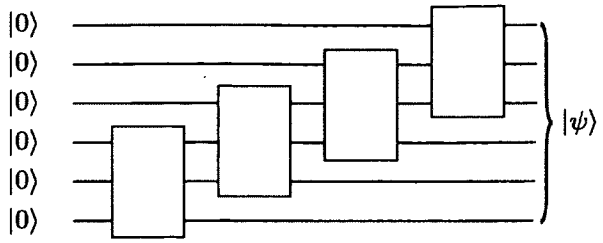


Figure 1: Matrix product state generation from a staircase circuit of local unitary gates.

with the number of gates.

The prime example is matrix product states (MPS). These states are the output of a quantum circuit whose gate layout is a staircase pattern shown in Fig. 1. Furthermore, MPS allow for efficient computation of expectation values of local observables and can reproduce physics of gapped 1D system. In [17], we and others have shown how to identify those gates efficiently. In this work, we turn our attention to MERA states, that are also described by quantum circuits but with a more elaborate gate layout.

III. MERA LEARNING WITH UNITARY CONTROL

A. Identifying the disentanglers

MERA states can be described as the output of a quantum circuit [18] whose structure is represented on Fig. 2 (as seen with inputs on the top and output at the bottom). For simplicity, we will focus in Sec. III on a one dimensional binary MERA circuit for qubits, but our method generalizes to all 1D MERA states, *i.e.*, particles could have more internal states—thus accounting for a larger MERA refinement parameter χ— and isometries could renormalize several particles to one effective particle. The circuit contains three classes of unitaries. Disentanglers (represented as $\square$) are two-qubit unitary gates; isometries (represented as $\triangle$) are also two-qubit gates but with one input qubit always in the $|0\rangle$ state; the top tensor (represented as $\bigcirc$) is a special case of isometry that takes as input two qubits in the $|00\rangle$ state. Each renormalization layer is made of a row of disentanglers and a row of isometries. Disentanglers remove the short-scale entanglement between adjacent blocks of two qubits while isometries renormalize each pair of qubits to a single qubit. Each renormalization layer performs theses operations on a different length scale. The quantum circuit thus mirrors the renormalization procedure that underlies the MERA.

Learning a MERA state amounts to identifying the various gates in that circuit. By inverting the time direction, we can think of this circuit as transforming a MERA state into the all $|0\rangle$ state, by sequentially disentangling

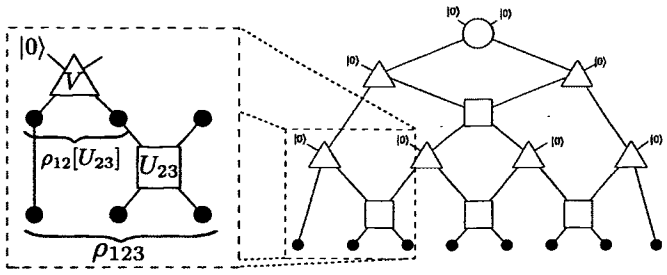


Figure 2: The optimal disentangler $\tilde{U}$ can be computed from the tomographic estimation of the density matrix ρ_{123} on the first three qubits. Once applied, the resulting state $\tilde{\rho}_{12}[\tilde{U}]$ is very close to a rank 2 matrix. Thus, there exist a unitary V that transform $\tilde{\rho}_{12}[\tilde{U}]$ into a state with the first qubit in the state $|0\rangle$.

"ancillary" qubits from the system. The intuitive idea behind our scheme is to proceed by varying the isometries and disentanglers until these ancillary qubits reach the state $|0\rangle$ for each row of isometries. We will exploit this feature to numerically determine each disentangler.

1. *Necessary condition for disentangler*

Consider the n qubits at the lowest layer of the MERA. Let ρ_{123} be the reduced density matrix on the first three qubits (see Fig. 2). If the state is exactly a MERA, there exists a unitary U_{23} acting on qubits 2 and 3 (see left of Fig. 2) for which

$$\rho_{12}\left[U_{23}\right] = \mathrm{Tr}_3\left[\left(\mathbb{I}_1 \otimes U_{23}\right)\rho_{123}\left(\mathbb{I}_1 \otimes U_{23}^{\dagger}\right)\right] \qquad (1)$$

has rank at most 2. Indeed, if the rank of $\rho_{12}[U_{23}]$ was strictly greater than 2, it would be impossible for the isometry V (see left of Fig. 2) to map the density matrix $\rho_{12}\left[U_{23}\right]$ to a state with one of the qubit in the state $|0\rangle$ because the dimension of the space $|0\rangle \otimes \mathbb{C}^2$ would be strictly smaller than the dimension of the support of the density matrix. Hence, we have the necessary criterion

$$\exists \tilde{U}_{23} \quad \rho_{12}\left[\tilde{U}_{23}\right] \text{ has rank less or equal than 2.} \qquad (2)$$

To find a unitary that fulfills this criterion, it is necessary to know the state ρ_{123}, and this can be achieved by brute-force tomography on these three qubits. Once the original state on the three qubits is known, one has to perform a search over the space of unitaries to find a suitable disentangler. To do this, we will define in Sec. III A 2 an objective function to minimize numerically.

Once this optimal unitary operator $\tilde{U}$ has been found numerically, it is necessary to consider how it modifies the quantum state before learning the other elements of the circuit. One obvious way to do so is to apply the unitary transformation to the experimental state and continue the procedure on the transformed state. This amounts to executing the circuit, and should in the end map the

experimental state to the all $|0\rangle$ state. For pedagogical reasons, we will first present our scheme assuming that the state is transformed at every step this way. Of course, such unitary control increases the complexity of the scheme and could be out of the reach of current technologies. However, in Sec. IV, we will explain how this unitary transformation can be circumvented at the cost of increasing the number of measurements.

After the optimal disentangler $\tilde{U}$ has been applied to the state, we need to identify the unitary V that rotates the density matrix on the first two qubits such that the first qubit is brought to the $|0\rangle$ state, c.f. Fig 2 left. This does not require any additional tomographic estimate since we already know the descriptions of the state on the first three qubits and the disentangler. We can thus compute the state on the first two qubits $\rho_{12}[\tilde{U}]$ and diagonalize it to obtain the eigenvectors corresponding to its two non-zero eigenvalues. The unitary V is chosen to map those two eigenvectors to the space $|0\rangle \otimes \mathbb{C}^2$, i.e., V rotates the qubits such that the support of the density matrix is mapped to a space where the first qubit is in the $|0\rangle$ state.

All other disentanglers of this layer can be found by recursing the above procedure. Once a disentangler has been identified, it is experimentally applied to the system and brute-force tomography is performed on the next block of three qubits.

Notice that for the last block of a layer of an open boundary MERA, a single unitary is responsible for minimizing the rank of two density matrices. One possible way to handle this is to get a tomographic estimate of the state on the last four qubits and to try to minimize the rank of both reduced matrices. Another way, for which we have opted in our numerical simulations, is to perform multiple sweeps over the layer. For instance, the disentanglers will first be identified from left to right and then the next sweep will be performed from right to left, using the disentanglers found in the first sweep as initial guesses in the space of unitaries (see Fig. 3). The number of sweeps can be increased for better accuracy. Additional sweeps requires either to extract the tomographic estimates after each new sweep or to perform tomography on blocks of slightly larger size. Multiple sweeps would also allow to apply our method to MERA states with periodic boundary conditions in 1D and could be useful for 2D-MERA states. While this would be an interesting continuation of our work, we focus on 1D-MERA for the rest of the article.

2. *Heuristic objective function*

One of the steps in our protocol consists in identifying the unitary $\tilde{U}$ that minimizes the rank of $\rho_{12}[U]$, c.f. Eq. (1). There are many distinct ways this can be done and in this section, we present a practical heuristic to accomplish this task. Minimizing the rank of the density matrix $\rho_{12}[U]$ is not a suitable numerical task because,

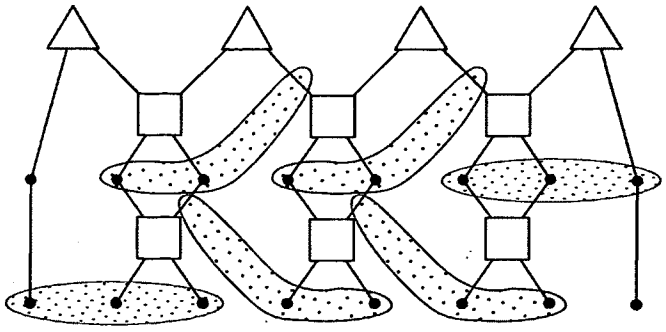


Figure 3: Identification of the disentanglers using two successive sweeps of the chain. Dotted regions cover particles on which brute-force tomography is performed. The first sweep (red dotted regions) finds unitaries starting from the left end of the chain. Those unitaries will be used as initial guesses for the second sweep (blue striped regions) that starts from the right end of the chain.

even if the experimental state is an exact MERA, the inferred density matrix will typically have full rank due to machine precision and the imperfect tomographic estimation of ρ_{123}. Thus, we turn the problem of finding $\tilde{U}_{23}$ into an optimization problem by considering the spectral decomposition of $\rho_{12}[U] = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k|$ where the eigenvalues are sorted in decreasing order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$. If $\rho_{12}[U]$ has most of its support on a two-dimensional space, it will have two small eigenvalues that are typically non-zero due to imperfections. We thus consider the objective function

$$f(U, \rho_{123}) = \sum_{k>2} \lambda_k \tag{3}$$

and we perform a minimization over the space of unitaries to determine the optimal unitary $\tilde{U}$. This objective function has a well-defined operational meaning—it is the probability of measuring the disentangled qubit in the state $|1\rangle$ after the isometry V has been applied. We will see in Sec. III B that this property can be used to certify the distance between the experimental and the reconstructed states.

Another way to think about this objective function is to consider the characteristic polynomial $P[X]$ of $\rho_{12}[U]$ which is of the form

$$P[X] = X^4 - X^3 + aX^2 - bX + c \tag{4}$$

where the coefficients a, b and c are positive since they correspond to the sum of product of the positive eigenvalues of the density matrix. In particular, coefficient b is the sum of all products of three eigenvalues, *i.e.*, $b = \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_3\lambda_4 + \lambda_2\lambda_3\lambda_4$. In order for the rank of the density matrix to be 2, it is sufficient for all 4 products of three eigenvalues to vanish, *i.e*,

$$\rho_{12}[U] \text{ of rank less than 2} \iff b = 0. \tag{5}$$

Thus, another suitable objective function is the positive coefficient b, which is a polynomial in the entries of

$\rho_{12}[U]$. Indeed, using Bocher formula, coefficient b can be expressed as $6b = 1 - 3\mathrm{Tr}A^2 + 2\mathrm{Tr}A^3$ where $A = \rho_{12}[U]$. Thus, b is a well-behaved function with respect to the density matrix. Note also that b can be expressed without diagonalizing the density matrix $\rho_{12}[U]$. We will focus on minimizing Eq. (3) in all subsequent discussions and numerical results.

3. *Numerical minimization over unitary space*

Minimization of Eq. (3) is performed using a conjugate gradient method. We first have to account for the fact that the unitary manifold is not a vector space. To get around this problem, we go to the tangent space by writing any unitary U as the result of a Hamiltonian evolution, *i.e.*, there exists a Hermitian matrix H such that $U = e^{iH}$. It is then possible to use the standard conjugate gradient method. Let us sketch the algorithm in more details.

First, we select a unitary U_0 either at random or from an initial guess (provided for instance by a previous sweep). We will search the unitary space by generating a sequence of unitaries $\{U_k\}$. At the k^{th} step of the minimization, the algorithm is the following.

1. We center the unitary space at point U_{k-1} by defining $\rho_k = (\mathbb{I} \otimes U_{k-1})\rho_{k-1}(\mathbb{I} \otimes U_{k-1})^\dagger$.
2. We compute the gradient $G^{(k)}$ by parametrizing the Hamiltonian H on 2 qubits by its decomposition on the Pauli group $H = \sum_{\mu\nu} h_{\mu\nu}\sigma_\mu \otimes \sigma_\nu$ where $\sigma_\mu \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ is a Pauli matrix. We successively evaluate the component of the gradient $G^{(k)}$ in the direction (μ, ν) by looking at the effect of the test unitary $U_{\mu,\nu} = \mathbb{I} + i\epsilon\sigma_\mu \otimes \sigma_\nu$ on the objective function, *i.e.*, $G^{(k)}_{\mu,\nu} = \frac{f(U_{\mu,\nu}\rho_k) - f(\mathbb{I},\rho_k)}{\epsilon}$ where ϵ is a small number.
3. Instead of following the gradient, which would generally undo some of the minimization performed in the previous steps, we use a conjugate gradient method where the new direction of search $\tilde{G}^{(k)}$ is optimized by taking into account the direction used in the previous step $\tilde{G}^{(k-1)}$ through the Polak-Ribière formula. More precisely, $\tilde{G}^{(k)} = G^{(k)} + \beta\tilde{G}^{(k-1)}$ in which the real parameter β is defined as $\beta = \max\left(0, \frac{G^{(k)}\cdot(\tilde{G}^{(k-1)} - G^{(k)})}{\tilde{G}^{(k-1)}\cdot\tilde{G}^{(k-1)}}\right)$.
4. We perform a line search along the direction $\tilde{G}^{(k)}$ by considering the family of unitaries $\exp\left(-it\sum_{\mu,\nu}\tilde{G}_{\mu,\nu}\sigma_\mu \otimes \sigma_\nu\right)$ and optimizing the parameter t to find t_{opt}. We then define

$$U_k = \exp\left(-it_{opt}\sum_{\mu,\nu}\tilde{G}_{\mu,\nu}\sigma_\mu \otimes \sigma_\nu\right) \qquad (6)$$

which ends the k^{th} iteration.

We iterate until the objective function is close enough to zero or that improvement has stopped. Note K the total number of iteration steps. The disentangler returned by the algorithm is $\tilde{U} = U_K U_{K-1} \ldots U_0$.

This method is *heuristic* since the objective functions present no characteristic that would ensure the convergence of the conjugate gradient method. In particular, our search over unitary space depends on the starting point, *i.e.*, the unitary chosen in the first iteration. Indeed, some starting points will lead the heuristic to a local minima where it will get stuck. In order to avoid this phenomenon, we can repeat the overall search by picking at random (according to the unitary Haar measure) different initial points which lead to potentially different minima and keep the smallest of those minima. In any case, this is a minimization problem over a space of *constant* dimension, so the method used to solve it does not affect the scaling with the number of particles n. Ultimately, we can always use a finite mesh over the unitary space and use brute-force search. Nevertheless, we found numerically that this heuristic works well.

A more serious problem is that a choice of unitary that is optimal *locally*, in the sense that it minimizes Eq. (3), could be sub-optimal *globally* as it might lead to a state for which it is impossible to find a disentangler obeying Eq. (3) elsewhere in the circuit. This is a phenomenon that is more likely to occur when the minimum is degenerate, *i.e.*, there exists several distinct (modulo gauge) exact disentanglers for the state. However, we have performed numerical experiments on randomly generated MERA states as well as physically motivated states and found that the conjugate gradient performs well (see Sec. III C).

B. Error analysis

In practice, due to numerical and experimental imperfections, the first disentangled qubit will not be exactly in the $|0\rangle$ state, but merely close to it. This situation arises from the conjunction of three causes : *i)* the experimental state of the system is not exactly a MERA, but merely close to one, *ii)* the tomographic estimate of the density matrices on blocks of three qubits are slightly inaccurate due to noisy measurements and experimental finite precision, *iii)* the numerical minimization did not find the exact minimum.

1. *Preventing error amplification by post-selection*

Our error analysis will show that the buildup of errors is linear in the number of disentanglers of the MERA circuit, which is itself linearly proportional to the number of particles in the experimental state. Essentially, the distance between the reconstructed state and the experimental state is the sum of the error made at each elementary step when estimating a disentangler and an

isometry. Fortunately, the error made at each elementary step is not amplified by errors made at previous steps. The key to isolate each step from the others is to measure the qubit that should have been disentangled in the computational basis. With high probability the qubit will be found in the $|0\rangle$ state. While the probability of measuring the $|0\rangle$ outcome depends on previous errors, the post-selected state is now free from previous errors. The interest of this post-selection is two-fold. First, it prevents errors in previous steps to contaminate the state and amplify the error made at the current step, thus limiting the error propagation. Second, by accumulating statistics on this measurement, we can estimate the probability of outcome $|0\rangle$ and use it to bound the distance of the reconstructed state to the actual state in the lab. Therefore, our procedure comes with a *built-in certification process.* We now describe the error analysis in more details.

2. *Error at each elementary step*

Recall the notation of Fig. 2. Due to numerical and experimental imperfections, the state on qubits 1, 2 and 3 after applying the disentangler $\tilde{U}_1$ and the isometry V_1 is not exactly in the $|0\rangle \otimes \mathbb{C}^{2(n-1)}$ subspace but contains a small component orthogonal to that space. Thus, it has the form

$$V_1\tilde{U}_1|\psi\rangle = \frac{|0\rangle|\eta_1\rangle + |e_1\rangle}{\sqrt{1+\langle e_1|e_1\rangle}} \tag{7}$$

where $|\eta_1\rangle$ is the normalized pure state on qubits 2 to n if qubit 1 had been completely disentangled from the chain and $|e_1\rangle$ is some *sub-normalized* vector supported on the subspace $|1\rangle \otimes \mathbb{C}^{2(n-1)}$. The isometry V_1 is chosen to minimize the norm of $|e_1\rangle$, i.e., to minimize $\epsilon_1 \equiv \langle e_1|e_1\rangle$.

Further along the layer, the state after applying k disentanglers and k isometries will be of the form

$$V_k\tilde{U}_k \ldots V_1\tilde{U}_1|\psi\rangle = \frac{|0\rangle^{\otimes k}|\eta_k\rangle + |e_k^{ac}\rangle}{\sqrt{1+\epsilon_k^{ac}}} \tag{8}$$

where the first term $|0\rangle^{\otimes k}|\eta_k\rangle$ is the normalized state had the k qubits in position $1, 3 \ldots 2k-3$ been completely disentangled from the chain and $|e_k^{ac}\rangle$ is the accumulated error vector orthogonal to the space where those k qubits are in the $|0\rangle^{\otimes k}$ state, whose square norm is $\epsilon_k^{ac} \equiv \langle e_k^{ac}|e_k^{ac}\rangle$. In order to find the optimal disentangler and isometry, we measure the last disentangled qubit in the computational basis and post-select on the $|0\rangle$ outcome, which occurs with probability $(1+\epsilon_k^{ac})^{-1}$. We then perform brute force tomography and identify numerically the disentangler and the isometry that minimizes the norm of the error vector $|e_{k+1}\rangle$ such that

$$V_{k+1}\tilde{U}_{k+1}|\eta_k\rangle = \frac{|0\rangle|\eta_{k+1}\rangle + |e_{k+1}\rangle}{\sqrt{1+\epsilon_{k+1}}}. \tag{9}$$

Applying this disentangler and isometry to the whole state of the chain, one gets

$$V_{k+1}\tilde{U}_{k+1} \ldots V_1\tilde{U}_1|\psi\rangle = \frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |e_{k+1}^{ac}\rangle}{\sqrt{1+\epsilon_{k+1}^{ac}}} \tag{10}$$

where the accumulated error vector at step $k+1$ is

$$|e_{k+1}^{ac}\rangle = |e_{k+1}\rangle + \sqrt{1+\epsilon_{k+1}}V_{k+1}\tilde{U}_{k+1}|e_k^{ac}\rangle \tag{11}$$

and the square of its norm $\epsilon_{k+1}^{ac} \equiv \langle e_{k+1}^{ac}|e_{k+1}^{ac}\rangle$ obeys the recurrence relation

$$1+\epsilon_{k+1}^{ac} = (1+\epsilon_{k+1})(1+\epsilon_k^{ac}) \tag{12}$$

since the elementary error vector $|e_{k+1}\rangle$, for which all previous ancillary particles have been disentangled, is orthogonal to the vector $V_{k+1}\tilde{U}_{k+1}|e_k^{ac}\rangle$. Thus,

$$1+\epsilon_{k+1}^{ac} = \prod_{i=1}^{k+1}(1+\epsilon_i). \tag{13}$$

3. *Global error*

After the choice of m disentanglers and m isometries, the reconstructed state is $|\phi\rangle = V_m^\dagger\tilde{U}_m^\dagger \ldots V_1^\dagger\tilde{U}_1^\dagger|0\rangle^{\otimes m+1}|\eta_m\rangle$. Its distance to the actual experimental state $|\psi\rangle$ can be stated in terms of the (in) fidelity as

$$\begin{aligned} 1-|\langle\phi|\psi\rangle|^2 &= 1-1/(1+\epsilon_m^{ac}) \\ &= 1-1/\prod_{i=1}^{m}1+\epsilon_i. \end{aligned} \tag{14}$$

Practically, one is interested in guaranteeing that the reconstructed state is close to the experimental state, up to global error E, textiti.e., to guarantee that $1-|\langle\phi|\psi\rangle|^2 \leq E$. Suppose that all error vectors are bounded, textiti.e. that for all step i, we have $\epsilon_i \leq \varepsilon$. Inverting Eq. (14), it suffices that

$$\varepsilon \leq (1-E)^{-1/m} - 1 \simeq E/m$$

in the limit where the tolerable global error E is small. Thus, we see that *errors accumulate linearly* and that a precision inversely proportional to the number of disentanglers is sufficient to ensure a constant global error. Furthermore, statistics on the post-selection performed at each step allows to estimate each ϵ_k^{ac}— and therefore ϵ_m^{ac} through Eq. (13)—that gives direct access to the distance between the reconstructed and experimental states.

Finally, from these estimates of ϵ_i, one can identify particular steps of the procedure that have gone wrong. This information can be used to turn the scheme into an *adaptive* one. Suppose the error is particularly large for a given step. This might be due to an important amount of entanglement concentrated in one region of space, *e.g.*,

near a defect, which can be accounted for by increasing the MERA refinement parameter χ locally, textiti.e. by using disentanglers acting on a larger number of qubits. In practice, χ could be increased until the error is below some target threshold.

C. Numerical performance

1. *Benchmarking results*

We have performed numerical simulations to benchmark the performances of the conjugate gradient method in our setting. We have generated random MERA states —by picking each unitary gate in the circuit from the unitary group Haar measure—, simulated the experiment on those states, and use our algorithm to infer the initial MERA state. We did not introduce noise in measurements to simulate experimental errors since the error analysis indicates how those errors would build up.

As mentioned before, there is no guarantee that our minimization procedure converges to the true minimum, resulting in small imperfections in the reconstructed state. Figure (4, top) shows the distance between the reconstructed state and the actual state. As indicated by the dashed line, these results are in good agreement with a linear scaling of the error, where the source of errors is due to finite machine precision and approximate minimization of the objective function.

The inference algorithm's complexity is dominated by the conjugate gradient descents, and therefore scales linearly with the number of disentanglers or the number of particles in the system. Figure (4, bottom) shows the actual run time of the inference algorithm for different randomly chosen MERA states and of various sizes. Once again, we see a good agreement with a linear dependence with the system size. Systems of up to 24 qubits can easily be handled in a few minutes of computation and requires 1197 different measurement settings for each sweep of the 24 qubit system. This is to be contrasted with the 656 100 experiments needed to reconstruct the state of 8 qubits in [3] and the post-processing of the data that took approximately a week [28]. Additional sweeps improve the convergence as showed on Fig. 5.

We also tested our method on a physical model, namely the 1D Ising model with transverse field at the critical point. The results obtained where coherent with what is expected from the approximation of the true ground state with a MERA with refinement parameter $\chi = 2$.

2. *Possible improvements*

Note the presence of isolated points on the graphs of Fig. 4 that achieve a lower fidelity and required a longer processing time. These cases appear because the heuristic fails to find a global minimum. It appears that in some cases, a unitary transformation U_{23} meeting criterion Eq. (3) is not sufficient to guarantee that it will be possible to find subsequent disentanglers obeying Eq. (3). Put another way, *locally* minimizing the objective function might not lead to a *global* optimum. Indeed, consider the following example. Let $|\psi\rangle$ be a MERA state whose first qubit is disentangled from the rest of the chain, textiti.e. $|\psi\rangle = |0\rangle|\phi\rangle$. The rank of the density matrix on the first two qubits is at most 2 and that remains true after *any* unitary is applied on qubits 2 and 3. Thus,

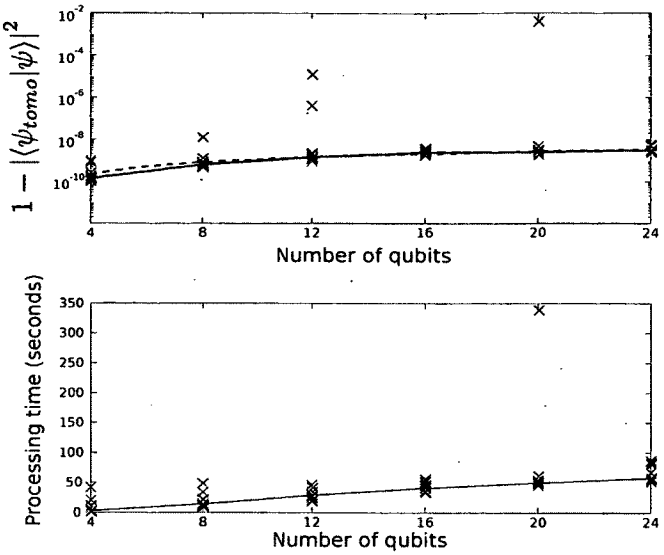


Figure 4: (top) Infidelity to the "experimental state", *i.e*, $1-|\langle\psi_{tomo}|\psi\rangle|^2$ where $|\psi\rangle$ is a random MERA on n qubits and $|\psi_{tomo}\rangle$ is the state reconstructed from the MERA tomography method using three sweeps. (bottom) Processing time (on a standard laptop) to perform MERA tomography using three sweeps. Both figures exhibit 10 runs for each number of qubits $n \in \{8, 12, 16, 20, 24\}$. In both figures, each $\times$ represents results for one random MERA. The full lines represent median for each number of qubits. The dashed line on the top figure is the linear approximation to the median. Notice that the numerical minimization can fail to converge as illustrated by the atypical data points. For instance, for one of the 20-qubit MERA, the processing time was one order of magnitude longer than the average and the infidelity six orders of magnitude larger than average.

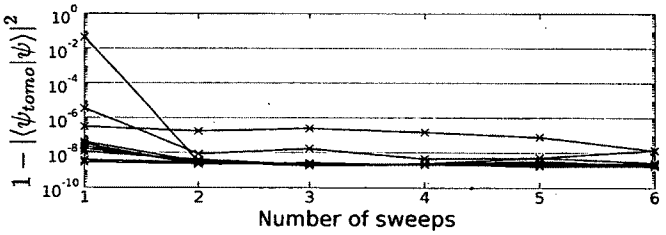


Figure 5: Infidelity to a 20 qubit state using a reconstructed method with a variable number of sweeps. Each line corresponds to a different random MERA.

any choice of disentangler minimizes the objective function Eq. (3)—we say that the minimum is degenerate. In this case, even the identity, *i.e.*, applying no disentangler at all, achieves the minimum. However, suppose the state $|\phi\rangle$ on qubits 2 to n is highly entangled and that removing part of this entanglement between qubits 2 and 3 was crucial to be able to reconstruct its MERA description. In this case, applying the identity on qubits 2 and 3, even if locally optimal, was not globally optimal. Hence, minimizing the objective function Eq. (3) seems to be necessary but not sufficient to successively identify all disentanglers.

Although our numerical simulations suggest that this situation is rather atypical and can be suppressed with additional sweeps (see Fig. 3), it is possible to overcome this problem by imposing additional constraint on the disentangler. For instance, one can demand that the second qubit be in a state as pure as possible, effectively minimizing the entanglement between the last qubit of one block and the first qubit of the next block. This corresponds to the following modified objective function

$$f\left(\tilde{\rho}_{12}\left[U\right]\right) = \sum_{k>2} \lambda_k + \epsilon\lambda_2 \tag{15}$$

i.e., we add a small perturbation that will only take action when the two smallest eigenvalues of $\tilde{\rho}_{12}[U_{23}]$ are very small and will further constrain the search. This slight modification solved the problematic situation we considered, and there exist many other heuristics to improve the method. This problem, and its heuristic solutions, are similar to those encountered when using the MERA to estimate the ground states of a Hamiltonian numerically.

IV. MERA LEARNING WITHOUT UNITARY CONTROL

For pedagogical reasons, we presented our learning method in a way that required disentanglers and isometries to be physically applied to the experimental state in order to unravel the circuit. In this section, we will show how to circumvent unitary control at the price of slightly more elaborate numerical processing and consuming more copies of the state. The main idea is to numerically simulate how measurements performed on the original, unaltered experimental system would be transformed if the unraveling circuit had been applied.

A. Simulating measurements on renormalized state

A MERA is an ansatz that corresponds to a renormalization procedure. Each renormalization step maps a state to another one on fewer particles and schematically corresponds to a layer of the MERA circuit. Applying the first layer and removing the ancillary particles that have been (approximately) disentangled maps the experimental state ρ_0 on n particles to a state ρ_1 on fewer particles (see Fig 6). Recursively, this procedure constructs a sequence of states $\{\rho_\tau\}_\tau$.

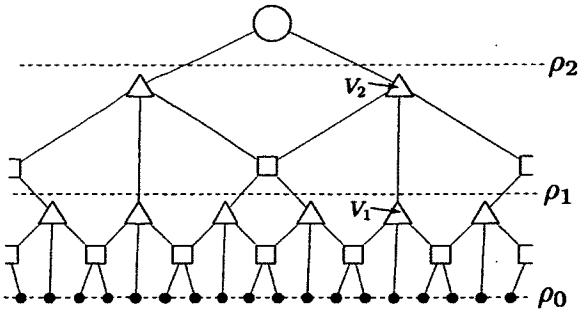


Figure 6: MERA as a renormalization procedure that creates a sequence of states $\{\rho_\tau\}_\tau$.

To get from $\rho_{\tau-1}$ to ρ_τ, one can either perform this mapping *physically* by experimentally applying the gates corresponding to the MERA layer, or one can *compute* the function mapping $\rho_{\tau-1}$ to ρ_τ from the description of the gates. As in [19], define an ascending superoperator $\mathcal{A}$ that maps an operator $O_{\tau-1}$ acting on layer $\tau-1$ to an operator O_τ acting on the next layer τ

$$O_\tau = \mathcal{A}_\tau(O_{\tau-1}) \tag{16}$$

such that

$$\mathrm{Tr}[\rho_\tau \mathcal{A}_\tau(O_{\tau-1})] = \mathrm{Tr}[\rho_{\tau-1} O_{\tau-1}]. \tag{17}$$

This recursively carries over to the experimental state ρ_0

$$\mathrm{Tr}[\rho_\tau \mathcal{A}_\tau \circ \cdots \circ \mathcal{A}_1(O_0)] = \mathrm{Tr}[\rho_0 O_0]. \tag{18}$$

Thus, in order to extract information from a density matrix ρ_τ, one can measure the expectation value of several observables O_0^i on the density matrix ρ_0. Measuring those observables will effectively amount to measuring the observables $O_\tau^i \equiv \mathcal{A}_\tau \circ \cdots \circ \mathcal{A}_1(O_0^i)$ on the density matrix ρ_τ.

The ascending superoperator can be computed from the knowledge of the disentanglers and isometries. Its exact form depends on the physical support of the observable. For instance, for ternary MERA, we can restrict to ascending superoperator that only depends on the isometries of the MERA [22] (see Fig. 7). This is a simple example where an experimental observable on one particle is mapped to observable on one renormalized particle. More generally, observables on many sites will be ascended to observables on fewer sites. Any choice of observables is valid as long as the renormalized observables $\left\{O_\tau^i\right\}_i$ span the support of the reduced density matrix ρ_τ.

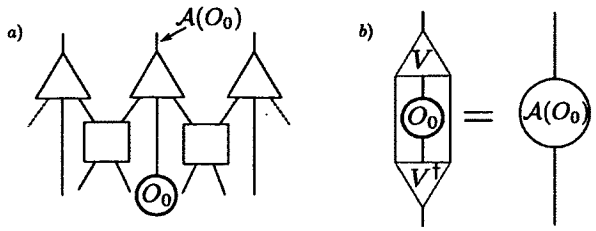


Figure 7: Ascending superoperator and renormalized observables for a ternary MERA. a) Ternary MERA with one site observable O_0 that is transformed into a renormalized observable $\mathcal{A}(O_0)$ on the renormalized state. b) Tensor contraction corresponding to the ascending superoperator $\mathcal{A}$.

B. Overhead in the number of measurements

This procedure leads to an overhead in the total number of measurements because renormalized observables are less efficient at extracting information. Suppose (for clarity) that we measure Pauli observables $\{O_0^i\}_i$ on the experimental state. These observables are orthonormal for the Hilbert-Schmidt inner product and thus maximize information extraction. However, the renormalized observables $O_1^i \equiv \mathcal{A}_1(O_0^i)$ need not be orthonormal. Consider their Gram matrix $G_{ij} = \mathrm{Tr}\left[O_1^i \left(O_1^j\right)^\dagger\right]$ which can be diagonalized by a unitary matrix Z. Its normalized eigenvectors $R_1^i = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} O_1^j$ are orthonormal observables but cannot be directly measured because they do not correspond to simple observables on the experimental state, but instead to linear combination of them. Thus, to reconstruct the density matrix $\rho_1 = \sum_i r_1^i R_1^i$, the expectation values $r_1^i = \mathrm{Tr}\rho_1 R_1^i$ have to be computed by taking a linear combination of the expectation values $o_0^j \equiv \mathrm{Tr}\rho_0 O_0^j$ on the experimental state

$$r_1^i = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} \mathrm{Tr}\rho_1 O_1^j = \frac{1}{\sqrt{\lambda_i}} \sum_j Z_{ij} o_0^j. \qquad (19)$$

Due to limited number of repeated measurements, estimation of each o_0^i will present a variance $\mathbb{V}(o_0^i)$. Suppose that measurements are repeated enough times to ensure that all variances are below a precision threshold, textiti.e., $\mathbb{V}(o_0^i) \leq \epsilon$. Since r_1^i is a linear combination of those measurements, it will have a variance $\mathbb{V}(r_1^i) = \frac{1}{\lambda_i} \sum_j |Z_{ij}|^2 \mathbb{V}(o_0^j) \leq \frac{\epsilon}{\lambda_i} \sum_j |Z_{ij}|^2$. Therefore, in order to ensure a precision ϵ on the estimate of r_1^i, this imprecision needs to be compensated by multiplying the number of repeated measurements by the *conditioning factor* $\lambda_i^{-1} \sum_j |Z_{ij}|^2$.

When scaling operators on τ layers, the conditioning factors for each layer will multiply (in the worst case) but we expect the conditioning for each layer to be a constant independent of system size. Thus, the total number of measurements will remain *polynomial* in the number of particles since there is only a *logarithmic* number of renormalization layers.

We can make this argument rigorous for critical systems that exhibit scale-invariance, precisely the physical systems for which MERA was introduced. Due to scale-invariance, the ascending operator $\mathcal{A}_\tau$ will not depend on the index of the layer and we refer to it as the scaling superoperator $\mathcal{S}$ [22]. Its diagonalization yields eigenvectors ϕ_α called scaling operators associated to eigenvalues μ_α. In [22], it was shown that those eigenvalues are related to the scaling dimensions Δ_α of the underlying conformal field theory (CFT) by $\Delta_\alpha = \log_3 \mu_\alpha$ where the basis of the log depends on the MERA type (here we consider a ternary MERA for clarity). Scaling operators ϕ_α can be used as observables to extract information about states in higher level of the MERA. Indeed, one can simulate a measurement of $\mathcal{S}^\tau(\phi_\alpha)$ on ρ_τ by measuring the observable ϕ_α on ρ_0. We can analyze the increase in the number of measurements by distinguishing two sources of imprecision. First, to reconstruct ρ_τ one has to use normalized operator $\phi_\alpha^{[\tau]} = 3^{\tau\Delta_\alpha} \mathcal{S}^\tau(\phi_\alpha)$ whose increased statistical fluctuations have to be compensated by performing additional measurements. Second, diagonalizing the Gram matrix of the $\phi_\alpha^{[\tau]}$ will introduce another conditioning factor. However, this Gram matrix is independent of the layer since $G_{\alpha\beta}^{[\tau]} = \mathrm{Tr}\left[\phi_\alpha^{[\tau]} \phi_\beta^{[\tau]}\right] = \mathrm{Tr}\left[\phi_\alpha \phi_\beta\right]$. Thus, the conditioning factor for layer τ will be the product of a factor exponential in the number of layers and a constant factor coming from the orthonormalization. Overall, this amounts to a conditioning factor that scales polynomially with system size.

C. Example

We have performed a numerical simulation to approximate the ground state of the critical Ising model $H = -\sum_{\langle i,j \rangle} \sigma_x^i \otimes \sigma_x^j - \sum_k \sigma_z^k$ based on the MERA structure illustrated on Fig. 6. Due to periodic boundary conditions of the MERA and the translational invariance of the Hamiltonian, all disentanglers and isometries on a given layer are identical. In addition, because the model is critical and hence (nearly) scale invariant, the gates are the same at each level of the MERA, except the first one, where scale invariance has not quite settled in.

Consider the isometry V_1 that maps three physical qubits to one renormalized particle, as illustrated on 7 b). Given V_1, we can compute how the one-site Pauli observables $\{O_0^i\} = \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ are transformed as one-site observables O_1^i on the renormalized particle (see Fig. 7). This linear transformation can be represented by a matrix M such that $O_1^i = \sum_j M_{ij} O_0^j$. In our particular example,

$$M = \begin{pmatrix} 1 & 0.062 & 0 & 0.460 \\ 0 & 0.779 & 0 & -0.213 \\ 0 & 0 & 0.401 & 0 \\ 0 & 0.330 & 0 & 0.294 \end{pmatrix}.$$

Notice that σ_y is mapped on itself so it is a scaling

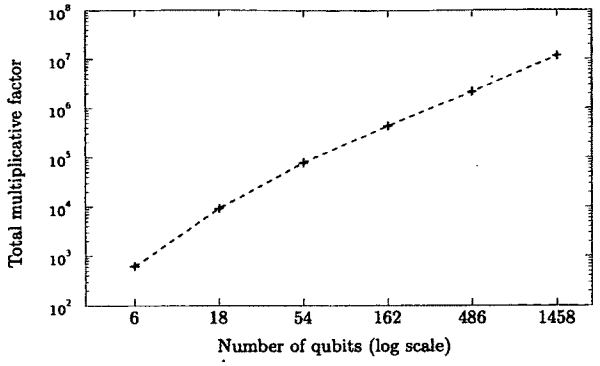


Figure 8: Total multiplicative factor to learn a ternary MERA corresponding to the ground state of the 1D Ising model with critical transverse field. The total number of measurements is the product of this factor with the number of measurements $N(\epsilon) \approx 1/\epsilon^2$ necessary to estimate the expectation value of a Pauli observable to accuracy ϵ on the physical particle. Notice that the curve approaches a straight line, which indicates that the total number of measurements scales *polynomially* with the number of particles.

operator, *i.e.*, the measurement of the observable σ_y on the physical particle is sufficient to simulate the measurement of σ_y on the renormalized particle since $\mathrm{Tr}\rho_1\sigma_y = (0.401)^{-1}\,\mathrm{Tr}\rho_0\sigma_y$. To estimate the expectation value of σ_y on the renormalized particle to some accuracy ϵ thus requires $(0.401)^{-2}N(\epsilon)$ measurements, where $N(\epsilon) \approx 1/\epsilon^2$ is the number of measurements required to estimate σ_y to accuracy ϵ on the physical particle.

Since M is invertible, one can similarly deduce the expectation values of Pauli observables on the renormalized state by taking linear combination of expectation values of observables on the experimental state ρ_0. The fact that the image by V_1 of one-site observables on ρ_0 spans the space of one-site observables on ρ_1 is quite favorable. In general, one could have to consider physical observables on a few sites to generate all renormalized observables. Furthermore, the choice of observables on the physical state could be optimized to minimize the amplification of repeated measurements.

Proceeding similarly for higher renormalization layers, the total number of measurement will be a product of the standard statistical overhead $N(\epsilon) \approx \epsilon^{-2}$ with a total multiplicative factor that results from the conditioning factors for each layers. Figure 8 shows this total multiplicative factor for the critical Ising model with the MERA structure shown at Fig. 6. The total number of qubits is $n = 2 \times 3^k$ for $k \leq 5$. Recall that quantum state tomography of the 8 qubit W state required more than 656 100 measurements in [3]. Using the same elementary precision, textiti.e., $N(\epsilon) = 100$, our method allows to learn a 18-qubit system with a comparable number of measurements (917 656 measurements).

D. Error propagation and certification without unitary control

In the scheme with unitary control, certification is made possible thanks to experimental measurements that directly project the experimental state on a particular MERA state, by making sure that the ancillary qubits are disentangled after each layer of renormalization. Keeping records on the success rate of these measurements thus yields information about the overlap. Not only does this allow for direct fidelity estimate, but the resulting post-selection also limits error propagation, c.f. Sec. III B 1.

Using a scheme similar to the one explained in Sec. IV B, one could simulate the physical measurement of the disentangled particles on the level of the experimental state. While this is possible, it requires in principle an exponential number of measurements since a single-qubit observable at a high level of the MERA will typically have support on its entire causal cone. That statement can seem contradictory to our claim of Sec. IV B that density matrices on high layers of the MERA can be estimated by using only a polynomial number of measurements. The apparent contradiction arises from the fact that efficient estimation is based on the assumption that the state is well-approximated by a MERA. Obviously, one cannot rely on that property to certify that the state is well approximated by a MERA.

Despite this difficulty, we will now argue that it is possible to certify the state without unitary control, using the 18-qubit ternary MERA of Fig. 6 as a concrete example. As a first step, we will study how errors propagate during the MERA learning process. Because a small change in the state can in principle alter the MERA structure substantially, it is difficult to rigorously bound this error propagation. Thus, this discussion, as well as appendix A, should be seen as plausible arguments for the scaling of errors rather than a rigorous proof.

We note $|\psi_0\rangle$ the experimental state and $|\psi_1\rangle$ the state after the first layer of renormalization. Note U_1 the unitary that gives rise to the isometry V_1, *i.e.*,

$$V_1 = (|0\rangle\langle 0| \otimes \mathbb{I} \otimes |0\rangle\langle 0|)\, U_1. \tag{20}$$

Note $\mathbb{U}_1$ the unitary that corresponds to the first layer of renormalization. Without post-selection, the state after that layer reads

$$|\psi_1\rangle = \mathbb{U}_1|\psi_0\rangle = \frac{|0\rangle^{\otimes 12}|\eta_1\rangle + |e_1\rangle}{\sqrt{1+\epsilon_1}} \tag{21}$$

where $|e_1\rangle$ is a sub-normalized error vector whose square norm is $\epsilon_1 \equiv \langle e_1|e_1\rangle$.

Our scheme to learn the MERA presented in Sec. IV B uses the identity $\mathrm{Tr}\left[O_1^i|\eta_1\rangle\langle\eta_1|\right] = \mathrm{Tr}\left[O_0^i|\psi_0\rangle\langle\psi_0|\right]$—valid when $\epsilon_1 = 0$ or when post-selecting on the all $|0\rangle$ state of ancillary qubits—to estimate the expected value of O_1^i on $|\eta_1\rangle$ through measurements of O_0^i on the physical state $|\psi_0\rangle$. Repeating with various O_1^i enabled us to reconstruct the density matrices σ_1^i that are marginals of

$|\eta_1\rangle$ on small blocks of particles. However, in the presence of errors and without post-selection, this identity does not hold. Using Eqs. (20) and (21), we obtain

$$\left|\mathrm{Tr}\left[O_1^i|\eta_1\rangle\langle\eta_1|\right] - \mathrm{Tr}\left[O_0^i|\psi_0\rangle\langle\psi_0|\right]\right| \leq 2\epsilon_1. \quad (22)$$

Thus the reconstruction algorithm will estimate the reduced density matrices $\sigma_1^{[i]}$ up to some error $E_1 \in O(\epsilon_1)$. The precise relation between E_1 and ϵ_1 depends on the method (linear inversion, maximum-likelihood, etc.) used to reconstruct the density matrix from the measurements. The disentangling algorithm will then take those error-prone contaminated reduced density matrices $\sigma_1^{[i]}$ as input to identify the disentanglers and isometries of the second renormalization layer. As a consequence, even in the absence of any other source of errors, these disentanglers and isometries will typically be chosen sub-optimally. Thus, we see that errors accumulate along the renormalization flow.

The accumulated error E_2^{ac} after the second layer of renormalization will contain a component inherited from the accumulated error E_1^{ac} of the previous layer and an intrinsic error ϵ_2. As before, the intrinsic error can be caused by a numerically sub-optimal choice of disentangler or simply because the state is not exactly a MERA. The distinction between this accumulated error E_k^{ac} and the accumulated error in the presence of post-selection, denoted ϵ_k^{ac} in Sec. III B, stems from the fact that the algorithm used to find the disentangler operates on a contaminated state. The analysis presented in appendix A shows that if there is an accumulated error E_k^{ac} after k layers of renormalization, the error E_{k+1}^{ac} is bounded by

$$1 + E_{k+1}^{ac} \leq \left(1 + E_k^{ac}\right)^3 \left(1 + \epsilon_{k+1}\right) \quad (23)$$

which results in an error that grows exponentially with the number of layers. The number of layers being logarithmic in the number of particles, the error thus scales polynomially with system size. The fidelity between the experimental state and the reconstructed state is given by the final accumulated error

$$|\langle\psi_0|\psi_{tomo}\rangle|^2 = \left(1 + E_{top}^{ac}\right)^{-1}. \quad (24)$$

For a ternary MERA with $n = 2 \times 3^k$ particles, we can use Eq. (23) and (24) to bound the fidelity as a function of the intrinsic error made at each layer of the MERA

$$|\langle\psi_0|\psi_{tomo}\rangle|^2 \geq \left(1+\epsilon_1\right)^{-n/3} \left(1+\epsilon_2\right)^{-n/3^2} \ldots \left(1+\epsilon_{top}\right)^{-1}. \quad (25)$$

Given this relation between fidelity and intrinsic errors, we now turn to the problem of certification.

To certify the fidelity of the reconstructed state, one therefore needs to estimate the intrinsic error ϵ_k made at each layer. This error is the probability that the disentangled particles are not in the $|0\rangle$ state. With unitary control, estimation of ϵ_k could be performed by experimentally projecting all those ancillary qubits in the $|0\rangle$ state and accumulating statistics. Without unitary control, estimation of ϵ_k has to be performed *locally* by estimating the projection orthogonal to $|0\rangle\langle 0| \otimes \mathbb{I} \otimes |0\rangle\langle 0|$ for each isometry. The expectation value of this projector, the *leakage error* f_k^i, is precisely the objective function used in our numerical procedure, c.f. Eq. (3). In order to estimate the error ϵ_k for the layer k, we use the union bound, which results in an estimate that scales with the number of isometries in the layer. Supposing that for each isometry the leakage error is smaller than ϵ, we get the upper bound

$$\epsilon_k \leq \sum_i f_k^i \leq n\epsilon/3^k. \quad (26)$$

Combining Eq. (25) and (8), we get

$$|\langle\psi_0|\psi_{tomo}\rangle|^2 \geq \prod_{k=1}^{\log_3 n} \left(1 + \frac{n}{3^k}\epsilon\right)^{-n/3^k}. \quad (27)$$

Assuming that ϵ i smaller than $1/n^2$, the right hand side of (27) reduces to $1 - \epsilon\sum_{k=1}^{\log_3 n}\left(\frac{n}{3^k}\right)^2 \sim 1 - \frac{n^2}{8}\epsilon$. Thus, for states that are at most $1/n^2$ away in fidelity from a MERA state, we can certify the distance between the reconstructed and the experimental states. This certificate is rather loose due to the use of the union bound and assuming that error accumulate in the worst possible way. A more accurate estimate of the fidelity can be performed using the Monte Carlo technique of [29] but requires a number of measurements that will scale exponentially with system size.

V. DISCUSSION

In this work, we have presented a tomography method that allows to efficiently learn the MERA description of a state by patching together tomography experiments on a few particles and using fast numerical processing. The method is heuristic but works very well in numerical simulations. A complete analytic understanding of how to find an optimal disentangler at each step would be desirable, but may well be intractable. With regards to experimental use, the method should be thought of as a proof of principle and is flexible enough to be adapted to accommodate many experimental constraints.

One issue of fundamental interest raised by our work is the relationship between the numerical tractability of a variational family of states and the ability to learn efficiently the variational parameters. In order to be interesting, variational family of states must not only be described by a small number of parameters, but also allow for the efficient numerical computation of quantities of interest, such as the energy of the system, correlation functions, or more generally expectation values of local observables. On its own, an efficient representation is of limited computational usefulness. For instance, the Gibbs state or ground state of a local Hamiltonian is

described by a few parameters — a temperature and a local Hamiltonian — but does not allow to extract physical quantities of interest efficiently. Another example is the variational family of projected entangled pair states or PEPS [30], the generalization of MPS to system in more than one dimension. While PEPS have been instrumental in better understanding of quantum many-body systems, they are in general intractable numerically [31].

Is there a relation between numerical tractability and efficient tomography? The method presented in [17] to learn a MPS from local measurements made explicit use of the energy minimization algorithm for MPS; namely DMRG [14, 15]. This example suggests that numerical tractability could imply that learning the variational parameters is possible. In that regard, MERA are intriguing states because they live at the frontier of tractability. Indeed, in more than 1 dimension, MERA states are a subclass of PEPS [32] with a bond dimension independent of system size [33]. While the computation of expectation values of local observables is believed to be intractable for PEPS, it is efficient for MERA. In one dimension, MERA can be seen as MPS if one allows the bond dimension to grow *polynomially* with the size of the system (while MPS are usually required to have a *constant* bond dimension). Thus, while MPS manipulations typically have a computational cost linear in the number of particles, 1D-MERA manipulations have a computational cost which is super-linear (but yet polynomial).

Beyond MPS and MERA, one could consider states obtained from a quantum circuit where the positions of the gates are known and try to identify those gates. An interesting question is then to characterize what topology of circuits makes it possible to learn gates efficiently. This could lead to formal methods for the testing and verification of quantum hardware.

Acknowledgments

This work was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). OLC acknowledges the support of NSERC through a Vanier scholarship. DP acknowledges financial support by the Lockheed Martin Corporation. We thank Marcus da Silva and Andy Ferris for stimulating discussions and Andy Ferris for sharing numerical data on scale-invariant MERAs.

[1] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
[2] S. Das Sarma, G. Gervais, and X. Zhou, Phys. Rev. B **82**, 115330 (2010).
[3] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chekal Kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, et al., Nature **438**, 643 (2005).
[4] J. T. Barreiro, P. Schindler, O. Gühne, T. Monz, M. Chwalla, C. F. Roos, M. Hennrich, and R. Blatt, Nature Physics **6**, 1 (2010).
[5] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, et al., Nature **460**, 240 (2009).
[6] L. DiCarlo, M. D. Reed, L. Sun, B. R. Johnson, J. M. Chow, J. M. Gambetta, L. Frunzio, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf, arXiv 1004.4324 (2010).
[7] S. Filipp, P. Maurer, P. Leek, M. Baur, R. Bianchetti, J. Fink, M. Göppl, L. Steffen, J. Gambetta, a. Blais, et al., Phys. Rev. Lett. **102**, 1 (2009).
[8] H. Mikami, Y. Li, K. Fukuoka, and T. Kobayashi, Phys. Rev. Lett. **95**, 2 (2005).
[9] K. Resch, P. Walther, and a. Zeilinger, Phys. Rev. Lett. **94** (2005).
[10] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987).
[11] M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).
[12] G. Vidal, Phys. Rev. Lett. **91**, 12 (2003).
[13] G. Vidal, Phys. Rev. Lett. **93**, 1 (2004).
[14] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992), URL `http://link.aps.org/doi/10.1103/PhysRevLett.69.2863`.
[15] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
[16] F. Verstraete and J. Cirac, Phys. Rev. B **73** (2006).
[17] M. Cramer, M. Plenio, S. Flammia, R. Somma, D. Gross, S. Bartlett, O. Landon-Cardinal, D. Poulin, and Y. Liu, Nature Communications **1**, 149 (2010).
[18] G. Vidal, Phys. Rev. Lett. **101**, 1 (2008).
[19] G. Evenbly and G. Vidal, Phys. Rev. B **79**, 144108 (2009).
[20] G. Evenbly, R. N. C. Pfeifer, V. Picó, S. Iblisdir, L. Tagliacozzo, I. P. McCulloch, and G. Vidal, Phys. Rev. B **82**, 161107 (2010).
[21] S. Montangero, M. Rizzi, V. Giovannetti, and R. Fazio, Phys. Rev. B **80**, 2 (2009).
[22] R. Pfeifer, G. Evenbly, and G. Vidal, Phys. Rev. A **79**, 2 (2009).
[23] L. Cincio, J. Dziarmaga, and M. Rams, Phys. Rev. Lett. **100**, 2 (2008).
[24] M. Aguado and G. Vidal, Phys. Rev. Lett. **100**, 1 (2008).
[25] R. König and E. Bilgin, Phys. Rev. B **82**, 1 (2010).
[26] R. Pfeifer, P. Corboz, O. Buerschaper, M. Aguado, M. Troyer, and G. Vidal, Phys. Rev. B **82**, 1 (2010).
[27] G. Vidal, arXiv **0707.1454v2** (2008).
[28] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).
[29] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).
[30] F. Verstraete, V. Murg, and J. Cirac, Advances in Physics **57**, 143 (2008).
[31] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **98**, 140506 (2007).
[32] F. Verstraete and J. Cirac, arXiv:cond-mat/0407066. (2004).
[33] T. Barthel, M. Kliesch, and J. Eisert, Phys. Rev. Lett. **105**, 6 (2010).

Appendix A: Error accumulation without post-selection

The modified scheme that circumvents the need for unitary control modifies the error propagation. Namely, the scaling of the overall error increases since the error at each step will depend on previous errors, that will accumulate and amplify subsequent errors. In this section, we present a heuristic argument to understand how this accumulation affects the overall precision of the scheme.

Because there is no post-selection, the algorithm used to find the disentanglers and isometries will not operate on the state $|\eta_k\rangle$ (see Eq. (8)) but instead will operate on the contaminated state

$$(1-\varepsilon)\,|\eta_k\rangle\langle\eta_k| + \varepsilon |E_k^{ac}\rangle\langle E_k^{ac}| \tag{A1}$$

where $\varepsilon \equiv \frac{E_k^{ac}}{1+E_k^{ac}}$ and $|E_k^{ac}\rangle$ is a subnormalized error vector resulting from the accumulation of all previous errors, and whose square norm is $E_k^{ac} \equiv \langle E_k^{ac}|E_k^{ac}\rangle$. Thus, the numerical minimization returns a unitary that is typically not the optimal disentangler for $|\eta_k\rangle$.

In the degenerate case—when the objective function Eq. 3 has many distinct minima (modulo gauge)—this might change the disentangling unitary drastically, either because the objective function is flat or because the solution jumps from one local minima to another. In the latter case, the errors is causing the algorithm to explore different local minima, which is actually an exploration that is desirable to find the global minimum. Degenerate minima correspond to hard instances of the problem, and it is conceivable that in these cases the state can neither be learned nor certified. In the non-degenerate case however, we can heuristically bound the accumulation of errors. We proceed in three steps. First, we analyze how the modification of the input state will affect the disentangling unitary returned by the algorithm. Second, we evaluate how this imperfect disentangler impacts the error propagation. Third, we bound the error at step $k+1$ in terms of the error at step k. This technical result is used in section IV D.

1. Disentangling unitary without post-selection

Let us denote by $\tilde{U} = e^{i\tilde{H}}$ the unitary that is returned by our algorithm in the presence of post-selection, *i.e.* the unitary that minimizes the objective function Eq. (3) for the post-selected state ρ. If we don't post-select on the ancillary particles being disentangled, this minimization is not performed on the perfect state $|\eta_k\rangle$ but rather on the contaminated state given by Eq. (A1). We want to know how much $\tilde{U} \doteq \arg\min_U f(U,\rho)$ changes when ρ goes from $|\eta_k\rangle$ to the state of Eq. (A1).

Using the chain rule, we formally write $\frac{\partial \tilde{U}}{\partial \rho} = \frac{\partial \tilde{U}}{\partial f}\frac{\partial f}{\partial \rho}$. The first term, $\frac{\partial \tilde{U}}{\partial f}$, quantifies how much $\tilde{U}$ changes when the objective function changes for a given ρ. In the non-degenerate case, we expect this term to be bounded in norm by a Lipschitz constant η. The second term, $\frac{\partial f}{\partial \rho}$, evaluates how the objective function changes when the state changes. Recalling that the objective function is a sum of eigenvalues and using non-degenerate perturbation theory, this term is going to be proportional to ε defined by Eq. (A1). Thus, instead of $\tilde{U} = e^{i\tilde{H}}$, the minimization algorithm returns $e^{i(\tilde{H}+\varepsilon\eta A)} \approx W\tilde{U}$ where the anomalous unitary

$$W = e^{i\varepsilon\eta A} \tag{A2}$$

quantifies the perturbation to the perfect disentangler due to the presence of error. Note that A is a Hermitian operator of norm of order 1.

2. Impact of the imperfect disentangler on error propagation

At step $k+1$, the anomalous unitary acts on the state of Eq. (10)

$$W_{k+1}\frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |e_{k+1}^{ac}\rangle}{\sqrt{1+\epsilon_{k+1}^{ac}}} \tag{A3}$$

where $\epsilon_{k+1}^{ac} \equiv \langle e_{k+1}^{ac}|e_{k+1}^{ac}\rangle$ is the error resulting from all previous steps but assuming the disentangler used at step $k+1$ is the exact one. Up to an overall phase, we can rewrite the state (A3) as

$$\frac{|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle + |E_{k+1}^{ac}\rangle}{\sqrt{1+E_{k+1}^{ac}}} \tag{A4}$$

where the error vector $|E_{k+1}^{ac}\rangle$ (whose square norm is E_{k+1}^{ac}) now takes into account the imperfect disentangler.

Comparing Eq. (A3) and (A4), we see that error E_{k+1}^{ac} relates to the error ϵ_{k+1}^{ac} through

$$1 + E_{k+1}^{ac} = 1 + \epsilon_{k+1}^{ac}/\beta^2 \tag{A5}$$

where $\beta = \left|\langle\eta_{k+1}|\langle 0|^{\otimes k+1} W |0\rangle^{\otimes k+1}|\eta_{k+1}\rangle\right|$.

3. Error propagation without post-selection

Using $W = e^{i\varepsilon\eta A}$, calculations show that

$$\beta^2 = 1 - \varepsilon^2\eta^2\left(\langle A^2\rangle - \langle A\rangle^2\right) = 1 - \varepsilon^2\eta^2\Delta^2 \tag{A6}$$

plus $O(\varepsilon^4\eta^4)$ terms, where the variance Δ^2 of A with respect to state $|0\rangle^{\otimes k+1}|\eta_{k+1}\rangle$ appears. Recalling that $\varepsilon = \frac{E_k^{ac}}{1+E_k^{ac}}$, we can bound β^2 by

$$\beta^2 = \frac{\left(1+E_k^{ac}\right)^2 - \left(E_k^{ac}\right)^2\eta^2\Delta^2}{\left(1+E_k^{ac}\right)^2} \geq \frac{1}{\left(1+E_k^{ac}\right)^2} \tag{A7}$$

for any E_k^{ac} if $\eta^2\Delta^2 \leq 1$ or for small E_k^{ac} otherwise.

The recurrence relation for errors given by Eq. (12) in the case with post-selection now relates the accumulated error ϵ_{k+1}^{ac} to the previous accumulated error E_k^{ac}

$$1 + \epsilon_{k+1}^{ac} = (1 + \epsilon_{k+1})(1 + E_k^{ac}). \quad \text{(A8)}$$

Combining Eq. (A8) and the bound on β given by Eq. (A7), Eq. (A5) becomes

$$1 + E_{k+1}^{ac} \leq (1 + E_k^{ac})^3 (1 + \epsilon_{k+1}) \quad \text{(A9)}$$

which indicates how errors proliferate when post-selection is not possible.

Appendix B: Comparing a reconstructed MERA to a predicted MERA

In this Section, we describe a polynomial algorithm to contract two MERA states, thus allowing to compute their fidelity. This algorithm is of practical interest for comparing a MERA whose parameters have been identified experimentally using our method to a predicted MERA state –found by numerical optimization for instance. Notice that contracting two different MERA states also allows to compute expectation values of tensor product of local observables $\bigotimes_i A_i$ since it suffices to contract the original state $|\psi\rangle$ and the modified state $|\phi\rangle = \bigotimes_i A_i|\psi\rangle$, which is also a MERA state.

Defining a method to contract two MERA states is equivalent to giving a prescription on how to sequentially contract the tensor network resulting to joining two MERA states. Recall that contracting two tensors $(M)_{i_a j_b}$ and $(N)_{k_b \ell_c}$ to obtain $T_{i_a \ell_c} = \sum_{j_b} M_{i_a j_b} N_{j_b \ell_c}$ has a computational cost of $a \times b \times c$ where a is the number of values that the index i_a can take b and c are defined in the same way with respect to j_b and ℓ_c. In a tensor network, every tensor is usually represented with a number of bonds that each represent an index that has the same maximal number of possible values. For a MERA, this maximal bond dimension is usually denoted by χ.

The main idea to contract efficiently two MERA states is essentially to turn them into two MPS before contracting them. We look at the MERA circuit as having $n/2$ columns of gates vertically and $\log_\chi n - 1$ renormalization layers horizontally. The sequence of contraction is to sequentially contract every tensor in the leftmost column to create a tensor with a large number of bonds that will then contract with every tensor in the next column. The maximal number of bonds that this leftmost tensor will have throughout the contraction of the network is given by the maximal number of bonds that are opened when taking a vertical cut in the tensor network. For a single MERA, cutting through each of the $\log_\chi n - 1$ layer opens up two bonds, one for the rightmost incoming edge of the isometry and one for the outgoing edge of the isometry. Thus, for the contraction of two MERAs, the maximum number of bonds for a vertical cut is bounded by $\max \# = 2 \times 2 \times \log_\chi n = 4 \log_\chi n$, which is verified numerically (see top of Fig. 9). Since at every contraction step, the leftmost tensor with a large number of bonds contract with another tensor that has at most two bonds in addition to the ones being contracted, the maximum cost of one contraction is $\chi^{\max \#} \chi^2 = \chi^2 n^4$. Finally, there are $O(n)$ disentanglers and isometries to contract so the total cost of contracting the network is bounded by $O(n^5)$. Actual numerical simulations show that this bound is probably not tight (see bottom of Fig. 9).

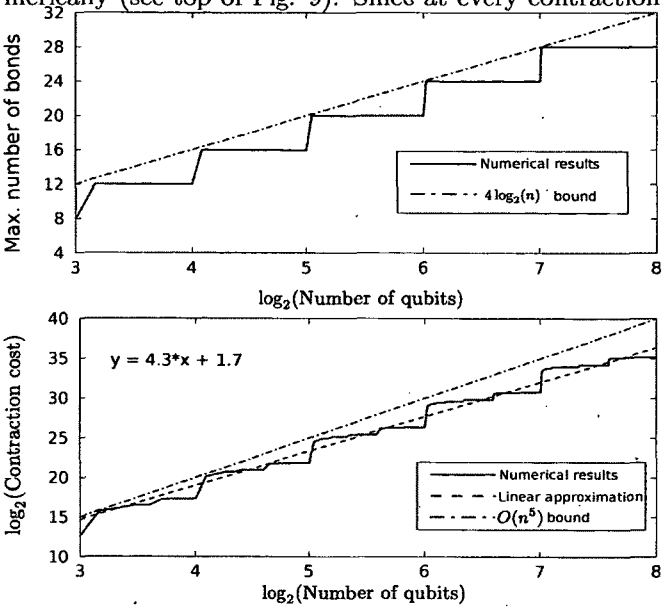


Figure 9: (top) Maximum number of bonds during the contraction procedure as a function of the logarithm of the number of qubits n. Numerical results (solid blue line) are consistent with the expected bound of $4 \log_2 n$. (bottom) Contraction cost C as a function of the number n of qubits on a log scale. Numerical results (solid blue line) are consistent with the $O(n^5)$ bound (dot dashed green line) but linear approximation (dashed red line) indicate that the cost scales like a smaller power of n, namely $C \simeq n^{4.3}$.

4.7 Discussion

4.7.1 Topologie des circuits « apprenables »

Pour aller au-delà de nos résultats sur les MPS et les MERA, il serait intéressant de déterminer quels sont les circuits « apprenables ». Le problème est le suivant : étant donné la topologie d'un circuit préparant un état $|\psi\rangle$, *i.e.* l'emplacement de ses portes unitaires (mais pas leur description), ainsi que plusieurs copies de $|\psi\rangle$, est-il possible d'apprendre le circuit préparateur ?

Une caractéristique essentielle utilisée pour les MPS et les MERA est l'apport progressif de qudits ancillaires dans l'état $|0\rangle$ dans le circuit. Cela permet de mettre en place des techniques variationnelles afin d'identifier les transformations unitaires qui agissent sur l'état expérimental afin de désintriquer ces qudits ancillaires. Ceci fournit un critère qui permet d'apprendre un circuit.

Une autre caractéristique des MPS et des MERA en 1D est que la séquence à suivre afin d'apprendre le circuit est claire, en raison de la structure 1D sous-jacente. Pour des états à réseau de tenseurs en dimension spatiale supérieure, cela n'est pas clair. Il serait toutefois intéressant de réfléchir à une méthode d'apprentissage pour les PEPS (*projected entangled pair states*), qui sont l'extension naturelle des MPS en 2D.

4.7.2 Autres classes variationnelles : PEPS

4.7.2.1 Extension 2D des MPS

Les PEPS (*projected entangled pair states*) ont été définis en étendant l'approche AKLT des MPS, présentée en annexe A.2, à des systèmes 2D [77]. Pour représenter un état sur $n = L^2$ qudits, l'idée est de construire un état ressource sur $\mathcal{O}(4n)$ quDits préparés dans des états maximalement intriqués, cf. éq (A.14). Ensuite, un opérateur projète 4 quDits virtuels vers 1 qudit réel, comme sur la figure 4.12.

Comme pour les MPS, on peut aussi voir les PEPS comme un cas particulier d'états à réseau de tenseurs. Pour les PEPS, les tenseurs sont de rang 5 avec un indice physique et 4 indices virtuels.

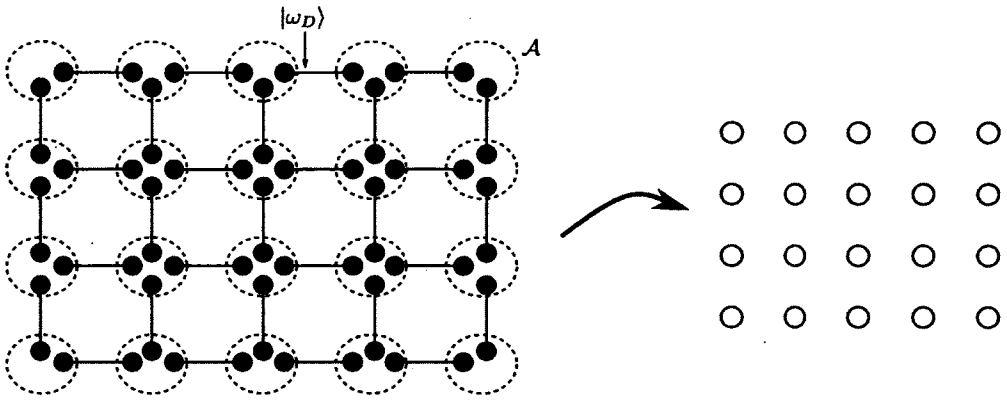


Figure 4.12 Construction d'un PEPS à partir d'un état ressource.

4.7.2.2 PEPS non-physiques

Les PEPS, bien qu'ils aient une représentation concise, n'ont pas une description efficace. En effet, il existe des PEPS qui ne permettent pas de calculer des quantités physiques (en supposant que P $\neq$ NP) [78]. Ce résultat découle de la difficulté [8] de contracter un PEPS, pour calculer sa norme par exemple. Toutefois, numériquement, il semble possible de contracter approximativement les PEPS et une meilleure compréhension de ce phénomène existe [79]. Il semble donc qu'il existe une sous-classe de PEPS qui ont une description efficace, *i.e.*, qui permettent le calcul efficace de quantités physiques. Il serait intéressant de mieux comprendre la structure de cette sous-classe de PEPS et de déterminer s'il existe un protocole d'apprentissage pour ces états.

8. Ce problème permet de calculer le nombre de chemins pour une machine de Turing non déterministe (#P-ardu).

Deuxième partie

Mémoires quantiques auto-correctrices 2D

Chapitre 5

Systèmes topologiques

Dans ce chapitre, nous allons nous intéresser à la notion d'ordre d'un système physique et plus particulièrement aux systèmes topologiquement ordonnés qui présentent des corrélations non-locales qui ne peuvent pas être capturées par un *paramètre d'ordre* local. Après avoir brièvement rappelé quelques notions de base sur la notion d'ordre, nous introduirons un premier exemple, très graphique, de système topologique : un liquide de spin. Avec cette intuition, nous énoncerons la définition et les propriétés d'un système topologique. Nous verrons en détails comment ces propriétés se manifestent dans l'exemple canonique d'un système topologique : le code torique, introduit par Kitaev en 1997 [80].

En particulier, nous verrons que les systèmes topologiques ont un espace fondamental dégénéré, ce qui permettrait d'encoder de l'information quantique dans cet espace. De plus, puisque l'information est encodée dans des degrés de liberté topologiques, elle risque d'être résistante à l'action locale de l'environnement. Ainsi, les systèmes topologiques fournissent les candidats à une mémoire quantique auto-correctrice, que nous définirons au chapitre 6.

5.1 Paramètre d'ordre local

Nous allons discuter brièvement de la notion de paramètre d'ordre local d'un système physique. En particulier, nous verrons sur l'exemple du modèle d'Ising quantique que la phase ordonnée présente un espace fondamental dégénéré, ce qui est essentiel afin d'encoder de l'information quantique. En effet, une mémoire quantique doit être capable d'encoder une superposition arbitraire d'états. Malheureusement, l'existence d'un paramètre d'ordre local qui distingue les états fondamentaux permettrait à un environnement de détruire la cohérence entre ces états. Ainsi, nous serons amenés à considérer les systèmes topologiques car ils ne possèdent pas de paramètres d'ordre locaux.

5.1.1 Paramètre d'ordre dans une transition de phase

Une phase est une classe d'états d'un système physique dont les propriétés sont uniformes et présentent un comportement similaire. On peut étendre cette notion à des classes de hamiltoniens en considérant la phase de leurs états fondamentaux. Suite à la modification d'un paramètre externe, le système peut passer d'une phase à une autre, c.-à-d. subir une *transition de phase*. Classiquement, le paramètre externe est typiquement la température, comme dans l'exemple courant de la transition eau-glace. Toutefois, à température nulle, il est aussi possible d'avoir une transition de phase dite quantique [81], liée aux fluctuations quantiques.

En effet, soit $H(g)$ un hamiltonien dépendant d'un couplage adimensionnel g. Suivant la valeur de g, les propriétés des états fondamentaux peuvent être très différentes. Par exemple, il peut apparaître deux phases distinctes, l'une désordonnée et l'autre ordonnée, séparées par une transition de phase pour une valeur critique g_c du paramètre de couplage.

Dans certains cas, la phase ordonnée correspond à un espace fondamental dégénéré, ce qui est exactement ce que l'on recherche afin de stocker de l'information quantique. Par exemple, pour une dégénérescence 2, il sera possible d'identifier deux états fondamentaux $|\Omega_1\rangle$ et $|\Omega_2\rangle$ comme les états $|0\rangle$ et $|1\rangle$ d'un qubit encodé de telle façon que tout état encodé $\alpha|0\rangle + \beta|1\rangle$ soit aussi un état fondamental.

Or, pour la plupart des systèmes, il existe un paramètre d'ordre local, *i.e.*, une observable locale L dont la valeur moyenne est nulle pour un fondamental de la phase désordonnée ($g > g_c$)

et non-nulle dans la phase ordonnée ($g < g_c$). De plus, ce paramètre d'ordre local distingue typiquement les différents états fondamentaux de la phase ordonnée car différentes valeurs du paramètre d'ordre sont associées à différents états fondamentaux). Or, dans ce cas, un champ externe couplée à ce paramètre d'ordre (p.ex. un champ magnétique) pourra lever la dégénéresence de l'espace fondamental et favoriser certains états initialement dégénérés. Plus généralement, un paramètre d'ordre local ouvrira la porte à des phénomènes de décohérence qui empêcheront de préserver la cohérence d'un état encodé dans l'espace fondamental. Afin de rendre ces notions plus concrètes, nous allons nous intéresser à la transition de phase dans le modèle d'Ising quantique.

5.1.2 Modèle d'Ising

Soit un réseau occupé par des particules de spin-1/2 dont le hamiltonien est

$$H_I = -Jg\sum_i \sigma_i^x - J\sum_{\langle i,j\rangle} \sigma_i^z\sigma_j^z \tag{5.1}$$

où σ^x et σ^z sont les matrices de Pauli, $J > 0$ est un cofficient réel et la deuxième somme est effectuée sur les premiers voisins $\langle i,j\rangle$. L'interaction $\sigma_i^z\sigma_j^z$ représente la tendance de deux spins adjacents de s'aligner dans le même sens suivant la direction z alors que le terme $\sum_i \sigma_i^x$ représente un champ magnétique externe orienté suivant la direction x.

5.1.2.1 Phase désordonnée

Plaçons-nous dans la limite où $g \to +\infty$, où le champ magnétique domine l'interaction spin-spin. Dans ce cas, l'état fondamental $|\psi\rangle$ est celui où chaque particule est dans l'état $|\rightarrow\rangle \equiv \frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle$, qui correspond à un spin orienté dans la direction x, c.-à-d.

$$|\psi\rangle = \bigotimes_i |\rightarrow\rangle_i. \tag{5.2}$$

Dans ce cas, chaque particule est indépendante. En particulier, la valeur moyenne de l'opérateur $\sigma_i^z\sigma_j^z$ est nulle si les sites i et j sont distincts. Évidemment, il s'agit ici d'un cas limite. En pratique, on aura plutôt $g \gg 1$ et les corrélateurs auront une décroissance exponentielle en fonction de la distance entre les sites

$$\langle\psi|\sigma_i^z\sigma_j^z|\psi\rangle \sim e^{-|x_i - x_j|/\xi} \tag{5.3}$$

qui fera apparaître une *longueur de corrélation* ξ.

Cette phase n'est pas intéressante pour le stockage d'information quantique car l'espace fondamental n'est pas dégénéré.

5.1.2.2 Phase ordonnée

Intéressons-nous maintenant à la limite $g = 0$. Dans ce cas, l'espace fondamental est dégénéré deux fois. Il est engendré par l'état $\bigotimes_i |\uparrow\rangle_i$ où chaque spin pointe dans la direction z et l'état $\bigotimes_i |\downarrow\rangle_i$ où chaque spin pointe dans la direction $-z$. En fait, cette dégénérescence survit pour un système infini même si g n'est pas strictement nul. Elle est liée au fait que le hamiltonien est invariant sous la symétrie $\sigma^z \mapsto -\sigma^z$ et $\sigma^x \mapsto \sigma^x$. Cette symétrie est idempotente puisque l'appliquer deux fois n'a aucun effet. On parle alors de symétrie $\mathbb{Z}_2$. Ainsi, les états $\bigotimes_i |\uparrow\rangle_i$ et $\bigotimes_i |\downarrow\rangle_i$ brisent la symétrie du système. Dans le cas plus raisonnable où g n'est pas strictement nul mais très petit devant 1, les corrélateurs ont une valeur propre non-nulle à des distances arbitraires

$$\lim_{|x_i - x_j| \to \infty} \langle \psi | \sigma_i^z \sigma_j^z | \psi \rangle = N_0^2 \tag{5.4}$$

où l'aimantation définie par $\langle \sigma_i^z \rangle$ joue le rôle de *paramètre d'ordre*. En effet, $\langle \sigma_i^z \rangle$ est nul dans la phase désordonnée ($g < g_c$) car $\langle \rightarrow |\sigma_z| \rightarrow \rangle = 0$. Par contre, il est non-nul pour la phase ordonnée ($g > g_c$) et permet de distinguer les états fondamentaux. L'état fondamental issu de $\bigotimes_i |\uparrow\rangle_i$ aura une aimantation positive $\langle \sigma^z \rangle = +N_0$ alors que l'état fondamental issu de $\bigotimes_i |\downarrow\rangle_i$ aura une aimantation négative $\langle \sigma^z \rangle = -N_0$. L'existence de ce paramètre d'ordre local permettra donc à l'environnement de détruire la cohérence d'une superposition arbitraire de ces deux états.

5.1.3 À la recherche de phases sans paramètre d'ordre local

Afin de mettre au point une mémoire quantique, on recherche donc un système qui ne possède pas de paramètre d'ordre local. Or, une classification très générale des phases due à Landau postule l'existence de ce paramètre d'ordre local. Plus spécifiquement, Landau s'intéressa aux transitions qui font intervenir des phases dont le groupe de symétrie de la phase ordonnée est contenu dans le groupe de symétrie de la phase désordonnée[1]. Il nota que le passage d'une symétrie à l'autre

1. Il peut paraître a priori étonnant que la phase désordonnée soit plus symétrique que la phase ordonnée. Considérons l'exemple de la transition eau-glace. La phase désordonnée, l'eau, est invariante pour toute translation. Au contraire, la phase ordonnée, la glace, n'est invariante que pour certains vecteurs de translation, qui définissent le

pouvait être quantifiée par un paramètre d'ordre, i.e une grandeur physique qui ne soit non-nulle que pour la phase ordonnée et qui permette de distinguer les états qui brisent la symétrie. Il faut donc aller au-delà de la classification de Landau pour s'affranchir d'un paramètre d'ordre local.

La classification de Landau, en particulier après l'incorporation des « fluctuations quantiques » avec Ginzburg, a connu un grand succès. Toutefois, suite à la découverte de la supraconductivité à haute température en 1986, les théoriciens ont commencé à étudier des états de liquide de spins [2]. Or, il fut rapidement remarqué que des états de liquide de spin différents avaient exactement la même symétrie [83]. Ainsi, la classification de Landau basée sur la symétrie ne parvenait pas à distinguer ces états. En particulier, ces états n'ont pas de paramètre d'ordre local et on parle alors d'ordre topologique [3]. La prochaine section va donc porter sur ces systèmes qui semblent prometteurs pour encoder de l'information quantique.

5.2 Ordre topologique

Certains modèles physiques présentent un ordre à longue portée ne correspondant *pas* à un paramètre d'ordre *local*. L'état fondamental des ces modèles exhibent des corrélations non-locales qui ne peuvent être capturées par des corrélateurs de type $\langle \sigma^z \sigma^z \rangle$. Seule la mesure d'observables faisant intervenir un grand nombre de particules (qui grandit avec la taille du réseau) permet de détecter un tel ordre, qualifié de *topologique* et parfois appelé ordre quantique.

5.2.1 Exemple introductif : liquide de spin

Afin de fixer graphiquement les idées, nous allons nous intéresser à un exemple de modèle topologique, celui des liquides de spin. Il a l'avantage de donner une intuition graphique de l'ordre topologique. Plus précisément, nous nous intéresserons aux modèles de dimères [84]. Ces modèles sont utiles pour décrire la dynamique de dimères présents dans les phases antiferromagnétiques désordonnées de spins-1/2. Deux phases génériques peuvent apparaître : un liquide de spin qui ne brise aucune symétrie et un cristal de dimères. Nous allons nous intéresser au liquide de spin, qui

réseau du cristal. Dans ce sens, l'eau est plus symétrique que la glace.

2. Pour une discussion historique, voir [82].

3. Historiquement, le terme topologique provient du fait que les modèles effectifs des hamiltoniens dont les fondamentaux présentent de l'ordre topologique sont des théories de champ quantique topologiques.

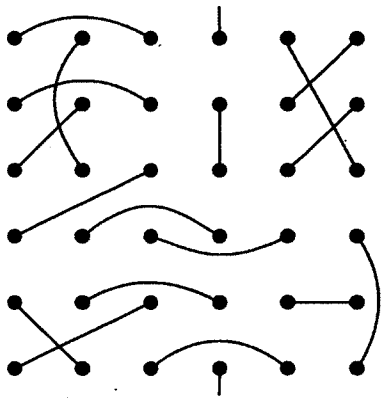


FIGURE 5.1 État VBS avec $\ell = \sqrt{5}$ sur un réseau $L = 6$ avec conditions périodiques.

exhibe de l'ordre topologique, en nous basant essentiellement sur l'article [85] et en nous inspirant de [86].

5.2.1.1 États à liaisons de valence

Un état à liaison de valence (*valence bond solid* ou VBS) est un état où chaque électron d'un réseau appartient à une liaison de valence, *i.e.*, appartient à une paire d'électrons dans un état singulet $|\psi^-\rangle$. Généralement, on restreint la portée des liaisons de valence en imposant une portée maximale ℓ. Un tel état est représenté graphiquement par un réseau de points où chaque liaison de valence est représentée par une arête (représentées en bleu sur la figure 5.1). L'ensemble des états à liaison de valence ne constituent ni une famille linéairement indépendante ni une famille orthogonale. Toutefois, cette base surcomplète est un outil intuitif pour décomposer les états. Plutôt que de considérer un état VBS, on peut considérer une superposition d'entre eux. Si la superposition ne favorise aucun des états VBS, *i.e.*, si toutes les amplitudes sont les mêmes, on obtient un état RVB, proposé en 1973 par Anderson [87] comme un candidat comme état fondamental d'un liquide de spin.

5.2.1.2 Invariants topologiques

En raison des contraintes géométriques, c.-à-d. le fait que les liaisons connectent deux sites et que chaque site appartient à exactement une liaison, les états RVB présentent des propriétés topologiques *qui ne peuvent pas être modifiées par une opération locale.* Ces propriétés peuvent être décrites par un invariant topologique appelé *parité d'intervalle* [88]. Considérons une réseau $L \times L$ avec conditions aux frontières périodiques, autrement dit un *tore.* Pour simplifier, supposons que

L soit pair. Soit un état VBS à courte portée sur ce réseau, par exemple celui de la figure 5.2. Traçons une ligne verticale entre la première colonne et la seconde colonne de points. Cette ligne

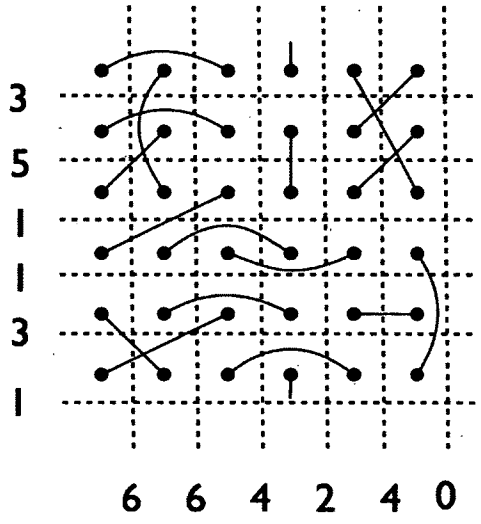


Figure 5.2 Calcul de la parité d'intervalle d'un état VBS.
Pour chaque ligne verticale, le nombre de liens coupés est pair et pour chaque ligne horizontale, le nombre de liens coupés est impair.

coupe six liaisons de valence. On peut alors recommencer cette procédure en traçant une ligne verticale ailleurs dans le réseau. On remarque que le nombre de liaisons coupées sera toujours *pair*. Le même raisonnement peut être appliqué avec des lignes horizontales et on constate que le nombre de liaisons coupées est toujours impair. On dira alors que la parité d'intervalle est paire verticalement et impaire horizontalement.

Ces corrélations sont déjà remarquables, mais elles sont de plus invariantes si on effectue un changement *local* dans la configuration des liaisons. Un tel changement local, représenté sur la figure 5.3, correspond à deux sites proches qui échangent leur extrémité d'une liaison de valence. Comme montré sur la figure 5.3, la parité d'intervalle est conservée. Plus généralement, tout changement local préserve la parité d'intervalle : il s'agit d'un invariant topologique.

Afin d'obtenir une intuition de ce résultat, rappelons-nous qu'un réseau carré avec conditions périodiques correspond à un tore. Les courbes fermées non-triviales sur un tore sont celles qui font le tour d'un des deux axes du tore : ce sont les seules qui ne peuvent pas être déformées jusqu'à devenir un point. Or, on peut remarquer que sur la figure 5.3, que la courbe obtenue par l'union des anciennes liaisons et des nouvelles liaisons est une courbe fermée triviale. Généralement, toute

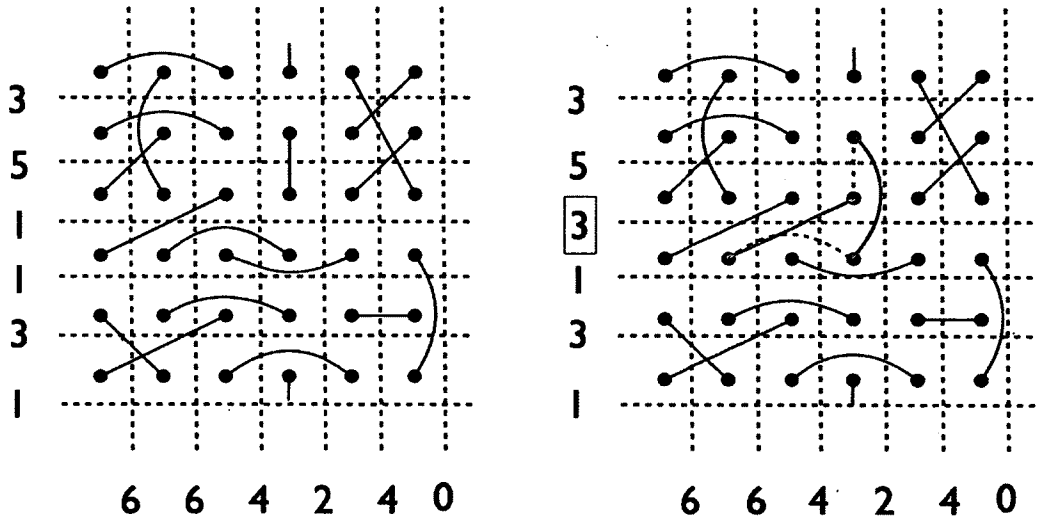


FIGURE 5.3 Invariance de la parité d'intervalle d'un état VBS sous changement local.
Les deux nouvelles liaisons (en rouge) sont obtenus à partir de deux anciennes liaisons (en pointillé) suite à une transformation locale. Considérons la droite horizontale qui coupent les deux nouvelles liaisons. Son nombre de liaisons coupées, inscrit dans le carré rouge, est passé de 1 à 3, cf. Fig. 5.2. Il est donc demeuré impair et la parité d'intervalle est inchangée.

modification locale correspond à une courbe triviale et ne change pas la parité d'intervalle. Au contraire, un grand nombre de changements locaux correspondant à une courbe fermée non-triviale peut changer la parité d'intervalle, comme représenté sur la figure 5.4 . Or, une telle transformations demandent un nombre macroscopique de changements locaux.

On définira une classe topologique comme l'ensemble des états VBS pouvant être obtenus grâce à un petit nombre de changements locaux, petit devant la dimension du tore. Or, nous venons de voir que la parité d'intervalle caractérise les différentes classes topologiques. Il existe donc quatre classes topologiquement distinctes représentées sur la figure 5.5, chacune caractérisée par la parité des liaisons dans les intervalle horizontaux et la parité des liaisons dans les intervalles verticaux. Un état RVB particulièrement intéressant est la superposition équi-amplitude de tous les états VBS appartenant à la même classe topologique.

5.2.1.3 Transition de phase et paramètre d'ordre

Il est possible d'écrire un hamiltonien dont l'espace fondamental est généré par les états RVB correspondant à chacun des secteur topologiques [84]. Ainsi, sur un géométrie torique, l'espace

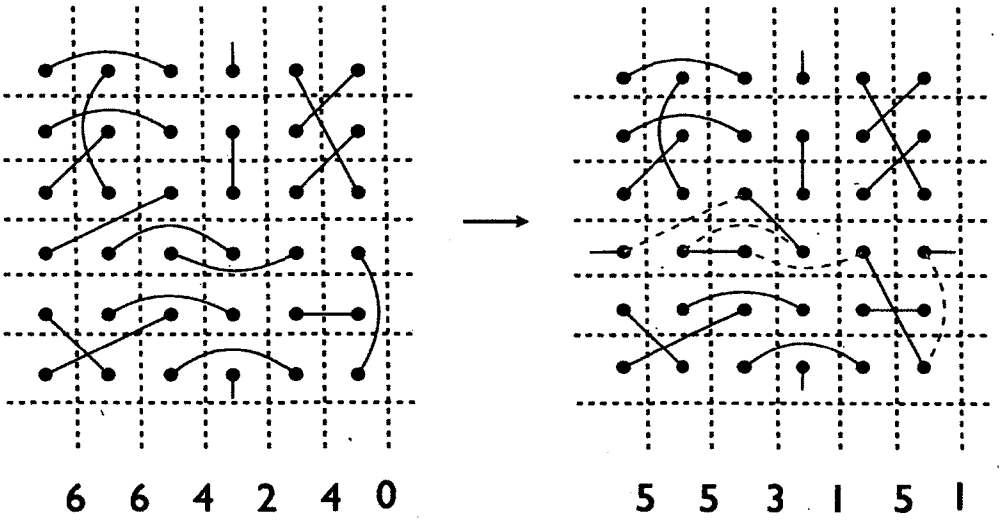


FIGURE 5.4 Changement de la parité d'intervalle d'un état VBS par un nombre macroscopique de changements locaux. La courbe fermée obtenue par union des anciennes liaisons (pointillées) et des nouvelles liaisons (pleines) fait le tour du tore et est donc non-triviale. La parité d'intervalle verticale est passée de paire à impaire, cf. Fig. 5.2.

fondamental est engendré par les quatre états correspondants à chacun des secteurs topologiques. Cet espace fondamental sera topologiquement ordonné, ce que nous allons définir rigoureusement dans la section suivante. En particulier, cet espace fondamental est prometteur afin d'encoder de l'information quantique.

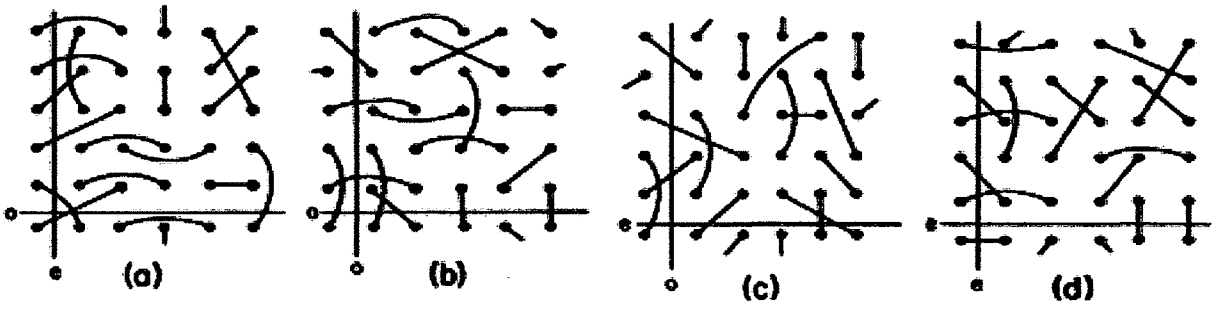


FIGURE 5.5 États VBS avec $\ell = \sqrt{5}$ sur un réseau $L = 6$. Chacun appartient à un secteur topologique différent. La classe topologique peut être déterminée par les parité d'intervalle (o=impair, e=pair) verticale et horizontale, appellée parité d'intervalle. Figure tirée de [85].

5.2.2 Définition(s) de l'ordre topologique

5.2.2.1 Indiscernabilité des états par des opérateurs locaux

Maintenant que nous avons vu un exemple d'ordre topologique, il convient de donner une définition formelle de cette notion. L'ordre topologique est la propriété d'un sous-espace vectoriel, typiquement l'espace fondamental d'un hamiltonien local dont les premières excitations sont séparés du fondamental par un gap d'énergie. Intuitivement, l'idée est qu'il existe des états orthogonaux qui ne peuvent pas être distingués localement et qu'aucune transformation locale ne permet de passer de l'un à l'autre. Formellement, on dira qu'un sous-espace est topologiquement ordonné s'il respecte la condition TQO (*topological quantum order*) donnée ci-dessous

Définition 3 (TQO). Un sous-espace $S \subset \mathcal{H}$ de dimension au moins 2, défini par son projecteur P, est topologiquement ordonné si l'action de tout opérateur local[4] O_{loc} dans le sous-espace est triviale, *i.e.*,

$$\forall O_{\text{loc}} \, \exists c \in \mathbb{C} \quad P O_{\text{loc}} P = cP \tag{5.5}$$

Cette définition formelle capture l'intuition énoncée plus haut. Soient deux états orthogonaux $|\psi\rangle$ et $|\psi^\perp\rangle$dans S, *i.e.*, tels que $\langle\psi|\psi^\perp\rangle = 0$. L'équation (5.5) projetée sur $|\psi\rangle\langle\psi^\perp|$ d'une part et $|\psi\rangle\langle\psi|$ d'autre part devient

$$\langle\psi^\perp|O_{\text{loc}}|\psi\rangle = c\langle\psi^\perp|\psi\rangle = 0 \tag{5.6}$$

$$\langle\psi|O_{\text{loc}}|\psi\rangle = \langle\psi^\perp|O_{\text{loc}}|\psi^\perp\rangle = c \tag{5.7}$$

L'équation (5.6) montre qu'il est impossible de passer de $|\psi\rangle$ à $|\psi^\perp\rangle$ à l'aide d'une transformation locale alors que l'éq. (5.7) montre que la valeur moyenne d'une observable locale est la même pour tout état dans le sous-espace S. Ceci élimine donc la possibilité qu'il existe un paramètre d'ordre local qui distingue les états fondamentaux.

En terme de correction d'erreur, la condition (5.5) n'est autre que la condition de Knill-Laflamme [89] pour l'ensemble des opérateurs locaux. Ainsi, aucune erreur locale ne peut corrompre l'information quantique encodée dans le sous-espace S. Ainsi, on espère pouvoir encoder de l'information quantique de façon robuste dans des degrés de liberté topologiques parce que l'action

4. La définition d'ordre topologique fait apparaître une notion de localité, typiquement définie par rapport à la géométrie du réseau.

locale de l'environnement ne pourra pas affecter cette information. L'espoir de faire une mémoire auto-correctrice à partir d'un système topologique réside en grande partie dans cette propriété. Nous reviendrons en détail sur les mémoires auto-correctrices dans le chapitre 6.

Dans les exemples habituels, l'ordre topologique se manifeste lorsqu'un hamiltonien local gappé présente un espace fondamental dégénéré qui obéit à la condition (5.5). Notons que la dégénérescence n'apparaît que lorsque le hamiltonien est défini sur une géométrie topologiquement non-triviale, p.ex. un tore.

Toutefois, la connexion entre ordre topologique et la structure du hamiltonien n'est pas claire. En effet, il est aussi possible d'argumenter que l'ordre topologique est la propriété d'un état, plutôt que du hamiltonien [90]. Dans ce cas, un état est dit topologiquement ordonné s'il existe un sous-espace topologiquement ordonné de dimension au moins 2 dont il fait partie. En particulier, on pensait que, étant donné un état topologique, la construction d'un hamiltonien local mènerait à un hamiltonien local gappé dont l'espace fondamental serait précisément le sous-espace topologique auquel il appartient. Or, cette intuition s'est révélée fausse : il est possible de construire un hamiltonien local sans gap dont le fondamental contient l'espace topologique, mais présente aussi un continuum d'états excités [91]. Ainsi, s'il est clair qu'un hamiltonien gappé dont l'espace fondamental obéit à la condition (5.5) suffit pour avoir de l'ordre topologique, les conditions nécessaires à l'ordre topologique sont encore mal comprises.

5.2.2.2 États à intrication courte portée

Une autre définition possible de l'ordre topologique est de définir les états topologiquement triviaux comme étant les états à intrication à courte portée. Formellement, un état à intrication de courte portée est obtenu à partir d'un état produit par une transformation unitaire locale. Une transformation unitaire locale (LU) est obtenue soit par évolution unitaire sous un hamiltonien local $U = \mathcal{T}e^{-i\int_0^1 H(g)dg}$ durant un temps constant, ou, de façon équivalente, par un circuit quantique de profondeur finie [5]. L'idée est alors de définir une relation d'équivalence pour laquelle deux états sont équivalents si une transformation LU transforme l'un en l'autre. Les états à intrication courte portée forment alors la classe d'équivalence des états produits, qui ne sont pas topologiquement ordonnés [90]. Au contraire, les états topologiquement ordonnés présentent de l'intrication longue portée. Un

5. Pour définir la profondeur d'un circuit quantique, on réécrit le circuit comme une séquence de produit tensoriel de portes unitaires locales : le nombre d'étages dans cette séquence est la profondeur du circuit. Ainsi, un MERA (cf. section 4.5) a une profondeur $\log n$.

sous-espace est topologiquement ordonné s'il contient des états topologiques appartenant à des classes d'équivalence différentes. Ces classes seront alors appelés des « secteurs topologiques ».

On croit que la présence d'intrication à longue portée dans les états topologiques est la raison pour laquelle la classification de Landau ne parvient à décrire ce type d'ordre [82]. Notons que la notion d'intrication longue portée peut être très différente de celle de corrélations longue portée. Ainsi, nous verrons que les états du code torique, introduit en 5.3, ont de l'intrication longue portée alors que leur longueur de corrélations, au sens des corrélations à deux corps, est strictement nulle.

5.2.2.3 Signatures de l'ordre topologique

Les définitions de l'ordre topologique données jusqu'à présent permettent difficilement de certifier qu'un état présente de l'ordre topologique, sauf pour des modèles simples comme le code torique. En effet, il faut soit montrer que pour toute observable locale la propriété (5.5) est satisfaite soit que tout circuit de profondeur finie ne transforme pas l'état considéré en un état produit. Ces définitions sont utiles pour prouver des résultats mathématiques ou montrer qu'un état particulier n'est pas topologiquement ordonné. Au contraire, on aimerait trouver des *signatures* de l'ordre topologique, *i.e.* des propriétés qui garantissent la présence d'ordre topologique.

Variation de la dégénérescence avec la topologie Une première signature, qui justifie l'appelation « ordre topologique » est la variation de la dégénérescence du fondamental avec la topologie de l'espace où agit le hamiltonien. Par exemple, nous verrons en 5.3.3.3 que le code torique a une dégénérescence qui varie comme 2^{2g} lorsque définit sur une surface de genre g. Le genre d'une surface correspond intuitivement au nombre de « trous » dans cette surface. Ainsi, une sphère est de genre 0 alors qu'un tore est de genre 1. En terme physique, un tore est obtenu en imposant des conditions périodiques aux frontières verticales et horizontales d'un plan 2D. Ainsi, le code torique définit sur une sphère a un fondamental unique alors qu'il est de dégénérescence 4 sur un tore.

Modèle anyonique Une autre signature de l'ordre topologique se manifeste dans les propriétés des excitations de basse énergie d'un hamiltonien gappé topologique. En effet, ces excitations se comportent typiquement comme des particules exotiques, appelés anyons [92]. Des anyons (abéliens) sont des quasi-particules dont la fonction d'onde à deux particules $|\psi_1, \psi_2\rangle$ prend une phase quelconque lorsque ces deux particules sont échangées , *i.e.*, $|\psi_2\psi_1\rangle = S|\psi_1\psi_2\rangle = e^{i\theta}|\psi_1\psi_2\rangle$.

Notons que ce comportement n'apparaît qu'en 2D : en dimension supérieure, on ne peut avoir que $\theta = 0$ (bosons) ou $\theta = \pi$ (fermions) car $S^2 = \mathbb{I}$.

Une propriété importante des anyons est leur règle de fusion [93]. Ces règles régissent l'apparition de nouveaux anyons lorsque deux anyons sont combinés. Intuitivement, ce phénomène ressemble à la combinaison des moments cinétiques de deux particules. Ainsi, deux particules de spin-1 donnent naissance à trois sous-espaces dont l'un est de spin-0, un autre de spin-1 et un autre de spin-2

$$1 \otimes 1 = 0 \oplus 1 \oplus 2 \tag{5.8}$$

En termes d'anyons, on dirait alors que la fusion de deux anyons de type 1 peut donner soit un anyon de type 0, soit un anyon de type 1, soit un anyon de type 2. Les anyons dont les règles de fusion ne sont pas déterministes sont appelés des anyons non-abéliens. Pour de tels anyons, l'échange de particules peut changer le type d'anyons. On a donc généralement

$$\psi_a(x_2, x_1) = \sum_b B_{ab} \psi_b(x_1, x_2) \tag{5.9}$$

où a, b indicent les types d'anyons (ou charges topologiques). B_{ab} est diagonale avec des valeurs propres $e^{i\theta}$ ($\theta \notin \mathbb{Z}\pi$) pour des anyons abéliens et non-diagonale pour des anyons non-abéliens. Dans le cas des anyons non-abéliens, l'amplitude de probabilité va donc changer suivant l'ordre dans lequel les particules sont échangées. Ainsi, le tressage d'anyons non-abéliens permet d'obtenir des portes quantiques et de réaliser un calcul quantique dit « topologique ».

Entropie topologique Finalement, une quantité calculable directement à partir d'un état propose de mesurer la présence (ou l'absence d'ordre topologique). Il s'agit de l'entropie topologique, proposée indépendamment par Kitaev & Preskill [94] et Levin & Wen en 2006 [95]. Dans la version la plus simple, il s'agit d'un terme de correction à la loi d'aire pour l'entropie (cf. section 4.1.3.2), *i.e.*,

$$S(\rho_A) = \alpha |\partial A| - \gamma + o(1) \tag{5.10}$$

où $o(1)$ est une correction qui tend vers 0 dans la limite $L \to \infty$. On peut montrer que l'entropie topologique γ est strictement positive que pour des états topologiques. De plus, elle est (i) un invariant topologique, *i.e.*, elle ne change pas si on déforme continûment la frontière de la région A et (ii) universelle, *i.e.*, elle ne change pas pour les hamiltoniens dans la même phase (pour lesquels il est possible de passer de l'un à l'autre continûment sans fermer le gap).

5.3 Modèle canonique : le code torique

L'ordre topologique est un sujet de recherche actif dont plusieurs aspects sont encore mal compris, p. ex. les conditions nécessaires pour avoir de l'ordre topologique. Afin de fixer les idées, il est utile d'avoir un modèle-type en tête [6]. Par exemple, le modèle canonique pour comprendre la notion de symétrie brisée est le modèle d'Ising avec champ transverse. Pour l'ordre topologique, le modèle canonique est celui du code torique, introduit par Kitaev en 1997 [80]. Nous allons donc présenter en détails ce modèle afin de développer une intuition de l'ordre topologique. De plus, le code torique sera utile afin de discuter des mémoires auto-correctrices au chapitre 6 et ses propriétés fournissent l'intuition de l'instabilité thermique qui afflige les mémoires topologiques, comme nous le verrons au chapitre 7. Le code torique est donc le fil conducteur afin de bien comprendre les notions développées dans les prochains chapitres.

Le code torique est un cas particulier d'un code stabilisateur, très utilisé en informatique quantique, que nous introduirons en 5.3.1. Ceci nous permettra de définir simplement le code torique en 5.3.2 avant d'explorer ses propriétés topologiques en 5.3.3.

5.3.1 Code stabilisateur

Un code de correction d'erreur quantique C (ou code correcteur) est un sous-espace vectoriel de l'espace de Hilbert où il est possible de stocker de l'information quantique. Ainsi, il est possible de corriger des erreurs, typiquement locales. Une classe très générale de codes correcteurs est celle des codes stabilisateurs [96]. Définie grâce à la structure du groupe de Pauli, elle regroupe les exemples historiques comme le code de Shor [97] ainsi que des codes plus récents, comme le code cubique [98].

5.3.1.1 Définition

La définition d'un code stabilisateur est intimement reliée à celle d'un état stabilisateur (cf. section 2.3.1.2) à la différence qu'un code n'est pas un état, mais un sous-espace vectoriel d'un espace de Hilbert de n qubits. Ainsi, il est aussi défini par un sous-groupe abélien $\mathcal{S}$ du groupe de Pauli qui ne contient pas $-\mathbb{I}$, mais ce sous-groupe n'est généré que par $\ell \leq n$ générateurs.

6. Il faut toutefois prendre garde à la tentation de prendre toutes les particularités d'un modèle-type pour des propriétés génériques.

Ainsi, alors que pour un état stabilisateur, on a $\ell = n$, ce qui fournit assez de contraintes pour spécifier un état unique, pour un code stabilisateur on aura $\ell \leq n$ afin de définir un sous-espace. Le code est le sous-espace propre commun associé à la valeur propre $+1$ de tous ces opérateurs qui commutent, *i.e.*,

$$\mathcal{C} = \{|\psi\rangle \quad \forall S \in \mathcal{S}\ S|\psi\rangle = +|\psi\rangle\}$$

Celui-ci sera donc de dimension $2^{n-\ell}$ et correspondra donc à $k = n - \ell$ qubits effectifs. Ainsi, à partir de n qubits physiques, on a encodé k qubits logiques.

5.3.1.2 Exemple : modèle d'Ising ferromagnétique sans champ

L'espace fondamental Vec $\{|\uparrow\rangle^{\otimes n}, |\downarrow\rangle^{\otimes n}\}$ du modèle d'Ising ferromagnétique sans champ, défini par son hamiltonien $H = -\sum_i \sigma_i^z \sigma_{i+1}^z$, peut être vu comme un code stabilisateur. En effet, l'espace fondamental est le sous-espace propre $+1$ des opérateurs de Pauli $Z_i Z_{i+1}$ qui génèrent bien un sous-groupe abélien ne contenant pas l'opérateur $-\mathbb{I}$.

Cet exemple permet aussi de mettre en lumière la connexion entre code stabilisateur et hamiltonien. En effet, il est toujours possible de voir un code stabilisateur $\mathcal{S} = \langle g_1, \ldots, g_\ell \rangle$ comme l'espace fondamental du hamiltonien non-frustré dont les termes commutent $H = -\sum_{i=1}^{\ell} g_i$.

Vérifions que l'on retrouve bien la dégénérescence du fondamental en comptant le nombre de générateurs indépendants. Pour une chaîne de n qubits, *i.e.* des conditions aux frontières ouvertes, on a $\ell = n - 1$ générateurs et on retrouve bien une dégénéresence 2. Pour n qubits disposés sur un cercle (donc avec conditions périodiques), il semble y avoir n opérateurs indépendants, mais il faut tenir compte de la contrainte globale $Z_1 Z_n = \prod_{i=1}^{n-1} Z_i Z_{i+1}$. Ainsi, il y a bien $\ell = n - 1$ générateurs indépendants.

Du point de vue informatique, l'espace fondamental du modèle d'Ising ferromagnétique correspond à un code de répétition. L'idée est d'encoder chaque bit classique 0 ou 1 dans un plus grand nombre n de bits. Afin de corriger les erreurs classiques qui sont des renversements de bits, correspondant à l'opérateur X, il suffit de regarder la majorité des bits classiques. Par exemple, il suffit d'encoder 1 bit logique dans 3 bits physiques pour corriger une erreur sur un bit physique. Ainsi, l'encodage se résume à $0 \mapsto 0001 \mapsto 111$ et la correction d'erreur, par exemple pour le renversement du 2e bit physique, à $010 \mapsto 000101 \mapsto 111$.

5.3.1.3 Correction d'erreur dans un code stabilisateur

Un code stabilisateur utilise n qubits physiques pour obtenir k qubits logiques. Le but de l'opération est de protéger l'information encodée dans le code face aux erreurs introduites au niveau physique par l'environnement. En supposant que l'environnement aie une action locale, il suffit de pouvoir gérer les erreurs locales, *i.e.* celles n'agissant que sur un petit nombre de qubits physiques.

Une erreur locale est formellement une transformation CPTP $\mathcal{E}$ qui n'agit non-trivialement que sur quelques qubits. Rappelons qu'une transformation CPTP est la transformation la plus générale sur l'espace des matrices densité, cf. 3.2.4.1. Grâce à la décomposition de Kraus $\mathcal{E}(\rho) = \sum_k E_k^\dagger \rho E_k$, il suffit donc de pouvoir corriger des opérateurs d'erreur $\{E_k\}_k$. De plus, puisque les opérateurs de Pauli forment une base d'opérateurs, il suffit de pouvoir corriger les opérateurs de Pauli de poids[7] faible.

Pour ce faire, il faut tout d'abord identifier ou « diagnostiquer » l'erreur. Pour ce faire, il suffit de mesurer chaque générateur du code stabilisateur afin d'obtenir une chaîne de valeurs propres ± 1 qui constitue le syndrome de l'erreur[8]. À partir de ce syndrome, il est possible d'identifier quelle serait l'opération à effectuer sur les qubits physiques afin de renverser l'erreur.

Par exemple, pour le code à répétition, le tableau de syndrome est donné dans le tableau 5.1.

Erreur	Syndrome		Correction
	Z_1Z_2	Z_2Z_3	
$\mathbb{I}$	+1	+1	$\mathbb{I}$
X_1	-1	+1	X_1
X_2	-1	-1	X_2
X_3	+1	-1	X_3

Table 5.1 Tableau de syndrome du code à répétition

5.3.1.4 Opérateur logique

Certains opérateurs L agissent à l'intérieur du code, *i.e.*, $\forall|\psi\rangle \in \mathcal{C},\ L|\psi\rangle \in \mathcal{C}$. Ainsi, ils effectuent des transformations sur les qubits logiques et ne les sortent pas du code. En particulier,

7. Le poids d'un opérateur de Pauli est le nombre de qubits sur lequel il agit non-trivialement.
8. En fait, cette étape a aussi l'effet crucial de projeter une erreur quelconque vers un nombre discret d'erreurs.

ils ne créent pas d'erreurs et ne sont pas détectables par les stabilisateurs du code. On qualifie de tels opérateurs de « logiques ». Formellement, il s'agit des opérateurs L qui commutent avec tous les stabilisateurs du code.

Définition 4 (Opérateur logique). L est un opérateur logique d'un code stabilisateur $\mathcal{S}$ si

$$\forall S \in \mathcal{S} \ [L, S] = 0 \tag{5.11}$$

En particulier, tous les stabilisateurs sont des opérateurs logiques qui agissent trivialement sur le code. On peut raffiner la définition afin d'obtenir les opérateurs logiques non-triviaux [9].

Définition 5 (Opérateur logique non-trivial). Un opérateur logique non-trivial L est un opérateur logique qui n'appartient pas au groupe stabilisateur.

Les opérateurs logiques forment un groupe, le groupe logique. Il sera donc utile de déterminer les générateurs du groupe logique d'un code stabilisateur.

Par exemple, pour le code d'Ising, un opérateur logique est $\bigotimes_{k=1}^{n} X_k$ qui échange $|\bar{0}\rangle = |\uparrow\rangle^{\otimes n}$ avec $|\bar{1}\rangle = |\downarrow\rangle^{\otimes n}$. Cet opérateur logique n'est autre que l'opérateur de Pauli X du qubit logique, et on le note $\bar{X} \equiv \bigotimes_{k=1}^{n} X_k$. En fait, $\bar{X}$ est le représentant canonique de la classe d'opérateurs $\bar{X}\mathcal{S}$ qui ont tous le même effet sur le code. On peut alors se demander quel serait l'opérateur $\bar{Z}$ qui déphase l'état logique $|\bar{1}\rangle$. Il s'agit tout simplement de la classe d'équivalence de n'importe quel Z_k. En particulier, il s'agit du paramètre d'ordre qui permet de détecter la transition de la phase désordonnée $\bigotimes_k |\rightarrow\rangle_k$ vers la phase ordonnée Vec $\{|\uparrow\rangle^{\otimes n}, |\downarrow\rangle^{\otimes n}\}$.

Du point de vue de la correction d'erreur, cela veut dire qu'une erreur Z sur un seul qubit ne peut pas être corrigé par le code d'Ising. Ainsi, il est facile pour l'environnement de détruire une superposition de $|\uparrow\rangle^{\otimes n}$ et de $|\downarrow\rangle^{\otimes n}$. En effet, une superposition $\alpha|\bar{0}\rangle + \beta|\bar{1}\rangle$ deviendra rapidement un mélange statistique $\rho = |\alpha|^2 |\bar{0}\rangle\langle\bar{0}| + |\beta|^2 |\bar{1}\rangle\langle\bar{1}|$. Ainsi, le code d'Ising, autrement dit le code à répétition, est un bon code classique mais un très mauvais code quantique car il ne protège pas la phase entre $|\bar{0}\rangle$ et $|\bar{1}\rangle$. Le premier code quantique découvert, celui de Shor [97], combine en fait un code à répétition pour les renversements de spin X avec un code à répétition pour les déphasages de spin Z, pour obtenir un bon code quantique.

9. Par abus de langage, on désigne souvent un opérateur logique non-trivial comme un opérateur logique tout court.

Nous allons maintenant nous intéresser à un code stabilisateur particulier, le code torique, dont nous verrons qu'il est ordonné topologiquement.

5.3.2 Définition du code torique

Le code torique est un code stabilisateur qui encode $k = 2$ qubits logiques dans n qubits physiques. Physiquement, il s'agit de l'espace fondamental d'un hamiltonien 2D agissant sur des qubits (ou spin-1/2) placés sur les arêtes d'un réseau carré 2D.

5.3.2.1 Générateurs

On peut définir le groupe stabilisateur du code torique à l'aide de deux familles d'opérateurs, représentés sur la Fig. 5.6.

- Les opérateurs étoiles A_s, dont un est représenté en rouge sur la Fig. 5.6, agissent sur les 4 qubits adjacents à un site s du réseau. Il s'agit d'une interaction à 4 corps défini par

$$A_s = \bigotimes_{i \in \mathcal{N}(s)} X_i \tag{5.12}$$

  Remarquons que le support géométrique de ces opérateurs est une plaquette du réseau réciproque.
- Les opérateurs plaquettes B_p, dont un est représenté en bleu sur la Fig. 5.6 agissent sur les 4 qubits appartenant à une plaquette p, *i.e.* ils sont adjacents à un site du réseau réciproque. Il s'agit d'une interaction à 4 corps défini par

$$B_p = \bigotimes_{i \in p} Z_i \tag{5.13}$$

5.3.2.2 Stabilisateurs

Par définition, les stabilisateurs sont des produits d'opérateurs étoiles et d'opérateurs plaquettes. Nous allons voir que ceux-ci ont une interprétation géométrique simple.

Par exemple, prenons le produit de deux opérateurs plaquettes pour des plaquettes adjacentes : le produit des opérateurs Z provenant de chaque générateur sur le qubit partagé se simplifie à l'identité $\mathbb{I}$ et on obtient un opérateur « rectangle » agissant sur six qubits, représenté sur la Fig. 5.7.

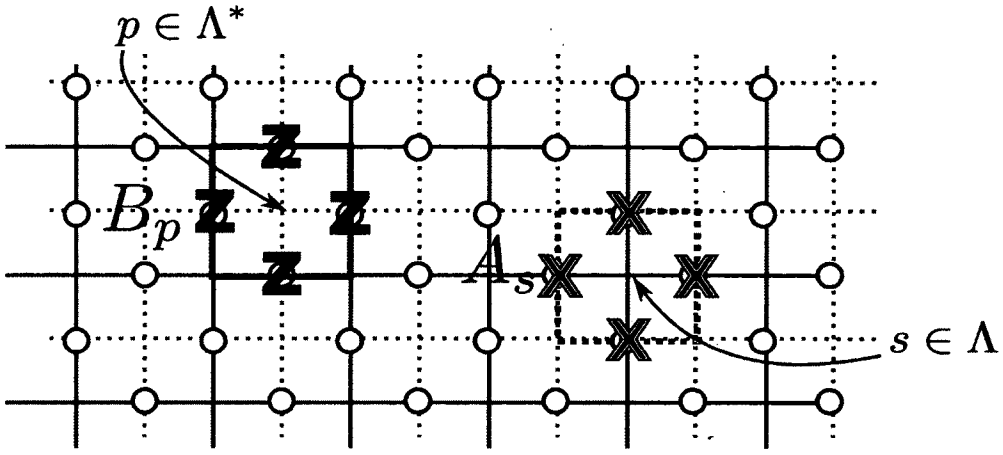


FIGURE 5.6 Définition du code torique.
Deux des générateurs du code, un opérateurs étoile A_s (en rouge) et un opérateur plaquette B_p (en bleu), sont représentés.

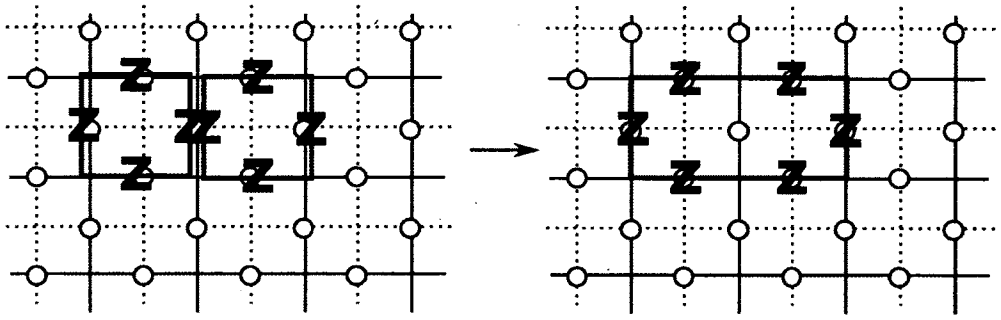


FIGURE 5.7 Le produit de deux opérateurs plaquettes est un opérateur à six corps sur un rectangle.

Plus généralement, en prenant le produit d'opérateurs plaquettes, on obtient des opérateurs boucles $Z_{\mathcal{B}} \equiv \bigotimes_{s\in\mathcal{B}} Z_s$ dont le support géométrique est une boucle fermée sur le réseau. De même, le produit d'opérateurs étoiles génèrent des opérateurs boucles $X_{\mathcal{B}^*} \equiv \bigotimes_{s\in\mathcal{B}^*} X_s$ dont le support géométrique est une boucle fermée sur le réseau réciproque. Deux exemples de tels opérateurs sont représentés sur la Fig. 5.8.

Ainsi, la multiplication d'un stabilisateur par un autre stabilisateur revient à unir leurs supports géométriques « modulo 2 », *i.e.* en retirant les sites qui apparaissent deux fois dans la partie X et la partie Z. Ainsi, le support des stabilisateurs sont toutes les boucles triviales, *i.e.*, celles qui sont la frontière d'une région qui peut se contracter à un point. Nous verrons que les boucles non-triviales correspondent aux opérateurs logiques.

De façon générale, un stabilisateur est donc un produit d'opérateurs boucles, *i.e.*, son support

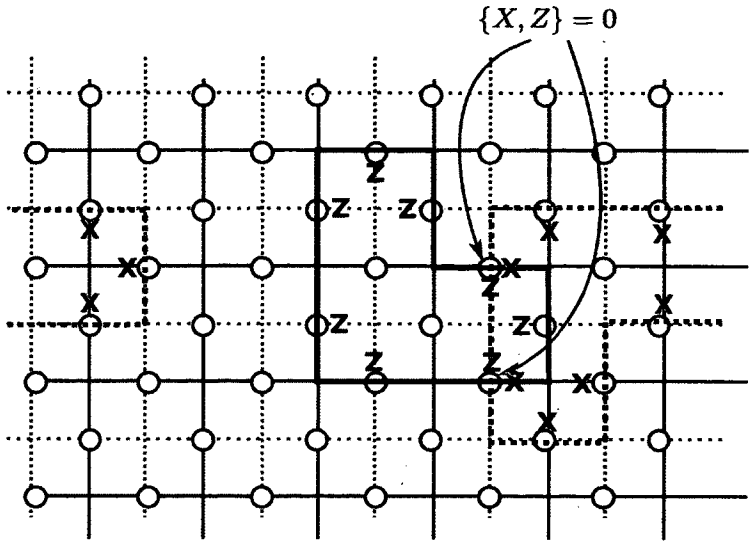


FIGURE 5.8 Stabilisateurs du code torique.
Deux stabilisateurs sont représentés : un opérateur boucle X_{B^*} (en rouge) et un opérateur boucle Z_B (en bleu). Ils commutent grâce à deux anti-commutations sur des qubits distincts.

géométrique est une réunion de boucles triviales.

5.3.2.3 Commutation

Afin de vérifier que les stabilisateurs commutent, il suffit de vérifier que les générateurs commutent. Remarquons tout d'abord que les opérateurs étoiles commutent trivialement entre eux. De même, les opérateurs plaquettes commutent deux à deux. Par ailleurs, deux opérateurs dont les supports géométriques n'ont pas d'intersection commutent trivialement.

Le cas intéressant est donc celui d'un opérateur étoile et d'un opérateur plaquette dont les supports se touchent, comme sur la figure 5.9. Cela ne se produit que si le site s est un coin de la plaquette p. Dans ce cas, le support commun est forcément formé de deux qubits. Or, les opérateurs X et Z anti-commutent. Il y aura donc deux anti-commutations, une pour chacun des qubits communs. Ainsi, les opérateurs A_s et B_p commutent.

Géométriquement, la commutation des stabilisateurs correspond au fait qu'un opérateur X_{B^*} et un opérateur Z_B anti-commutent sur les qubits où leurs supports se coupent. Or, les boucles triviales du réseau direct et du réseau réciproque se coupent forcément sur un nombre pair de sites, comme par exemple sur la figure 5.8.

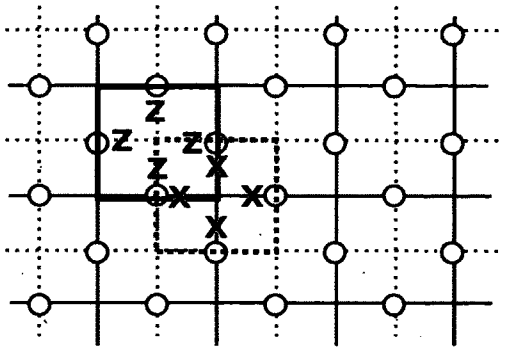


FIGURE 5.9 Commutation d'un opérateur plaquette et d'un opérateur étoile dont les supports ont une intersection non nulle.

5.3.2.4 Code torique

Le code torique est donc le code stabilisateur associé aux opérateur étoiles et aux opérateurs plaquettes, *i.e.*, l'ensemble des états $|\psi\rangle$ tels que

$$\forall s \; A_s|\psi\rangle = +|\psi\rangle \qquad\qquad \forall p \; B_p|\psi\rangle = +|\psi\rangle \tag{5.14}$$

ou encore l'espace fondamental du hamiltonien

$$H = -\sum_s A_s - \sum_p B_p \tag{5.15}$$

qui est local, gappé et sans-frustration et dont nous allons voir qu'il admet un espace fondamental dégénéré.

5.3.2.5 Dégénérescence

Le code torique est défini sur un réseau carré avec conditions périodiques, autrement dit un tore. Afin de calculer la dimension du code, il suffit de compter le nombre de générateurs indépendants. Pour ce faire, il faut compter le nombre de sites s et le nombre de plaquettes p. Il est possible d'associer deux qubits à chaque site. Ainsi, il a $n/2$ sites[10]. De même, il y a $n/2$ plaquettes. On obtient ainsi n générateurs... mais qui ne sont pas tous indépendants !

En effet, les conditions périodiques ajoutent des contraintes sur ces opérateurs. Considérons le produit de tous les opérateurs étoiles $\prod_{s\in\Lambda} A_s$. Chaque qubit est affecté par deux opérateur étoile,

10. Le nombre n de qubits est toujours pair par construction.

l'un à sa gauche, l'autre à sa droite. Ainsi, la contribution sur chaque qubit est $X \times X = \mathbb{I}$. On en déduit que le produit des opérateurs étoiles agit trivialement sur tous les qubits, *i.e.*,

$$\prod_{s \in \Lambda} A_s = \mathbb{I}. \tag{5.16}$$

De même, le produit de tous les opérateurs plaquettes est trivial

$$\prod_{p \in \Lambda^*} B_p = \mathbb{I}. \tag{5.17}$$

On en déduit qu'il n'y a que $\ell = n - 2$ générateurs indépendants et que la dimension du code est donc 4. Il permet donc d'encoder $k = 2$ qubits logiques.

5.3.2.6 Opérateurs logiques

Les opérateurs logiques du code torique sont complètement délocalisés, contrairement au code d'Ising où un d'entre eux n'était que l'aimantation Z_k d'un seul spin. Afin de déterminer les opérateurs logiques, il faut trouver des opérateurs qui commutent avec tous les stabilisateurs mais qui n'en sont pas. Le plus simple est d'exhiber 4 générateurs du groupe logique. On choisit alors les 4 opérateurs représentés sur la figure 5.10

Considérons par exemple $\bar{X}_1$. Il commute trivialement avec tous les opérateurs étoiles et a deux qubits en commun avec tout opérateur plaquette dont il intersecte le support, ce qui implique qu'il commute avec les opérateurs plaquette. De plus, il n'est pas dans le groupe stabilisateur car son support n'est pas une boucle triviale. De même, les opérateurs $\bar{X}_2$, $\bar{Z}_1$ et $\bar{Z}_2$ sont des opérateurs logiques. Comment s'assurer qu'ils sont indépendants ? Il suffit de s'intéresser aux relations d'anti-commutation par ces opérateurs. Remarquons que $\bar{X}_1$ et $\bar{Z}_1$ ne se croisent qu'au niveau d'un seul qubit et anti-commutent donc. De même, $\bar{X}_2\bar{Z}_2 = -\bar{Z}_2\bar{X}_2$. De plus, les opérateurs $\bar{X}_1$ et $\bar{X}_2$ (resp. $\bar{Z}_1$ et $\bar{Z}_2$) commutent car leurs supports ont une intersection vide. Ainsi, ces opérateurs ont les mêmes relations d'anti-commutation que les matrices de Pauli. En fait, il s'agit des matrices de Pauli des qubits logiques. Ces 4 opérateurs logiques génèrent donc l'espace des opérateurs logiques.

On pourrait se poser la question du choix du support de $\bar{X}_1$: pourquoi choisir cette ligne verticale du réseau réciproque plutôt qu'une autre, voire une ligne qui zigzague ? Ce choix est complètement arbitraire. En effet, en multipliant $\bar{X}_1$ par un opérateur étoile, on peut déformer son

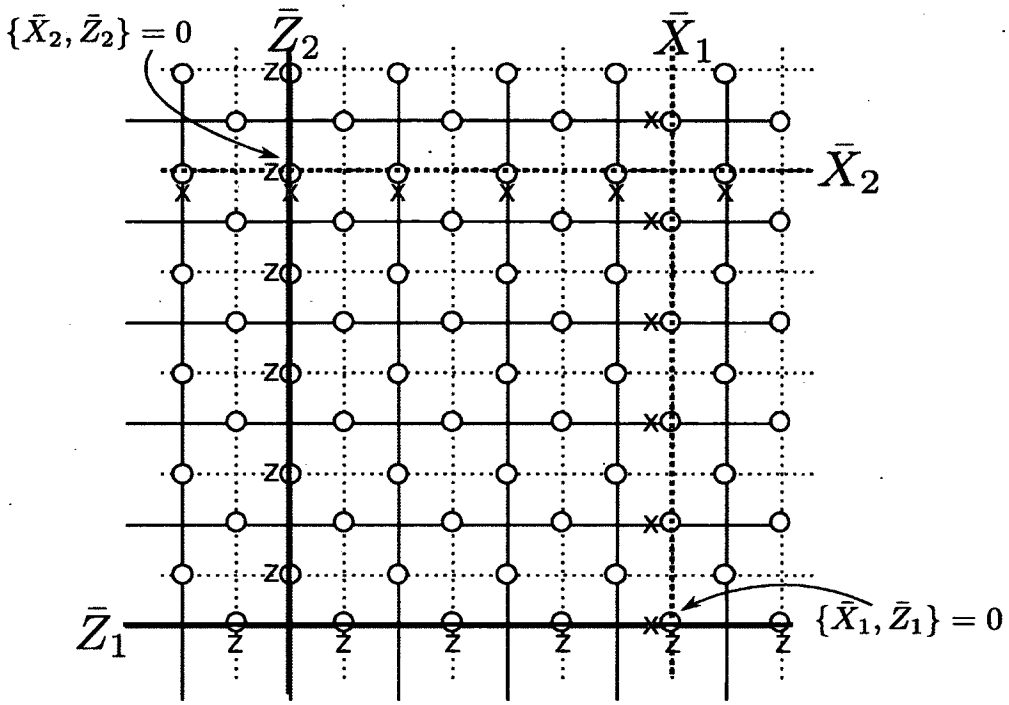


Figure 5.10 Générateurs du groupe logique du code torique.

support afin d'obtenir une autre courbe, mais qui fera toujours le tour du tore. Formellement, on a défini une relation d'équivalence sur les opérateurs par $A \sim B \Leftrightarrow \exists S \in \mathcal{S} \quad A = BS$. Ainsi, deux opérateurs équivalents ont exactement le même effet dans l'espace fondamental puisque, par définition, les stabilisateurs agissent trivialement sur le code. Les quatre opérateurs $\bar{X}_{1,2}$, $\bar{Z}_{12}$ sont donc des représentants canoniques qui génèrent l'algèbre du groupe logique.

Voilà la nature essentielle de l'opérateur logique : son support géométrique est une boucle non-triviale sur le tore, *i.e.*, une courbe qui ne peut être contractée vers un point. En effet, les stabilisateurs correspondent à toutes les boucles triviales d'opérateurs et ont un effet trivial sur le code. Ainsi, les quatre opérateurs logiques correspondent à des boucles non-triviales suivant les deux directions du tore, soit avec des opérateurs X, soit avec des opérateurs Z. On voit ainsi poindre la nature topologique du code torique. Quels sont les autres propriétés qui manifestent sa nature topologique ?

5.3.3 Propriétés topologiques

Dans cette section, nous passerons en revue les propriétés topologiques du code torique. En premier lieu, nous vérifierons que le code torique possède un fondamental topologiquement ordonné, selon la condition (5.5). Nous montrerons ensuite que les excitations de basse énergie du code torique sont décrites par un modèle anyonique. Finalement, nous verrons que la dégénérescence du code dépend de la topologie de la surface 2D.

5.3.3.1 Indiscernabilité locale des états fondamentaux

Nous allons vérifier que le fondamental du code torique obéit à la condition d'ordre topologique donnée par l'éq. (5.5). Soit O_{loc} un opérateur local. On veut connaître l'action de l'opérateur local à l'intérieur du fondamental, *i.e.*, on s'intéresse à l'opérateur $PO_{\text{loc}}P$ où P est le projecteur sur le fondamental. Pour vérifier la condition TQO, nous allons

- expliciter la structure du projecteur sur l'espace fondamental P
- « nettoyer » l'opérateur local en le multipliant par un stabilisateur
- montrer que les opérateurs dont le support est une réunion de courbes ouvertes créent des états excités

Structure du projecteur fondamental L'espace fondamental est l'espace propre commun associée à la valeur propre $+1$ de tous les stabilisateurs $S \in \mathcal{S}$. Le projecteur sur l'espace fondamental est donc l'intersection des projecteurs sur l'espace propre +1 des stabilisateurs. En fait, il suffit de considérer les projecteurs sur l'espace propre $+1$ d'une famille génératrice de $\mathcal{S}$, par exemple les opérateurs étoiles et les opérateurs plaquettes. Ainsi, le projecteur P est le produit des projecteurs associés à chacun de ces opérateurs, *i.e.*,

$$P = \prod_{s\in\Lambda} \frac{\mathbb{I} + A_s}{2} \prod_{p\in\Lambda^*} \frac{\mathbb{I} + B_p}{2} \tag{5.18}$$

Or, chacun de ces produits a une structure bien particulières. Le produit $P_+ = \prod_{s\in\Lambda} \frac{\mathbb{I}+A_s}{2}$ est proportionnel à la somme sur tous les opérateurs de Pauli de type X dont le support est une réunion de boucles fermées $\mathcal{B}$ contractibles sur le réseau réciproque

$$P_+ = \prod_{s\in\Lambda} \frac{\mathbb{I} + A_s}{2} = \frac{1}{2^{n/2}} \sum_{\mathcal{B}\subset\Lambda^*} X_{\mathcal{B}}. \tag{5.19}$$

Voyons que c'est le cas sur l'exemple de trois opérateurs étoiles situés sur trois sites adjacents s_1, s_2 et s_3. On a

$$\prod_{i=1}^{3} \mathbb{I} + A_{s_i} = \mathbb{I} + A_{s_1} + A_{s_2} + A_{s_3} + A_{s_1}A_{s_2} + A_{s_1}A_{s_3} + A_{s_2}A_{s_3} + A_{s_1}A_{s_2}A_{s_3} \tag{5.20}$$

Chaque terme correspond à boucle différente ou à une réunion de boucles (pour $A_{s_1}A_{s_3}$). Par exemple, les termes $A_{s_1}A_{s_2}$, $A_{s_1}A_{s_3}$ et $A_{s_1}A_{s_2}A_{s_3}$ sont représentés sur la figure 5.11.

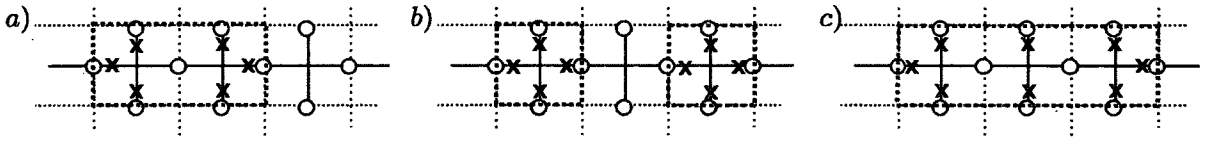


FIGURE 5.11 Trois exemples d'opérateurs boucles sur 3 sites.
En a), le produit $A_{s_1}A_{s_2}$ devient un opérateur rectangle sur 6 sites. En b), le produit $A_{s_1}A_{s_3}$ est supporté sur deux boucles disjointes. En c), le produit $A_{s_1}A_{s_2}A_{s_3}$ est un opérateur rectangle sur 8 sites.

De même, le produit $P_\square \equiv \prod_{p\in\Lambda^*} \frac{\mathbb{I}+B_p}{2}$ est proportionnel à la somme de tous les opérateurs de Pauli de type Z dont le support est une réunion de boucles fermées contractibles du réseau direct

$$P_\square = \prod_{p\in\Lambda^*} \frac{\mathbb{I}+B_p}{2} = \frac{1}{2^{n/2}} \sum_{B'\subset\Lambda} Z_{B'}. \tag{5.21}$$

En fait, les stabilisateurs sont exactement les opérateurs de la forme $X_B Z_{B'}$. Autrement dit,

$$P = \sum_{S\in\mathcal{S}} S. \tag{5.22}$$

Procédure de nettoyage Ainsi, lorsque l'on s'intéresse à $PO_{\text{loc}}P$, on peut « nettoyer » O_{loc} en le mutlipliant par un stabilisateur S afin d'obtenir un nouvel opérateur $O'_{\text{loc}} = O_{\text{loc}}S$ qui agit de la même façon dans l'espace fondamental puisque $PO'_{\text{loc}}P = PO_{\text{loc}}SP = PO_{\text{loc}}P$. En particulier, on peut éliminer tous les opérateurs de Pauli dont le support est une réunion de boucles fermées, *i.e.*, les stabilisateurs. Ainsi, on peut remplacer tout opérateur O_{loc} par un opérateur équivalent pour le code de la forme

$$O_{\text{loc}} = \sum_{P_i \in \mathcal{S}} \alpha_i P_i + \sum_{P_i \notin \mathcal{S}} \beta_i P_i \quad \rightarrow \quad \tilde{O}_{\text{loc}} = \left(\sum_i \alpha_i \right) \mathbb{I} + \sum_{P_i \notin \mathcal{S}} \beta_i P_i \tag{5.23}$$

et les opérateurs de Pauli locaux qui ne sont pas dans le groupe stabilisateur ont un support géométrique qui est une réunion de courbes ouvertes.

Courbes ouvertes créent des excitations De tels opérateurs peuvent être écrits comme le produit tensoriel d'opérateurs $P_{\mathcal{O}}$ agissant chacun sur une courbe ouverte $\mathcal{O}$, *i.e.*, ils sont de la forme

$$\bigotimes_{\mathcal{O} \subset \text{supp}(P_i)} \gamma_{\mathcal{O}} P_{\mathcal{O}}. \tag{5.24}$$

Un tel opérateur $P_{\mathcal{O}}$ anti-commute avec les générateurs $g_{(i)}$ du groupe stabilisateur situés aux extrémités de la courbe ouverte. On a donc, pour tous états fondamentaux $|\Omega\rangle$, $|\Omega'\rangle$

$$\langle \Omega' | P_{\mathcal{O}} | \Omega \rangle = \langle \Omega' | P_{\mathcal{O}} g_{(i)} | \Omega \rangle = -\langle \Omega' | g_{(i)} P_{\mathcal{O}} | \Omega \rangle = -\langle \Omega' | P_{\mathcal{O}} | \Omega \rangle \tag{5.25}$$

ce qui montre que $\langle \Omega' | P_{\mathcal{O}} | \Omega \rangle = 0$. On en déduit que

$$P P_{\mathcal{O}} P = 0 \tag{5.26}$$

et donc que l'action d'un opérateur de Pauli dont le support est une réunion de courbes ouvertes est nul à l'intérieur de l'espace fondamental. Ils ne font apparaître que des états excités.

Preuve de l'indiscernabilité par une observable locale En combinant les résultats (5.23) et (5.26), on obtient que l'action d'un opérateur local à l'intérieur de l'espace code se réduit donc être proportionnel à l'identité

$$P O_{\text{loc}} P = \left(\sum_i \alpha_i \right) P = c(O_{\text{loc}}) P. \tag{5.27}$$

5.3.3.2 Modèle d'anyons

Premiers états excités = chaîne d'erreur Après nous être intéressés à la structure des états fondamentaux du code torique et avoir montré que l'espace fondamental est topologiquement ordonné, nous aimerions mieux comprendre la structure des états excités. Un premier pas dans cette compréhension a été de constater que les opérateurs de Pauli dont le support est une réunion de courbes ouvertes ne produisent que des états excités, cf. éq. 5.26. Nous allons maintenant déterminer quels sont les premiers états excités et montrer qu'il s'agit de quasi-particules anyoniques. Le comportement anyonique fournit l'intuition de l'instabilité thermique, démontrée au chapitre 7, qui apparaît pour les mémoires topologiques 2D.

Le hamiltonien du code torique est particulièrement simple. Il est non-frustré et composé d'une somme d'opérateurs qui commutent deux à deux, dont les valeurs propres sont ± 1. On peut ainsi penser à chaque opérateur étoile ou plaquette comme à une contrainte locale. Si elles sont toutes respectés, l'état est dans le fondamental, mais la violation d'une d'entre elle, *i.e.*, être dans l'espace propre -1 d'un opérateur étoile ou plaquette correspond à une énergie additionnelle de $\Delta E = 2$.

Est-il possible d'avoir un état d'énergie ΔE? Supposons qu'une seule contrainte soit violée, par exemple un opérateur étoile, *i.e.* $A_{s^*}|\phi\rangle = -|\phi\rangle$. Dans ce cas, la valeur moyenne du produit des tous les opérateurs étoiles sur $|\phi\rangle$ est -1, ce qui contredit le fait que ce produit donne l'identité, cf. éq (5.16)

Par contre, il est possible d'avoir des états d'énergie $2\Delta E$ qui sont précisément les premiers états excités. Un exemple d'un tel état est celui obtenu à partir d'un état fondamental $|\psi\rangle$ en appliquant une erreur Z_i sur un qubit i. Cette erreur anti-commute avec les opérateurs étoiles des sites s_i adjacents à ce qubit, modifiant leur valeur propre de $+1$ à -1. En effet, $Z_i|\psi\rangle = Z_i A_{s_i}|\psi\rangle = -A_{s_i} Z|\psi\rangle$. Ainsi, du point de vue de l'énergie deux quasi-particules viennent d'apparaître, chacune associée à une énergie ΔE, à partir du vide.

Supposons maintenant qu'on applique une autre erreur Z_j sur un qubit j adjacent au qubit i. Cette nouvelle erreur va anti-commuter avec les opérateurs des sites adjacent. Or, un site est adjacent aux deux qubits i et j. L'opérateur étoile de ce site change de signe deux fois et sa contrainte est donc satisfaite. En d'autres termes, seuls les opérateurs aux extrémités de la chaîne $Z_i Z_j$ voient leur contrainte violée. En terme de quasi-particules, on a donc déplacé celle-ci sans payer une énergie additionnelle, ce qui est représenté sur la figure 5.12. On peut ainsi bouger les quasi-particules, ce qui revient à allonger la chaîne d'erreur Z. Une telle erreur est représentée sur

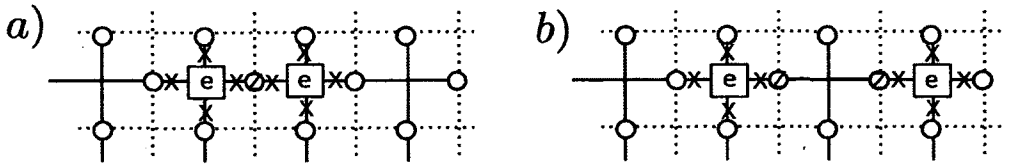


FIGURE 5.12 Propagations de quasi-particules électrique.

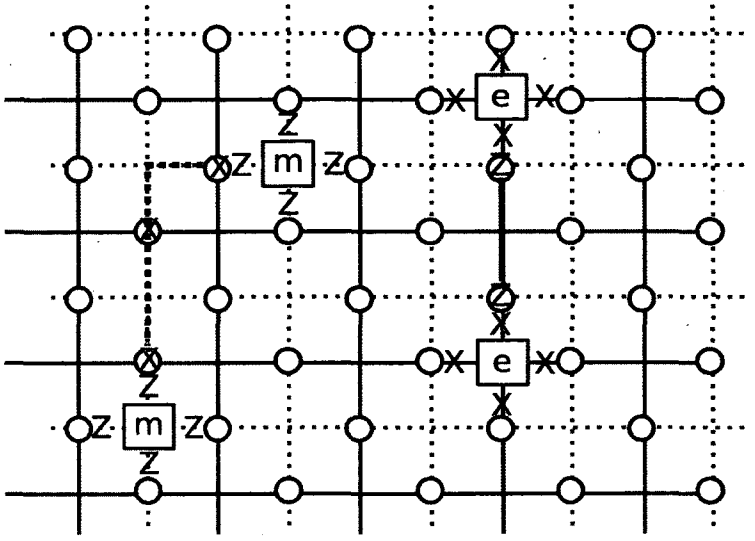


FIGURE 5.13 Anyons élémentaires du code torique : erreurs élémentaires de type magnétique (m) et électriques (e).

la figure 5.13.

Il apparaît donc deux types de quasi-particules, l'une associée aux chaines d'erreur Z, que nous appellerons des excitations électriques, l'àutre associée aux chaînes d'erreur X, que nous appellerons excitations magnétiques. Notons que ces excitations électriques, associées à des violations d'opérateurs étoiles, vivent sur des sites alors que les excitations magnétiques, associées à des violations d'opérateurs plaquette, vivent sur des plaquettes, i.e, des sites du réseau réciproque. Il s'agit de quasi-particules qui apparaissent en paires à partir du vide. Une erreur de type Y va faire apparaître une paire d'excitations magnétiques et d'excitations électriques. On peut aussi considérer qu'une excitation magnétique accompagnée d'une excitation électrique correspond à une quasi-particule composite.

Nous montrerons dans l'annexe C quelles sont les statistiques anyoniques de ces quasi-particules. Ce qu'il est important de retenir pour la suite est que ces anyons se propagent sans coût énergétique. De façon remarquable, tous les hamiltoniens locaux gappés dont le fondamental

obéit à la condition (5.5) ont des états excités qui semblent obéir à un modèle anyonique [11]. La relation entre les deux est encore mal compris, mais notre travail sur la stabilité des mémoires quantiques est un premier pas vers une meilleure compréhension de ce lien.

5.3.3.3 Dégénérescence en fonction de la topologie

La dégénérescence du fondamental du hamiltonien varie selon la topologie de la surface où il est appliqué. Ainsi, pour une sphère, le fondamental est unique car il y a exactement autant de stabilisateurs indépendants que de qubits. Sur un tore, l'espace fondamental est de dimension 4. Plus généralement, sur une variété de genre g, *i.e.*, un tore à g anses, il y aura $2g$ contraintes globales sur les générateurs. En effet, pour chaque anse, deux boucles non-triviales distinctes apparaissent. Ainsi, la dégénérescence du fondamental sera alors de 2^{2g}.

Cette dégénérescence est très intéressante du point de vue du stockage de l'information quantique. En effet, la dégénérescence 2^{2g} indique qu'il est possible de définir $2g$ qubits logiques. De plus, la condition d'ordre topologique (5.5) montre que toute erreur locale peut être corrigée. En effet, il s'agit simplement de la condition de Knill-Laflamme [89] pour une erreur locale. Ainsi, il est possible de mesurer les syndromes (les valeurs propres de chaque générateur du groupe stabilisateur) et d'en déduire l'opération de correction appropriée. Toutefois, diagnostiquer l'erreur la plus probable est un problème numérique difficile [12] pour lequel plusieurs décodeurs existent.

Ainsi, la correction d'erreur dite active demande beaucoup de ressources. Sur une échelle de temps courte (devant le délai typique d'apparition d'une erreur), il faut mesurer les syndromes, trouver l'erreur la plus probable, appliquer l'opérateur de correction . Il s'agit donc d'une mémoire qui doit être rafraichie en permanence, similaire à une mémoire RAM sur un ordinateur actuel. Or, il serait utile de disposer d'un disque dur quantique, qui encoderait de l'information quantique pour une longue durée, sans demander d'effectuer de la correction d'erreur active. Dans le domaine de l'informatique quantique, on parle alors de mémoire auto-correctrice, et le prochain chapitre s'intéressera aux propriétés désirées pour un tel système.

---

11. Notons que le caractère anyonique des excitations ne dépend pas de la topologie de la surface sur lequel est défini le hamiltonien. Ainsi, on aurait le même modèle anyonique si on s'était placé sur une sphère ou une variété de genre plus élevée.

12. Pour une excellente introduction aux décodeurs topologiques, rédigée en français, lire le mémoire de maîtrise de mon collègue Guillaume Duclos-Cianci [99].

Chapitre 6

Mémoires auto-correctrices

Dans ce chapitre, nous discuterons de la notion de mémoire auto-correctrice, l'équivalent quantique d'un disque dur. Nous verrons que deux propriétés sont attendues du hamiltonien d'un tel dispositif : la stabilité du spectre qui résulte de la robustesse à des perturbations et la stabilité thermique qui quantifie la robustesse à l'interaction avec un environnement. Nous verrons que les systèmes topologiques introduits dans le chapitre précédent sont de bons candidats en raison de leur stabilité spectrale. Nous nous interrogerons alors sur leur stabilité thermique. À l'aide d'un modèle simplifié de l'interaction mémoire-environnement, nous formulerons un critère de stabilité thermique, l'existence d'une barrière d'énergie qui grandit avec la taille de la mémoire. Malheureusement, nous montrerons dans le chapitre suivant que les mémoires topologiques basées sur des systèmes 2D ne présentent pas de telles barrières d'énergie.

Ce chapitre sert donc à mettre en place les notions nécessaires afin de comprendre l'instabilité thermique des mémoires topologiques 2D qui sera discutée dans le prochain chapitre. Notre article, présenté dans le prochain chapitre, prouve que les ingrédients de la stabilité spectrale mènent à une instabilité thermique.

6.1 Motivation

Une mémoire auto-correctrice serait l'équivalent quantique du disque dur ou de la mémoire flash que l'on retrouve sur nos ordinateurs actuels. Il s'agit d'un système physique encodant de l'information quantique pour une longue période de temps sans nécessiter d'intervention extérieure. La caractéristique clé par rapport aux codes de correction d'erreur est l'absence d'intervention active afin de le contrôler. Par exemple, pour les codes stabilisateurs, il faut mesurer les syndromes et diagnostiquer les erreurs sur une durée de temps très courte. Au contraire, une mémoire auto-correctrice serait capable de préserver l'information sur une longue durée en raison de l'interaction mémoire-environnement.

Un tel dispositif serait un atout majeur pour le calcul quantique, de la même façon qu'un dispositif de stockage (disque dur ou mémoire flash) est très précieux pour les ordinateurs actuels. En particulier, il permettrait de stocker des états quantiques correspondant à des étapes intermédiaires de calcul ou de stocker un résultat pour pouvoir l'utiliser plus tard. De même, il aurait de profondes conséquences pour la communication quantique et la cryptographie quantique [100]. En effet, la capacité de stocker des messages quantiques ouvre beaucoup de possibilités pour un adversaire quantique [101]. Du point de vue de la science fondamentale, la possibilité qu'un système préserve la cohérence quantique d'une superposition arbitraire d'états macroscopiquement distincts sur une durée arbitrairement longue est une question théorique majeure.

6.1.1 Propriétés désirées pour une mémoire auto-correctrice

Une mémoire auto-correctrice quantique vise à reproduire les propriétés d'un disque dur classique. Quelles sont-elles ?

À bien y penser, un disque dur est fascinant : on y encode de l'information puis on éteint son ordinateur. Quand on l'allume de nouveau, parfois plusieurs jours plus tard (oui, c'est possible, même pour des chercheurs), l'information est toujours présente. Durant la période de stockage, la mémoire a pourtant été laissée à elle-même, en interaction avec un environnement. Ainsi, pour une mémoire classique construite grâce à un matériau magnétique, les rayons cosmiques, par exemple, peuvent altérer l'état de la mémoire. Pourtant, l'information est encodée de façon robuste, par les propriétés physiques intrinsèques du système, puisqu'aucune intervention extérieure n'a été nécessaire.

Les propriétés désirées pour une mémoire quantique sont le reflet de ces propriétés d'un disque dur classique. Il s'agirait donc d'un système physique possédant un sous-espace vectoriel $\mathcal{C}$ où serait encodée l'information quantique. Initialement, l'information sera encodée sous la forme d'un état $|\psi_i\rangle \in \mathcal{C}$ puis le système sera laissé en interaction avec son environnement, ce qui sera une source d'erreur $\mathcal{E}$ sur la mémoire. Aucune intervention extérieure n'est permise durant cette période de stockage, d'une durée macroscopique. On aimerait que ce temps de stockage augmente polynomialement, voire exponentiellement, avec la taille de la mémoire. Finalement, la mémoire sera récupérée dans un état $\rho = \mathcal{E}(|\psi_i\rangle\langle\psi_i|)$. Une étape finale de correction d'erreur permettrait alors d'obtenir un état $|\psi_f\rangle \in \mathcal{C}$ à partir de ρ et on veut évidemment que $|\psi_f\rangle \simeq |\psi_i\rangle$.

La plupart de ces desiderata dépendent crucialement de l'erreur globale $\mathcal{E}$ qui demande de modéliser l'interaction système-environnement. Nous verrons en 6.4 qu'en faisant des hypothèses très générales sur le modèle de bruit reflétant l'interaction système-environnement, il est possible d'obtenir une définition formelle qui constitue une approximation de ce que pourrait être une mémoire auto-correctrice.

Afin de nous guider dans ce que pourrait être une mémoire auto-correctrice quantique, nous allons maintenant explorer un modèle simplifié d'une mémoire classique : le modèle d'Ising ferromagnétique.

6.1.2 Mémoire auto-correctrice classique : modèle d'Ising

Afin de modéliser un disque dur classique, nous allons considérer le modèle d'Ising ferromagnétique

$$H = -J \sum_{\langle i,j \rangle} \sigma_i^z \otimes \sigma_j^z \tag{6.1}$$

soit sur une chaîne 1D, soit sur un réseau carré 2D. Il s'agit bien sur d'une idéalisation de ce qui se passe réellement dans un matériau magnétique. Toutefois, il permettra de mettre en lumière des considérations importantes, en particulier le rôle crucial du paramètre d'ordre.

L'espace fondamental du modèle d'Ising est généré par les états $|\bar{0}\rangle \equiv |\uparrow\rangle^{\otimes n}$ et $|\bar{1}\rangle \equiv |\downarrow\rangle^{\otimes n}$, et peut donc stocker un bit classique en étant préparé dans un de ces deux états. Supposons que le bit encodé soit dans l'état 0 et que ce soit donc l'état $|\uparrow\rangle^{\otimes n}$ qui soit préparé. La mémoire est

ensuite placée en contact avec un réservoir de chaleur à une température finie T. Les fluctuations thermiques vont alors faire basculer certains spins, ce qui mènera à la formation de domaines où les spins sont dans l'état $|\downarrow\rangle$. Chacun de ces domaines correspond à une énergie proportionnelle à sa frontière. Le comportement devient alors très différent selon qu'on se place en 2D ou en 1D.

6.1.2.1 (In)stabilité thermique en 1D/2D

En 1D, la frontière des domaines est un mur entre deux spins adjacents. Supposons qu'un seul spin bascule dans la chaîne : cela entraîne l'apparition de deux murs de domaines et donc d'une énergie $4J$. Supposons maintenant qu'un autre spin adjacent à un mur de domaine bascule. Cela n'a pour effet que de déplacer un des murs de domaine et l'énergie ne change pas. Autrement dit, les murs de domaines sont des excitations localisées qui se déplacent sans payer d'énergie. En particulier, des domaines arbitrairement grands peuvent se former dès qu'il y a assez d'énergie pour basculer un spin. Intuitivement, on comprend bien que l'influence de l'état initial disparaît rapidement et que l'information encodée sera perdue.

En 2D, la frontière des domaines est associée à un coût énergétique qui grandit comme leur périmètre. Ainsi, plus le domaine est grand, plus il sera associé à une énergie élevée.

Plaçons-nous à température donnée et considérons les configurations possibles à l'énergie β, ce qui revient à fixer le nombre de spins basculés. Deux cas extrêmes se présentent, le cas où tous les spins basculés sont regroupés et le cas où tous ces spins sont disséminés. A première vue, une configuration avec un grand domaine basculé sera favorisée, puisqu'elle minimise l'énergie. Or, il faut aussi tenir compte de l'entropie. En effet, à température finie, on cherche à minimiser l'énergie libre $F = E - TS$, et le terme d'entropie aura tendance à favoriser les très nombreuses configurations où il y a plusieurs petits domaines.

À faible température, il faut donc faire un compromis entre les deux. On montre qu'en dessous d'une température critique (la température de Curie), aucun domaine macroscopique n'apparaît et le paramètre d'ordre, l'aimantation M est non-nulle. De plus, la valeur de M garde la mémoire de l'état initial.

Ainsi, le modèle d'Ising en 2D constitue une mémoire stable thermiquement car elle préserve l'information sur l'état initial à température assez faible. Ainsi, en ayant préparé initialement la mémoire dans l'état $|\bar{0}\rangle$ ou $|\bar{1}\rangle$, si on refroidit le système après une longue période de temps, soit grâce à de la correction d'erreur, soit en abaissant la température, on retrouvera l'information

encodée. Au contraire, le modèle d'Ising en 1D serait une mémoire instable thermiquement car l'information sur l'état initial est perdue. Le modèle d'Ising ferromagnétique 2D est donc protégé par une transition de phase et est donc une bonne mémoire classique. Ferait-il une bonne mémoire quantique ?

6.1.2.2 Stabilité du spectre

Supposons maintenant qu'on veuille utiliser le modèle d'Ising pour stocker de l'information quantique. Dans ce cas, il est primordial de préserver la dégénérescence du fondamental afin de protéger les superpositions arbitraires d'état. Or, une perturbation $J\epsilon\sigma_k^z$ sur n'importe quel spin k suffit à détruire la cohérence entre $|\bar{0}\rangle \equiv |\uparrow\rangle^{\otimes n}$ et $|\bar{1}\rangle \equiv |\downarrow\rangle^{\otimes n}$. En effet, les niveaux d'énergie auront alors un déplacement $\delta E = 2J\epsilon$. En supposant que ce champ fluctue, un état initial $|\psi(0)\rangle = \alpha|\bar{0}\rangle + \beta|\bar{1}\rangle$ deviendra rapidement un mélange statistique $\rho = |\alpha|^2 |\bar{0}\rangle\langle\bar{0}| + |\beta|^2 |\bar{1}\rangle\langle\bar{1}|$. Ainsi, l'état quantique de la mémoire est détruit.

Plus généralement, s'il existe un paramètre d'ordre qui distingue les différents états fondamentaux, il suffit à un champ extérieur de se coupler à ce paramètre d'ordre pour lever la dégénérescence de l'espace fondamental. Ainsi, tout système qui possède un paramètre d'ordre local serait déstabilisé par ce genre de perturbations !

Pour une mémoire auto-correctrice quantique, on cherche donc un hamiltonien dont l'espace fondamental dégénéré est à la fois stable thermiquement, comme celui du modèle d'Ising 2D, et dont le spectre est robuste face aux perturbations. Plus précisément, on cherche un hamiltonien local gappé dont (i) le gap ne se ferme pas sous l'action d'une perturbation locale et (ii) la dégénérescence du fondamental n'est pas levée par une perturbation locale. En particulier, on vient de montrer que le critère (ii) n'est pas satisfait pour un système possédant un paramètre d'ordre local. Heureusement, on a vu au chapitre précédent des hamiltoniens à l'espace fondamental dégénéré et sans paramètre d'ordre local : les systèmes topologiquement ordonnés !

6.2 Codes à projecteurs commutatifs

Dans cette section, nous allons définir une vaste classe de codes correcteurs, appelés codes à projecteurs commutatifs ou CPC, susceptibles de fournir une mémoire auto-correctrice 2D. Ils regroupent la plupart des modèles d'ordre topologique. Les codes CPC topologiques seront nos candidats pour une mémoire auto-correctrice.

Une façon de les introduire est de considérer qu'ils sont une extension des codes stabilisateurs, cf. 5.3.1. Un code stabilisateur défini par une famille génératrice $g_1 \dots g_\ell$ est l'espace fondamental du hamiltonien $H = -\sum_{k=1}^{\ell} g_k$ où les g_k commutent deux à deux et qui ne présentent pas de frustration. Nous maintenant définir les codes CPC en gardant ces propriétés, mais en levant la restriction d'être lié à un sous-groupe des opérateurs de Pauli.

6.2.1 Hamiltoniens locaux, non-frustrés, commutatifs

Soit n particules quantiques à d niveaux, ou qudits, placées sur les sites d'un réseau $\Lambda = (V, E)$ composés de vertex $v \in V$ et d'arêtes $e \in E$ en dimension D. Dans le chapitre suivant, nous nous restreindrons à $D = 2$, mais dans ce chapitre, D ne sera pas fixé. On va s'intéresser à une famille très générale de hamiltoniens locaux, sans frustration et dont les termes commutent de la forme

$$H = -\sum_{X \subset V} h_X \tag{6.2}$$

où h_X est un opérateur hermitien qui agit non trivialement sur les sites X. Ce hamiltonien respecte de plus les propriétés suivantes :

1. La norme d'opérateur de chacun des termes est bornée, *i.e.*, $\|h_X\| \equiv \max_\psi \sqrt{\frac{\langle\psi|h_X^2|\psi\rangle}{\langle\psi|\psi\rangle}} \leq 1$. Puisque h_X est hermitien, cela revient à dire que sa plus grande valeur propre en valeur absolue est bornée par 1. Intuitivement, on impose cette condition afin de ne pas tricher en « cachant » une grande quantité d'énergie dans un des termes du hamiltonien.
2. h_X est un terme local au sens où h_X est forcément nul si le diamètre de la région X est supérieur à une portée d'interaction w. Le diamètre $\mathrm{diam}(X)$ est la plus grande distance entre deux sites de X. On suppose en effet qu'une distance soit définie sur le réseau Λ, par exemple la distance euclidienne sur un réseau carré.
3. Les termes du hamiltonien commutent deux à deux, *i.e.* $[h_X, h_Y] = 0$.

4. Le hamiltonien est sans-frustration. Ainsi, un état fondamental global $|\Omega\rangle$ est aussi un état fondamental de chacun des termes élémentaires $-h_X$.

Les critères 3 et 4 permettent d'obtenir des hamiltoniens solubles dont on connaît exactement le spectre. Ainsi, ils sont simples à manipuler et il est souvent possible de prouver des résultats rigoureux. L'espoir est que les résultats s'étendent à tous les hamiltoniens de la même phase. Par exemple, un résultat valide pour un espace fondamental topologiquement ordonné sera vrai pour tous les états fondamentaux des hamiltoniens dans la même phase car on passe de l'un à l'autre grâce à des transformations locales. La question intéressante devient alors d'identifier les phases qui possèdent des hamiltoniens locaux, non-frustrés et commutatifs. Le folklore dans la communauté est que tous les hamiltoniens locaux et gappés peuvent être renormalisés vers un hamiltonien local, non-frustré et commutatif, sauf s'ils appartiennent à une phase chirale.

6.2.2 Codes à projecteurs commutatifs

Dans le contexte d'une mémoire quantique, on ne s'intéresse qu'à l'espace fondamental de H et à la variation du gap d'énergie en fonction de la taille N du système. Il est alors utile de définir un nouveau hamiltonien, noté H_{CPC}, qui a le même espace fondamental que H et dont le gap est minoré par 1. Cette simplification mathématique permet de simplifier la structure du spectre en déplaçant le zéro d'énergie, redéfinissant l'échelle d'énergie et en regroupant les états excités selon le nombre de termes élémentaires dont ils ne sont pas l'état fondamental. Toutefois, les conclusions que nous obtiendrons ne sont pas affectées par cette simplification et restent valides pour tous les hamiltoniens de la forme (6.2). Pour définir H_{CPC}, on décompose spectralement chaque terme élémentaire h_X pour ne garder que le projecteur P_X associé à sa plus grande valeur propre λ_1 que l'on redimensionne à 1

$$h_X = \sum_{\lambda} \lambda \Pi_\lambda = \lambda_1 \Pi_1 + \sum_{\lambda_i < \lambda_1} \lambda_i \Pi_i \mapsto P_X \equiv \Pi_1. \tag{6.3}$$

L'hamiltonien obtenu appartient alors à la classe des codes à projecteurs commutatifs ou CPC qui sont de la forme

$$H_{CPC} = -\sum_{X \subset V} P_X \tag{6.4}$$

où les P_X sont des projecteurs, *i.e.*, $(P_X)^2 = P_X$, locaux, qui commutent $[P_X, P_Y] = 0$ et un fondamental $|\Omega\rangle$ de H_{CPC} vérifie $\forall X \; P_X|\Omega\rangle = +|\Omega\rangle$. En particulier, le projecteur sur l'espace

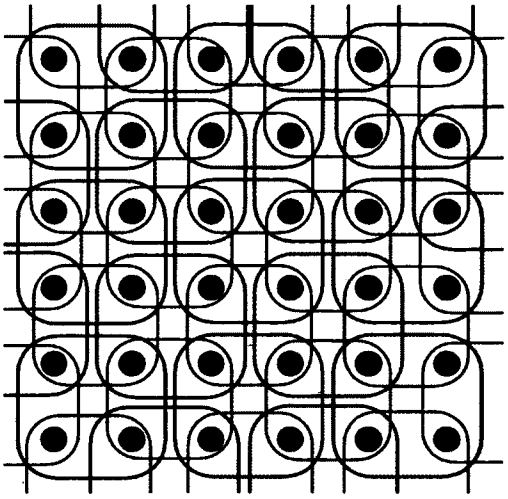


FIGURE 6.1 Structure géométrique d'un code CPC. Avec un regroupement approprié des sites et des projecteurs, on peut supposer que Λ est un réseau carré régulier et que chaque projecteur agit sur une cellule 2x2. Ainsi, chaque rectangle aux coins arrondis représente le support géométrique d'un projecteur et chaque site est représenté par un cercle. Sur la figure, des conditions aux frontières périodique sont utilisées.

fondamental $\mathcal{C}$, souvent appelé code, est

$$P_{\mathcal{C}} = \prod_{X} P_X. \tag{6.5}$$

Un opérateur logique est un opérateur L qui agit dans le code, *i.e.*, $[L, P_{\mathcal{C}}] = 0$. Graphiquement, ces codes sont représentés sur la Fig. 6.1.

Cette classe de hamiltonien est générale et possède une structure assez simple afin de prouver rigoureusement des résultats généraux. En particulier, dans le but de définir une mémoire quantique auto-correctrice, nous nous intéressons à la sous-classe de ces codes qui présentent de l'ordre topologique, au sens de la condition TQO. D'ailleurs, la plupart des exemples connus de modèles topologiques sont des codes à projecteurs commutatifs. En particulier, les codes stabilisateurs peuvent être mis sous cette forme. Plus généralement, le code torique [80], les codes de couleur [102], les modèles de réseaux de boucles de Levin-Wen [103], les modèles Turaev-Viro [104], les modèles doubles de Kitaev [80] font partie de cette catégorie.

Nous allons maintenant voir que l'on dispose d'un résultat très puissant afin de garantir la stabilité du spectre des modèles CPC qui présentent de l'ordre topologique.

6.3 Robustesse aux perturbations : stabilité spectrale

La robustesse du spectre aux perturbations est essentielle pour (au moins) deux raisons[1]. La première est due à des contraintes pratiques liées à la réalisation d'une mémoire quantique. Cette mémoire sera décrite en pratique par un hamiltonien qui n'aura sans doute pas exactement la forme du hamiltonien théorique. Donc, si les propriétés désirées ne sont vraies que pour le hamiltonien théorique, il sera très difficile, voire impossible de mettre au point cette mémoire. Autrement dit, la robustesse aux perturbations quantifie l'écart qui peut être toléré entre le hamiltonien théorique et sa réalisation expérimentale. La seconde est due aux perturbations cohérentes de l'environnement. En effet, on cherche à avoir une mémoire qui préserve l'information même en interaction avec un environnement. Par exemple, il n'est pas exclu qu'un champ magnétique vienne perturber ce hamiltonien. Il faut donc que ses propriétés soient robustes à cette perturbation.

Intuitivement, la nature topologique des degrés de liberté d'un code topologique entraîne que le spectre soit stable. Nous explorerons en détail le cas du code torique, où le gap ne se ferme pas et où la levée de dégénérescence n'est qu'exponentiellement petite. Cela constituera donc notre définition de la stabilité spectrale. Nous nous interrogerons alors sur les mécanismes d'instabilité et les propriétés à éviter dans un hamiltonien. Cela nous amènera à montrer que la condition d'ordre topologique ne suffit pas à garantir la stabilité spectrale. Nous verrons toutefois qu'une condition additionnelle, dite de cohérence locale, permet de prouver la stabilité spectrale. Ce résultat essentiel [105] sera rappelé.

6.3.1 Robustesse spectrale du code torique

Avant de présenter un résultat général, intéressons-nous à la robustesse du code torique. Le comportement du code torique nous servira de guide pour définir la stabilité spectrale d'un code CPC topologique. Considérons donc le spectre de $H + V$ où $V = \sum_X v_X$ est une somme de termes locaux dont la norme d'opérateur est bornée, *i.e.* $\|v_x\| \leq J$. Nous allons d'abord nous intéresser à la variation des niveaux d'énergie fondamentaux puis à celle du gap.

1. La distinction entre ces deux raisons est artificielle. Elle revient à déplacer la frontière entre la mémoire et son environnement.

Levée exponentiellement petite de la dégénérescence Intéressons-nous à la levée de la dégénérescence δ entre deux états fondamentaux orthogonaux $|\Omega\rangle$ et $|\Omega^\perp\rangle$. À l'ordre m en théorie des perturbations, celle-ci sera proportionnelle à l'élément de matrice $\langle\Omega|V^m|\Omega^\perp\rangle$ ou à la différence $\langle\Omega|V^m|\Omega\rangle - \langle\Omega^\perp|V^m|\Omega^\perp\rangle$. Ces deux termes ne sont non-nuls que si V^m fait apparaître des boucles non-triviales d'opérateurs, de la forme $\bigotimes_{X\subset\mathcal{B}} v_X$ où $\mathcal{B}$ est une boucle non-triviale. Or, V est local donc V^m ne fera apparaître de telles boucles que si $m = \alpha L$ où L est la plus petite dimension du tore. La levée de dégénérescence sera donc de la forme $\delta \sim J^{\alpha L}$ qui est exponentiellement petite en L.

Robustesse du gap Notons tout d'abord que même si $J \ll 1$, la perturbation V contient assez d'énergie pour fermer le gap. En effet, la norme d'opérateur de la perturbation est

$$\|V\| \leq \sum_i \|V_i\| = |\Lambda|\,\|V_i\| \tag{6.6}$$

et il s'agit donc d'une grandeur extensive. Par exemple, dans le cas du modèle d'Ising, un champ magnétique $h\sum_i \sigma_i^z$ a une norme d'opérateur $h \times n$ qui est comparable au gap dès que $h \sim 1/n$, ce qui correspond à un champ très faible. Ainsi, on ne peut pas exclure a priori qu'une perturbation locale ferme le gap car l'énergie amenée par la perturbation est extensive. Toutefois, nous allons donner l'intuition que pour le code torique, seuls des termes locaux peuvent intervenir.

Intéressons-nous à la variation des niveaux d'énergie des états excités. Il est commode de penser à ceux-ci comme étant obtenus à partir du vide à partir de chaîne d'erreur X ou Z. Supposons qu'on s'intéresse à des premiers états excités obtenus à partir de chaîne d'erreurs X. La variation de leur niveau d'énergie sera alors de la forme $\langle e_1|V^m|e_2\rangle$ où $|e_1\rangle = X_{\mathcal{C}_1}|\Omega^{(1)}\rangle$ et $|e_2\rangle = X_{\mathcal{C}_2}|\Omega^{(2)}\rangle$. Même à l'ordre 1, cette contribution peut être non-nulle si $\mathcal{C}_1 \cup \mathrm{supp}(V) \cup \mathcal{C}_2$ forment une boucle non-triviale. Toutefois, seuls quelques termes locaux de V vont contribuer. Ainsi, ces termes seront bornés par la norme d'opérateur des termes locaux J qu'il suffit de choisir plus petite que le gap afin qu'il ne se ferme pas.

6.3.2 Définition de la stabilité spectrale

Par analogie avec le code torique, on dira qu'un hamiltonien gappé H a un spectre stable si pour toute perturbation locale V dont la norme d'opérateur est assez petite

1. le gap Δ de $H + V$ est une constante, indépendante de la taille du système
2. la levée de dégénérescence δ est exponentiellement petite dans la taille du système.

6.3.2.1 Mécanismes d'instabilité

Afin de mieux cerner les propriétés qui rendent un hamiltonien instable spectralement, nous allons nous présenter deux exemples simples : l'un où la dégénérescence des états fondamentaux est levée, l'autre où le gap est fermé par une perturbation.

Levée de dégénérescence L'exemple-type de la levée de dégénérescence est le modèle d'Ising en présence d'un champ magnétique externe.

$$H = -\sum_{\langle i,j \rangle} \sigma_i^z \otimes \sigma_j^z \qquad V = -h \sum_i \sigma_i^z \tag{6.7}$$

Alors que le hamiltonien non-perturbé présente un fondamental dégénéré généré par $|\uparrow\rangle^{\otimes n}$ et $|\downarrow\rangle^{\otimes n}$, l'état fondamental de $H + V$ est $|\uparrow\rangle^{\otimes n}$ et son premier état excité est $|\downarrow\rangle^{\otimes n}$ d'énergie $E_0 + hn$. Ainsi, il suffit d'un champ d'une intensité $h \sim 1/n$ pour lever la dégénérescence. Notons que ce champ vient se coupler avec le paramètre d'ordre σ_i^z. A priori, ce mécanisme d'instabilité est donc éliminé pour des hamiltoniens topologiques.

Fermeture du gap La fermeture du gap est plus subtile. Elle contient les ingrédients qui permettront de montrer que l'ordre topologique seul ne suffit pas à garantir la stabilité spectrale. Considérons le modèle d'Ising avec un défaut magnétique sur le site i^* donné par l'éq. (6.8)

$$H = -\sum_{\langle i,j \rangle} \sigma_i^z \otimes \sigma_j^z - \sigma_{i^*}^z \qquad V = h \sum_i \sigma_i^z \tag{6.8}$$

Le hamiltonien non perturbé H possède un unique état fondamental $|\uparrow\rangle^{\otimes n}$ alors que $|\downarrow\rangle^{\otimes n}$ est un état excité d'énergie $E_0 + 1$. Or, ce gap en faveur de $|\uparrow\rangle^{\otimes n}$ peut être fermé par un faible champ magnétique externe qui favorise $|\downarrow\rangle^{\otimes n}$. Ainsi, le gap du hamiltonien perturbé $H + V$ se ferme pour $h \sim 1/n$.

Ce mécanisme d'instabilité n'est pas tant du à un paramètre d'ordre local, qu'à la présence d'une région particulière, le site i^*. L'interaction d'Ising entre premier voisin, bien que locale, va étendre l'influence du défaut à tous les sites. Nous allons voir que ce mécanisme d'instabilité n'est pas éliminé par l'ordre topologique.

6.3.3 TQO ne garantit pas la stabilité du spectre

6.3.3.1 Insuffisance de la condition TQO

En s'inspirant du résultat pour le code torique, on pourrait penser que la condition TQO suffit pour garantir la stabilité du spectre. Or, ce n'est pas le cas, comme le montre le contre-exemple suivant, tiré de [105].

Considérons le hamiltonien suivant, qui donne une interaction de type Ising entre les opérateurs plaquette du code torique avec une plaquette-défaut p^*

$$H_{IC} = -\sum_{\langle p,p' \rangle} B_p \otimes B_{p'} - B_{p^*} - \sum_s A_s \tag{6.9}$$

L'espace fondamental de H_{IC} est celui des états pour lesquels $\forall p\ B_p = +1$ et $\forall s\ A_s = +1$. Il s'agit donc exactement du fondamental du code torique, qui respecte la condition TQO. On en déduit que H_{IC} respecte aussi la condition TQO. Par contre, il est possible de fermer le gap de ce hamiltonien, à l'aide d'un « champ magnétique de plaquette »

$$V = \frac{1}{|\Lambda|} \sum_p B_p \tag{6.10}$$

Le hamiltonien $H_{IC} + V$ contient maintenant le secteur $A_s = +1$, $B_p = -1$ qui étaient les premiers états excités de H_{IC}.

6.3.3.2 Condition additionnelle : cohérence locale

Ainsi, il faut ajouter une deuxième condition afin d'obtenir la robustesse spectrale. Il s'agit de la condition de cohérence locale. Elle demande que le projecteur global sur le code $P_C = \prod_X P_X$ soit une extension de tous les projecteurs élémentaires locaux. Formellement, pour toute région A

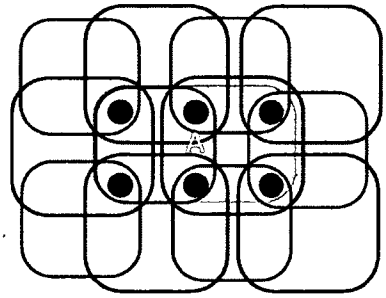


FIGURE 6.2 Support du projecteur P_A qui regroupe tous les projecteurs dont le support intersecte la région A (en rouge).

topologiquement triviale, on définit une version locale du projecteur sur le code par

$$P_A = \prod_{X \cap A \neq \emptyset} P_X \tag{6.11}$$

tel que représenté sur la Fig. 6.2. On veut comparer P_A au projecteur global P_C. Pour ce faire, on pense à $P_C = \sum_i |\Omega^{(i)}\rangle\langle\Omega^{(i)}|$ comme étant un étant quantique, plus précisément le mélange statistique équiprobable de tous les états du code, et on considère sa matrice densité réduite (non-normalisée) sur la région A

$$\rho_A = \mathrm{Tr}_{\bar{A}}\,[P] \tag{6.12}$$

On définit de même une matrice densité réduite de la version locale du projecteur sur le code

$$\rho_A^{\mathrm{loc}} = \mathrm{Tr}_{\bar{A}}\,[P_A]\,. \tag{6.13}$$

Par construction, on sait que

$$\ker \rho_A^{\mathrm{loc}} \subset \ker \rho_A \tag{6.14}$$

(où ker désigne le noyau) puisque P_C contient plus de contraintes que P_A. On dira que ces deux états sont compatibles s'ils ont le même noyau, i.e.,

$$\ker \rho_A = \ker \rho_A^{\mathrm{loc}} \tag{6.15}$$

Définition 6 (Cohérence locale ou LC). Le code respecte la condition de cohérence locale si $\ker \rho_A = \ker \rho_A^{\mathrm{loc}}$ pour toute région topologiquement triviale A.

Intuitivement, un état $|\psi\rangle$ dans $\ker \rho_A \setminus \ker \rho_A^{\mathrm{loc}}$ est un état sur la région A qui ne peut pas

être étendu à un état $|\Psi\rangle$ global dans C. Autrement dit, il s'agit d'un état local qui est en dehors du code, mais que les projecteurs locaux qui intersectent A, ne peuvent pas éliminer. Typiquement, cette situation apparaît lorsque le hamiltonien contient un terme qui privilégie une région du réseau, par exemple la plaquette défectueuse du hamiltonien (6.9). Ainsi, les projecteurs locaux ne tiennent pas compte de l'influence du défaut alors que celui-ci modifie la structure du hamiltonien.

Pour des codes stabilisateurs, on peut montrer que la condition de cohérence locale demande que tout stabilisateur dont le support géométrique est strictement contenu dans la région A est le produit de générateur dont le support intersecte A. Ainsi, le modèle d'Ising avec défaut au site i^* est un code stabilisateur qui ne respecte pas la condition de cohérence locale. En effet, considérons un site k loin du défaut i^*. Z_k est un stabilisateur qui s'écrit $Z_k = Z_k \otimes Z_{k+1} \times \cdots \times Z_{i^*+1,i^*} \times Z_{i^*}$ qui fait apparaître une chaîne de générateurs. Notons que la condition de cohérence locale dépend du choix des générateurs. En effet, du point de vue stabilisateur, une autre famille génératrice est tout simplement $\{Z_i\}_{i\in V}$ qui n'est pas topologiquement ordonné par contre. Il n'est pas clair s'il est toujours possible de définir une nouvelle famille de projecteurs afin qu'un hamiltonien respectant la condition TQO respecte la condition de cohérence locale.

6.3.4 Résultat sur la stabilité spectrale

Bravyi, Hastings et Michalakis ont montré en 2010 que tout hamiltonien CPC qui respecte les conditions TQO et LC ont un spectre stable [105]. Elle est valable en dimension D quelconque. Ils considèrent une perturbation qui est une somme de termes centrés sur chaque site du réseau et qui ont une décroissance exponentielle. Formellement, elle est de la forme $V = \sum_{i\in\Lambda} V_i$ où chaque terme V_i, centré sur le site i, a une force J et une décroissance exponentielle μ

$$V_i = \sum_{r\geq 0} V_i(r) \tag{6.16}$$

où $V_i(r)$ agit non-trivialement que sur les sites situés à une distance inférieure à r de i et $\|V_i(r)\| \leq Je^{-\mu r}$.

Notons que même si la force J est très petite, la norme d'opérateur de la perturbation est extensive et donc que pour J petit mais constant, il serait a priori possible que le gap se ferme. Toutefois, ce n'est pas le cas.

En effet, le résultat sur la stabilité du spectre montre que les niveaux d'énergie $E_k(J)$ de

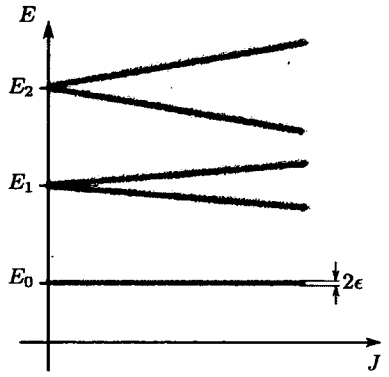


FIGURE 6.3 Bornes de la variation des niveaux d'énergie avec la force de la perturbation. L'effet de flou correspond au terme d'origine topologique ϵ qui est exponentiellement petit alors que la pente des droites reflète la variation des niveaux d'énergie qui est proportionnelle à la force J des termes locaux. Les niveaux d'énergie de $H + V$ se trouvent à l'intérieur de la région délimités par les droites.

$H + V$ vont s'élargir avec J croissant , mais rester à l'intérieur d'un intervalle $I_k = [\![E_k(1 - c_1 J) - \epsilon; E_k(1 + c_1 J) + \epsilon]\!]$ où c_1 est une constante ne dépendant que de μ et D, E_k est l'énergie non-perturbée et ϵ est un terme exponentiellement petit. Plus précisément, il s'écrit

$$\epsilon = \text{poly}(L) e^{-c_2 L^{3/8}} \tag{6.17}$$

avec c_2 une constante qui ne dépend que du facteur de décroissance μ et de la dimension D. Ce résultat est illustré sur la Fig. 6.3. En particulier, la levée de dégénérescence est $\delta \leq 2\epsilon$ qui est bien exponentiellement petite et le gap du hamiltonien perturbé est borné par $\Delta(J) \leq \Delta(1 - c_1 J) - 2\epsilon$.

Ainsi, ce résultat fondamental oriente la recherche de mémoires auto-correctrices vers ces modèles qui respectent les conditions TQO et LC. Malheureusement, nous avons montré, mon directeur et moi, que la condition de cohérence locale entraîne forcément l'apparition d'une instabilité thermique, permettant à l'environnement de corrompre l'état de la mémoire. Ce résultat est celui démontré dans l'article présenté dans le prochain chapitre, à la section 7.3. Afin de formuler ce résultat, il faut se doter d'une définition de la stabilité thermique, ce que nous allons maintenant présenter.

6.4 Thermalisation et barrière d'énergie

Dans cette section, nous proposerons une définition pour la stabilité thermique d'une mémoire. Cette capacité de préserver l'information en étant placé en interaction avec un réservoir thermique, sans aucune intervention extérieure, constitue précisément la propriété d'auto-correction tant recherchée pour ces mémoires. Or, la thermalisation est un phénomène compliqué qui demanderait de modéliser l'interaction entre la mémoire et son environnement. Bien que cette description soit possible pour un système et un environnement particuliers, notre objectif est plutôt de prouver un résultat qui soit valable pour une grande classe de modèles. Ainsi, nous devrons donc nous rabattre sur quelques principes fondamentaux plutôt qu'un modèle microscopique. Ceci nous permettra néanmoins de proposer un critère pour la stabilité thermique, l'existence d'une barrière d'énergie.

La protection de l'information quantique par une mémoire est un phénomène hors-équilibre : la mémoire est préparée dans un état fondamental, mis en contact avec un réservoir thermique durant une longue période avant qu'on tente de récupérer l'information. Dans le cas classique, la protection dans le modèle d'Ising provenait de l'existence d'une phase ferromagnétique sous la température de Curie. Il s'agit d'un critère fort de stabilité thermique. Dans le cas d'une mémoire quantique, nous nous contenterons d'un critère plus faible, celui de l'existence d'une barrière d'énergie, motivé par un modèle simplifié de la thermalisation que nous présentons maintenant.

6.4.1 Modélisation minimale de l'environnement

6.4.1.1 Action locale de l'environnement

L'idée de base de la protection topologique est de supposer que l'environnement ne sera pas capable d'affecter les degrés de liberté topologiques car son action est locale. Un traitement complet demanderait de modéliser l'interaction mémoire-environnement. Nous nous contenterons de définir un modèle d'erreur qui caractérise les transformations sur la mémoire résultant de cette interaction. Plus précisément, à chaque pas de temps k, une transformation locale sera appliqué sur la mémoire. Formellement, il s'agit d'une transformation CPTP $\mathcal{E}_k$ locale, *i.e.*, qui est une somme de termes n'agissant que sur un petit nombre de particules qui sont proches les unes des autres. Ainsi, l'évolution de la mémoire sera donné par une suite $\{\mathcal{E}_k\}_{k=1}^{T}$ où T joue le rôle de la durée d'évolution.

Ainsi, à partir d'un état initial $\rho_0 = |\psi_0\rangle\langle\psi_0|$, la mémoire se retrouvera successivement dans différents états $\rho_{t+1} = \mathcal{E}_{t+1}(\rho_t) = \prod_{k=1}^{t+1} \mathcal{E}_k(\rho_0)$, ce qui définit une suite d'états intermédiaires $\{\rho_t\}_{t=1}^{T}$.

6.4.1.2 Pénalisation des états de haute énergie

La question est maintenant de définir quelles sont les séquences d'états les plus probables. Pour se faire, on va considérer que l'environnement utilise une approche similaire à celle d'un algorithme Monte-Carlo. Ainsi, il pénalise les états de haute énergie et privilégie ceux de basse énergie, de façon similaire au facteur de Boltzmann qui stipule que la probabilité d'être dans une configuration d'énergie E à température inverse β est proportionnelle à $e^{-\beta E}$. Ainsi, les séquences d'états intermédiaires qui font apparaître des états de haute énergie seront moins probables que celles ne contenant que des états de basse énergie.

6.4.2 Barrière d'énergie

Si l'environnement amène deux états fondamentaux distincts $|\Omega\rangle$ et $|\Omega'\rangle$ au même état bruité ρ^*, il sera impossible de récupérer l'information quantique encodée dans la mémoire. En effet, l'opération de récupération finale ne dépend que de l'état final ρ^* de la mémoire. Une autre façon de voir les choses est de considérer la séquence d'états intermédiaires résultant du modèle de bruit et de la thermalisation. Si celle-ci fait passer d'un état fondamental $|\Omega\rangle$ à un état fondamental différent $|\Omega'\rangle$, alors l'état récupéré sera différent de l'état initialement encodé et une erreur logique sera apparue sur le code. Ainsi, la séquence d'erreurs $\{\mathcal{E}_k\}_{k=1}^{T}$ aura réussi à simuler l'opérateur logique L qui relie ces deux états fondamentaux. Dans ce cas, on parlera d'une séquence d'erreurs logique.

On va donc s'intéresser tout particulièrement à l'énergie maximale des états intermédiaires d'une telle séquence logique, que nous appellerons une barrière d'énergie. On aimerait que pour toute séquence de transformations locales qui font passer la mémoire d'un état fondamental à un état fondamental orthogonal, la barrière d'énergie soit grande, *i.e.*, qu'elle grandisse avec la taille du système. Autrement dit, on aimerait que les bassins d'attraction associés à des états fondamentaux différents soient séparés par une barrière d'énergie.

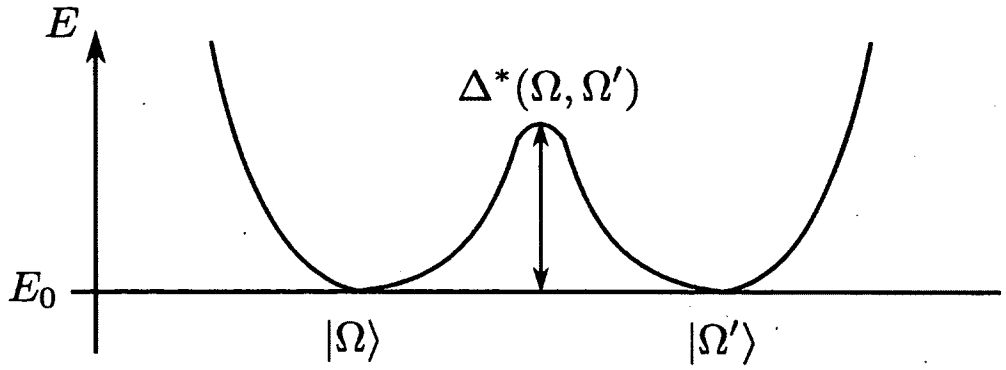


FIGURE 6.4 Paysage d'énergie entre deux bassins d'attractions d'états fondamentaux $|\Omega\rangle$ et $|\Omega'\rangle$ L'axe des ordonnées représente l'énergie alors que celui des abscisses correspond intuitivement à des transformations locales.

6.4.2.1 Paysage d'énergie

Intuitivement, une telle barrière d'énergie qui séparerait les différents états fondamentaux du hamiltonien garantirait la préservation de l'information encodée face à l'action de l'environnement. Nous allons donc décrire ce qui se passe dans un paysage d'énergie où le marcheur est l'état de la mémoire et l'altitude correspond à l'énergie des états intermédiaires. Cette image est forcément imparfaite, en particulier parce qu'on encode de l'information quantique et donc que le code est un sous-espace vectoriel et non pas un ensemble discret de mots-code. Toutefois, elle a le mérite de donner une intuition (peut-être dangereuse) sur ce qui se passe.

Initialement, la mémoire est préparée dans un état fondamental, et toute transformation locale aura pour effet d'augmenter l'énergie. Ainsi, on se trouve au fond d'une vallée, comme sur la figure 6.4. Il existe d'autres vallées correspondant aux autres états fondamentaux. Le tunnel qui permet de passer d'une vallée à l'autre est un opérateur logique. Or, pour les codes topologiques, ces opérateurs sont non-locaux et il est donc interdit de les emprunter. Une autre façon de voir les choses est que le marcheur est limité à faire de petits pas et l'opérateur logique demanderait des bottes de sept lieux.

La question qui se pose est de savoir s'il est possible de passer d'une vallée à l'autre grâce à une suite de petits pas (transformations locales) et surtout si cette séquence est probable. Une mesure approchée de cette probabilité est de regarder l'état intermédiaire de plus haute énergie.

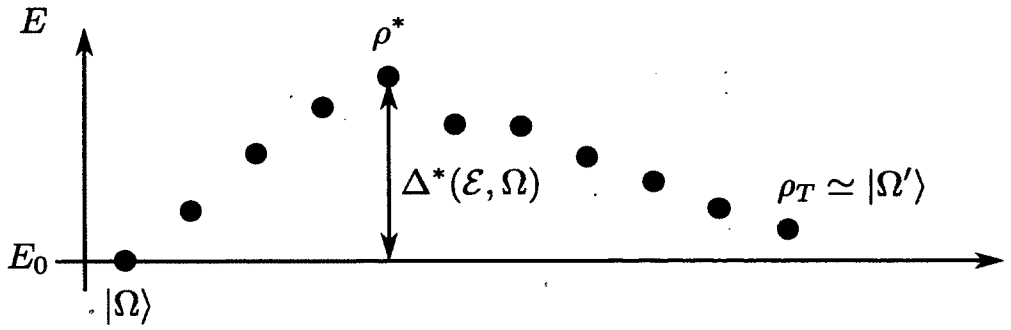


FIGURE 6.5 Séquence d'erreurs logique entre deux états fondamentaux $|\Omega\rangle$ et $|\Omega'\rangle$
L'axe des ordonnées représente l'énergie alors que chaque point est un état intermédiaire $\rho_t = \mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_t \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]$.

6.4.2.2 Définition formelle

Plus précisément, une séquence d'erreurs logique $\{\mathcal{E}_k\}_{k=1}^{T}$ prend un état fondamental et après une durée polynomiale dans la taille du système, *i.e.* $T \in \mathrm{poly}(n)$, l'amener dans un état ρ_T dont la projection sur l'état fondamental soit différente de l'état initial. L'idée est de modéliser une situation où deux états fondamentaux distincts $|\Omega\rangle$ et $|\Omega'\rangle$ aboutissent au même état bruité ρ^*, comme sur la figure 6.5. On va donc s'intéresser à une séquence d'erreurs logique $\{\mathcal{E}_k\}_{k=1}^{T}$ qui à partir d'un état fondamental $|\Omega\rangle$ aboutit à un état final $\rho_T = \mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_T \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]$ qui soit (i) proche d'un état fondamental et (ii) dont la projection sur l'espace fondamental $\Omega' \equiv P_C \rho_T P_C$ soit distincte de l'état initial. Formellement, on a donc

$$\mathrm{Tr}\left[P_C \,\mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_T \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]\right] \geq 1 - \epsilon_1 \tag{6.18}$$

$$\mathrm{Tr}\left[\Omega \mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_T \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]\right] \leq \epsilon_2 \tag{6.19}$$

où apparaît un système ancillaire $\mathcal{A}$. Celui-ci sert à modéliser les effets non-Markoviens de l'environnement. En effet, l'erreur $\mathcal{E}_{t+1}$ peut dépendre non-seulement de l'état ρ_t, mais aussi de l'histoire de l'évolution de la mémoire. Ce point technique sera important pour le résultat démontré dans notre article.

Notons que ϵ_1 et ϵ_2 n'ont pas besoin d'être petits. Pour que la séquence d'erreurs logique atteigne un état $\rho_T \equiv \mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_T \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]$ dont la projection sur l'espace fondamental $\Omega' = P_C \rho_T P_C$ soit différente de l'état fondamental initial Ω, il suffit que $\epsilon_1 + \epsilon_2 < 1$ car

$$(1 - \epsilon_1)\,\mathrm{Tr}\left[\Omega\Omega'\right] \leq \mathrm{Tr}\left[\Omega P_C \rho_T P_C\right] \leq \epsilon_2 \tag{6.20}$$

Étant donnée une séquence d'erreurs logique $\mathcal{E} \equiv \{\mathcal{E}_k\}_{k=1}^{T}$ et un état fondamental initial $|\Omega\rangle$, on définit la barrière d'énergie, représentée sur la figure 6.5, comme l'énergie maximale atteinte par les états intermédiaires au-dessus de l'énergie E_0 du fondamental

$$\Delta^*(\mathcal{E}, \Omega) = \max_{k \in [\![1;T]\!]} \mathrm{Tr}\left[H \,\mathrm{Tr}_{\mathcal{A}}\left[\mathcal{E}_k \ldots \mathcal{E}_1(\Omega \otimes \rho_{\mathcal{A}})\right]\right] - E_0. \tag{6.21}$$

La barrière d'énergie du hamiltonien H est alors la plus petite barrière d'énergie pour toutes les séquences d'erreurs logiques et tous les états fondamentaux initiaux $|\Omega\rangle$

$$\Delta^*(H) = \min_{\mathcal{E}, \Omega} \Delta^*(\mathcal{E}, \Omega) \tag{6.22}$$

6.4.2.3 Critère de stabilité

La barrière d'énergie permet de formuler un critère suffisant pour la stabilité thermique, celui d'avoir une barrière d'énergie qui grandit avec la taille du système. En effet, si $\Delta(H) \sim L^{\alpha}$, avec $\alpha > 0$, cela veut dire que toute séquence d'erreurs logique fait apparaître un état intermédiaire de très grand énergie, ce qui est peu probable.

Au contraire, si la barrière d'énergie de la mémoire est une constante, il est possible pour l'environnement de passer d'un fondamental à un autre fondamental orthogonal grâce à une séquence d'erreur locale. Ainsi, une accumulation d'erreur locale peut affecter l'état global de la mémoire, ce qui laisse penser qu'elle est instable.

Des résultats d'instabilité qui montrent que tout code topologique 2D ne présente qu'une barrière d'énergie indépendante de la taille du système ont été obtenus par Bravyi et Terhal [106] en 2009 pour les codes stabilisateurs. Mon directeur, David Poulin, et moi-même avons étendu ce résultat en 2013 à tous les codes CPC topologiques avec cohérence locale [4]. Tous ces résultats reposent sur l'existence d'opérateurs logiques supportés sur une bande 1D, qualifiés d'opérateurs logiques en ruban. Ces contraintes sur l'existence d'une mémoire auto-correctrice en 2D seront l'objet du chapitre suivant.

Chapitre 7

Instabilité thermique des mémoires topologiques 2D

Dans ce chapitre, nous verrons que les mémoires topologiques 2D dont on sait prouver que leur spectre est stable présentent une instabilité thermique. Cette instabilité repose sur l'existence d'opérateurs logiques non-triviaux en ruban, *i.e.* dont le support est une région 1D. Nous montrerons tout d'abord pourquoi l'existence de tels opérateurs est source d'instabilité thermique, en se basant sur l'exemple du code torique. Comme démontré en annexe D, ces opérateurs ruban existent non seulement pour les codes stabilisateurs, comme démontré par Bravyi et Terhal [106], mais aussi pour tout code à projecteurs commutatifs (CPC) [107, 108]. Ceci nous amènera à notre contribution majeure à ce domaine, l'article

> Local topological order inhibits thermal stability in 2D
> Olivier Landon-Cardinal et David Poulin.
> *Physical Review Letters*, **110**, 090502 (2013)

qui montre l'existence d'une séquence d'erreurs logique dont la barrière d'énergie est indépendante de la taille du système pour tout hamiltonien CPC topologiquement ordonné qui respecte la condition de cohérence locale. Ainsi, pour tous les modèle topologiques 2D dont on sait prouver la stabilité du spectre, nous montrons qu'ils ne sont pas protégés par une barrière d'énergie grandissant avec la taille du système.

7.1 Opérateurs en ruban et instabilité thermique

7.1.1 Instabilité du code torique

Comme nous l'avons vu en 5.3.2.6, le code torique admet des opérateurs logiques qui sont des opérateurs de Pauli supportés sur des boucles non-triviales. Par exemple, $\bar{Z}_1$ est un produit tensoriel d'erreur suivant une boucle du réseau qui fait le tour du tore, cf. Fig. 5.10. Les générateurs logiques $\bar{X}_1$, $\bar{X}_2$, $\bar{Z}_1$ et $\bar{Z}_2$ sont des opérateurs non-locaux qui agissent sur un nombre macroscopique de qubits. Toutefois, ils peuvent être imités par une séquence d'erreurs locales car ils sont des produits tensoriels d'opérateurs à un corps. Nous allons maintenant voir ce phénomène de deux points de vue différents : d'abord en tant que séquence d'erreurs locale puis en terme d'anyons qui se propagent et percolent.

7.1.1.1 Séquence d'erreurs

Une séquence d'erreurs locales peut imiter un opérateur logique en appliquant successivement des erreurs sur un qubit suivant une boucle non-triviale. Par exemple, numérotons les qubits du support de $\bar{Z}_1$ de 1 à L. On remarque alors que $\bar{Z}_1 = \prod_{k=1}^{L} \mathcal{E}_k$ où $\mathcal{E}_k$ applique une erreur Z sur le qubit k

$$\mathcal{E}_k(\rho) = Z_k \rho Z_k^\dagger \tag{7.1}$$

Or, tous les états intermédiaires ne sont que des premiers états excités, dont l'énergie est la valeur du gap, ici 2. En effet, seuls les opérateurs étoiles situés aux extrémités de la chaîne qui supporte l'opérateur partiel $\bigotimes_{k=1}^{\ell} Z_k$ anti-commutent avec cet opérateur. Pour tous les opérateurs étoiles au milieu de la chaîne, il est impossible de distinguer l'opérateur partiel de l'opérateur logique $\bar{Z}_1$. Ainsi, la barrière d'énergie de cette séquence d'erreurs logique est très faible et surtout indépendante de la taille du réseau. De la même façon, il est possible de construire une séquence d'erreurs à partir de n'importe lequel des autres générateurs logiques $\bar{X}_1$, $\bar{X}_2$, ou $\bar{Z}_2$, c'est-à-dire suivant n'importe quelle boucle non-triviale.

Évidemment, l'environnement qui appliquerait successivement des erreurs Z sur des qubits adjacents suivant une boucle non-triviale n'est pas naturel. On parle alors de modèle adversaire. Toutefois, même pour un modèle plus réaliste où les erreurs apparaissent au hasard sur n'importe

quel qubit, une erreur logique finira par apparaître. Pire, cela se fait alors dans un temps indépendant de la taille du système [109]. Nous allons voir que le plus simple pour comprendre ce phénomène est de penser aux erreurs en terme de création et de propagation d'anyons.

7.1.1.2 Percolation d'excitations thermiques

L'application d'une erreur Z sur un qubit sur un état fondamental $|\Omega\rangle$ correspond à créer une paire d'anyons à partir du vide, ce qui demande une énergie de valeur 2. Or, une fois créés, ces anyons peuvent se propager sans coût en énergie et se retrouver arbitrairement loin l'un de l'autre. Si leur ligne de vie constitue une boucle non-triviale lorsqu'ils fusionnent pour redonner le vide, ils auront effectué une action non-triviale dans l'espace fondamental. En plus de se déplacer, ces anyons peuvent aussi fusionner avec des anyons de même type provenant d'une autre paire (et il s'agit d'anyons abéliens pour lesquels la fusion est déterministe). Il suffit alors que des anyons percolent suivant une boucle non-triviale afin qu'une erreur logique apparaisse. Cela demande une durée de temps indépendante de la taille du système [109].

7.1.2 Intuition : propagation d'anyons et opérateurs ruban

Pour des systèmes topologiques autres que le code torique, la propagation d'anyons suivant une boucle non-triviale devrait avoir une action non-triviale dans l'espace fondamental. Ainsi, ces modèles devraient aussi être instables thermiquement. Toutefois, il n'est pas évident qu'un hamiltonien CPC topologiquement ordonné, défini à partir d'une famille de projecteurs élémentaires, admette un modèle anyonique pour ses excitations de basse énergie. C'est bien évidemment le cas si le hamiltonien est construit à partir d'un modèle anyonique, comme pour les modèles habituels d'ordre topologique [80, 103, 104], mais cela n'est pas évident en général. L'intuition est toutefois qu'il existe bien un modèle anyonique pour tout hamiltonien CPC topologiquement ordonné. Notre article vient d'ailleurs conforter cette impression et peut être vu comme un premier pas vers la démonstration rigoureuse de ce résultat.

Ainsi, afin de formaliser que la barrière d'énergie des hamiltoniens CPC 2D est constante, nous utiliserons l'approche qui consiste à exhiber une séquence d'erreurs logique. Dans le cas du code torique, le raisonnement reposait sur deux ingrédients : l'existence d'un opérateur logique supporté sur une boucle et le fait que cet opérateur soit un produit tensoriel. Plus généralement, nous

parlerons d'opérateur ruban pour un opérateur logique non-trivial supporté sur une boucle non-triviale avec une largeur indépendante de la taille du système. Nous allons voir dans la prochaine section qu'un tel opérateur logique en ruban existe pour tout code stabilisateur topologique, et plus généralement pour tout hamiltonien CPC topologique.

La première étape afin de montrer l'existence de tels opérateurs logiques en ruban est de montrer qu'il existe un ruban non-corrigible. Il s'agit d'un travail technique que nous avons préféré placer dans l'annexe D car il n'est pas strictement nécessaire afin de comprendre notre article. Afin de replacer notre travail dans le contexte de la littérature, il est toutefois important de rappeler les articles de la littérature permettant d'atteindre ce résultat. Le premier[1] résultat général sur l'existence d'opérateurs ruban pour les codes stabilisateurs 2D est du à Bravyi et Terhal [106]. Ils ont montré comment déformer un opérateur logique afin qu'il n'agisse que sur un ruban grâce à un résultat clé, le lemme de de nettoyage (cf. section D.1.1). Cet outil technique a été étendu au cas plus général des codes CPC par Bravyi, Poulin et Terhal [107] pour devenir le lemme de désintrication holographique (cf. section D.2.2). Il utilise crucialement le fait que les projecteurs élémentaires qui définissent le code commutent. Ensuite, Haah et Preskill ont utilisé cet outil afin de prouver l'existence d'opérateurs ruban pour les codes CPC [108].

L'existence de tels opérateur ruban suffit à prouver que la barrière d'énergie est indépendante de la taille du système dans le cas particulier des codes stabilisateurs. En effet, les opérateurs logiques pour ces codes sont des opérateurs de Pauli qui sont des produits tensoriels de transformations unitaires à un corps. Par exemple, pour le code torique, un opérateur logique est $\bar{Z}_1 = \bigotimes_{k=1}^{\ell} Z_k$ où k indice les qubits le long du tore. Il est représenté sur la Fig. 7.1 a). L'environnement peut donc construire séquentiellement l'opérateur en appliquant un opérateur de type $\bigotimes_{k=1}^{m} Z_k$, i.e, une chaîne ouverte sur m sites, représentée sur la Fig. 7.1 b). Dans ce cas, seuls les opérateurs étoiles situés aux extrémités de la chaîne seront violés, i.e., l'état sera un état propre -1 de ces opérateurs. Cela correspond à une quantité constante d'énergie au-dessus du niveau fondamental. Ce raisonnement est très général pour tout opérateur logique en ruban qui est de plus un opérateur produit, ce qui est le cas pour tout code stabilisateur.

Pour des codes non stabilisateurs, on ne pense pas qu'il existe toujours un opérateur logique qui soit un opérateur produit. Ainsi, l'application séquentielle d'un opérateur logique n'est pas évidente, tel que remarqué dans [108]. La contribution technique de notre article est d'exhiber une séquence d'erreurs logique qui applique séquentiellement l'opérateur ruban pour tout code CPC.

1. Un résultat similaire a été obtenu indépendamment pas Kay et Colbeck [110].

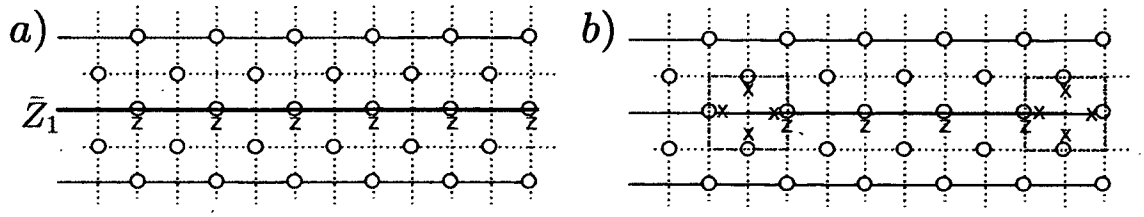


FIGURE 7.1 Application partielle d'un opérateur logique pour un code stabilisateur. Dans un code stabilisateur, les opérateurs logiques agissent sur une ligne et sont des opérateurs de Pauli, en particulier des opérateurs produits. Par exemple, l'opérateur logique $\bar{Z}_1 = \bigotimes_{k=1}^{\ell} Z_k$ est représenté en a). Il est possible de construire séquentiellement cet opérateur en appliquant progressivement des opérateurs Z selon une ligne, tel que l'opérateur intermédiaire est une chaîne partielle, de la forme $\bigotimes_{k=1}^{m} Z_k$ avec $m < \ell$, comme représenté en b). Dans ce cas, le système possède une énergie qui n'est qu'une constante au-dessus de l'énergie fondamentale puisque seuls les opérateurs étoiles situées aux extrémités de la chaîne sont violés, i.e., l'état du système est un état propre -1 de ces opérateurs.

7.2 Opérateur ruban

Dans le cas du code torique, certains opérateurs logiques non-triviaux étaient supportés sur une boucle non-triviale de particules. Il s'agit d'un cas extrême d'opérateur ruban. En effet, dans d'autres modèles topologiques où la longueur de corrélation est non-nulle, les opérateurs logiques seront supportés sur une bande non-triviale, *i.e.* un boucle non-triviale élargie. Formellement, on définit un ruban $R_{\mathcal{B}}$ de largeur b comme l'ensemble des sites situés à une distance inférieure à b de la boucle $\mathcal{B}$

$$R_{\mathcal{B}}(b) = \{s \in \Lambda \quad \text{dist}(s, \mathcal{B}) \leq b\}\,. \tag{7.2}$$

Un opérateur ruban est un opérateur logique non-triviale supporté sur un ruban $R_{\mathcal{B}}(b)$. On verra que pour un code CPC, b est une constante indépendant de la taille du système, qui ne dépend que la portée w des générateurs définissant le code. De plus, pour un code topologique, $\mathcal{B}$ est une boucle non-triviale.

7.2.1 Existence d'un ruban non-corrigible

L'annexe D reprend les raisonnements, provenant essentiellement des articles [106, 107, 108], afin de montrer qu'il existe un ruban non-corrigible pour un code CPC. Intuitivement, cela signifie qu'il existe un opérateur logique dont le support est inclus dans cette région.

Nous allons voir que dans le cas d'un code stabilisateur, l'opérateur logique non-trivial est toujours un opérateur produit et qu'il est donc clair d'imaginer comment construire une séquence d'erreurs logique à partir d'un tel opérateur. Par contre, pour un code CPC général, on a des bonnes raisons de croire que l'opérateur logique non-trivial n'est pas un opérateur produit et son application séquentielle n'est pas évidente.

7.2.2 Opérateur ruban dans un code stabilisateur

Pour un code stabilisateur, les opérateurs logiques sont des opérateurs qui commutent avec le groupe stabilisateur. Ils sont non-triviaux si, de plus, ils ne font pas partie du groupe stabilisateur. Or, le groupe logique est toujours généré par des opérateurs de Pauli. Ainsi, on peut toujours choisir l'opérateur logique non-trivial pour que ce soit un opérateur de Pauli. De plus, le raisonnement fourni en annexe D.1 montre que son support est un ruban.

Puisqu'il s'agit d'un opérateur produit, il est possible de construire une séquence d'erreurs logique qui applique progressivement cet opérateur, comme dans l'exemple du code torique. De plus, cette séquence d'erreurs logique crée des états intermédiaires dont l'énergie au-dessus de l'énergie fondamental n'est qu'une constante indépendante de la taille du système. En effet, seuls les contraintes correspondant aux générateurs dont le support intersecte les extrémités de l'opérateur partiellement construit ne sont pas respectées. La barrière d'énergie d'un code stabilisateur est indépendante de la taille du système.

Nous allons maintenant voir qu'il existe des opérateurs logiques non-triviaux en ruban pour tout code CPC. Toutefois, un tel opérateur n'est plus en général un opérateur produit. Il faudra donc trouver une façon astucieuse de construire une séquence d'erreurs logique.

7.2.3 Opérateur ruban dans un code CPC

Soit M un ruban non-corrigible dont l'existence est prouvée en D.2. Selon la condition de Knill-Laflamme [89], il existe donc un opérateur $\mathcal{O}_M$ supporté sur M tel que

$$P_C \mathcal{O}_M P_C \not\propto P_C \tag{7.3}$$

Toutefois, l'opérateur $P_C \mathcal{O}_M P_C$ présente deux problèmes : (i) rien ne garantit qu'il soit une transformation unitaire et (ii) il agit a priori sur l'ensemble des particules du réseau parce que P_C a une action globale. Le problème (ii) peut se résoudre facilement en se rappelant que le projecteur $P_C = \prod_X P_X$ est un produit de projecteurs élémentaires locaux. Ainsi, il suffit de conjuguer $\mathcal{O}_M$ par $P_{\bar{M}} = \prod_{X \cap M \neq \emptyset} P_X$ pour obtenir un opérateur qui agit à l'intérieur du code. Notons que cette procédure peut en fait être appliquée à n'importe quel opérateur local de support A afin de produire un opérateur qui agit à l'intérieur du code et de support légèrement plus grand $\bar{A}$ qui est la réunion des supports de tous les projecteurs élémentaires qui intersectent la région A.

Attaquons-nous maintenant au problème (i) : on va essayer de construire une transformation unitaire à partir de $\mathcal{O}_M$. Pour ce faire, on décompose $\mathcal{O}_M$ sur la base des opérateurs de Pauli $\mathcal{O}_M = \sum_i c_i O_i$. Il existe au moins un opérateur de Pauli O_{i_0} dans cette décomposition qui agit non-trivialement sur l'espace code, i.e, $P_C O_{i_0} P_C \not\propto P_C$. Comme remarqué plus haut, il suffit de conjuguer O_{i_0} par $P_{\bar{M}}$ pour obtenir un opérateur hermitien logique supporté sur $\bar{M}$

$$E = P_{\bar{M}} O_{i_o} P_{\bar{M}} \tag{7.4}$$

qui est non-trivial car E et O_{i_0} ont la même action à l'intérieur de l'espace fondamental car $P_C E P_C = P_C O_{i_o} P_C$. Ainsi, il a au moins deux espaces propres associées à des valeurs propres distinctes $\lambda_1 \neq \lambda_2$ à l'intérieur du code.

Il suffit alors de prendre son exponentielle pour obtenir un opérateur unitaire

$$U_\tau = \exp(-i\tau E) \tag{7.5}$$

De plus, U_τ est un opérateur logique, et il est non-trivial si τ n'est pas un multiple de $\frac{2\pi}{\lambda_1 - \lambda_2}$.

Contrairement au cas des codes stabilisateurs, rien ne garantit que U_τ soit un opérateur produit. Or, le fait que l'opérateur logique non-trivial soit un opérateur produit était crucial dans le résultat pour les stabilisateurs. En effet, il est clair qu'un opérateur produit peut être construit séquentiellement par l'environnement. Ainsi, pour les codes CPC, la question technique était de savoir s'il était possible de construire séquentiellement un opérateur qui n'est pas un produit tensoriel.

Haah et Preskill ont proposé une première approche [108], basée sur le fait que l'opérateur logique non-trivial ne crée que peu d'intrication. Nous allons présenter cette tentative et expliquer brièvement pourquoi elle échoue en 7.2.4. Ensuite, nous présenterons notre approche en 7.3.

7.2.4 Instabilité thermique des mémoires CPC 2D ?

Nous allons passer en revue la première approche tentée par Haah et Preskill [108]. Elle repose sur le fait que l'opérateur logique non-trivial en ruban ne crée que peu d'intrication. Techniquement, il s'agit d'un *matrix product operator* ou MPO, l'équivalent opérateur d'un MPS. Nous définirons d'abord ce qu'est un MPO, avant de montrer que l'opérateur hermitien E crée peu d'intrication et enfin prouver que U_τ peut être approximé par un MPO. Finalement, nous reproduirons le raisonnement proposé par Haah et Preskill qui conclut que l'environnement ne peut pas appliquer ce MPO séquentiellement.

Afin de définir un MPO, il suffit d'utiliser la décomposition de Schmidt, introduite en annexe A.1.1 au niveau des opérateurs. Ainsi, tout opérateur agissant sur deux régions AB peut s'écrire

$$O_{AB} = \sum_{i=1}^{\chi} \lambda_i O_A^i \otimes O_B^i \tag{7.6}$$

où le nombre de termes est le rang de Schmidt χ de l'opérateur O_{AB} par rapport à la partition AB. Un opérateur qui crée peu d'intrication est un opérateur dont le rang de Schmidt est borné par une constante, indépendante de la taille des régions A et B.

Revenons donc à l'opérateur hermitien E qui agit sur la bande étendue $\bar{M}$. Nous allons voir qu'il crée peu d'intrication. En fait, nous allons montrer que $P_{\bar{M}}$ possède cette propriété et que E en hérite.

On peut découper $\bar{M}$ en ℓ régions de largeur w (où w est le diamètre maximal du support des projecteurs élémentaires), numérotées de $k = 1$ à $k = \ell$, comme sur la figure 7.2. En raison de leur largeur, tout projecteur élémentaire agit au plus sur deux régions consécutives k et $k + 1$. On définit alors un nouveau projecteur, noté $P_{k,k+1}$, qui est le produit de tous les projecteurs dont le support intersectent k et $k + 1$ et aussi des projecteurs élémentaires entièrement contenus [2] dans la région k

$$P_{k,k+1} = \prod_{\substack{X \subset \{k,k+1\} \\ X \not\subset \{k+1\}}} P_X \tag{7.7}$$

2. Ce choix est arbitraire, on pourrait aussi mettre ces projecteurs dans $P_{k-1,k}$.

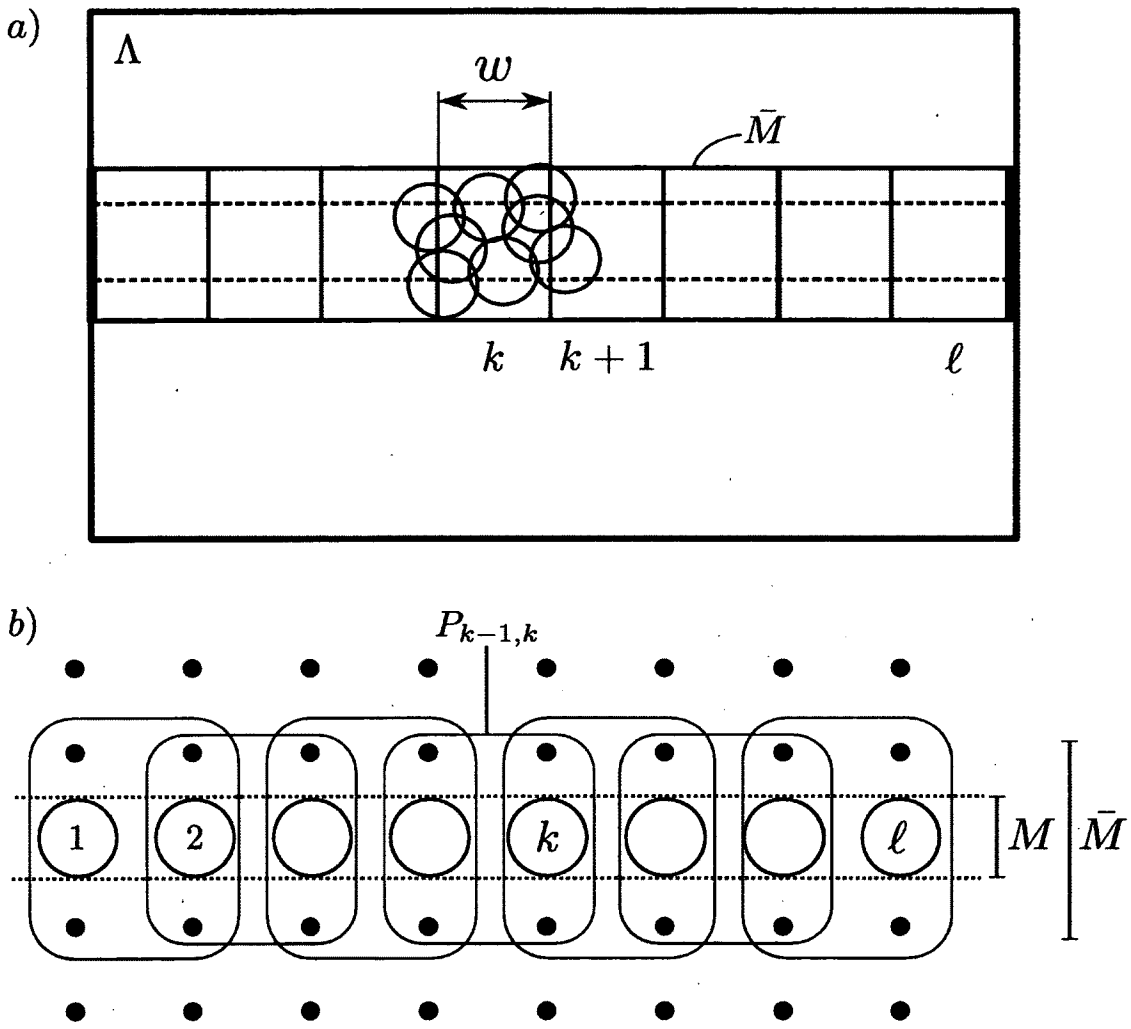


FIGURE 7.2 Décomposition du ruban non-corrigible M. a) Découpage de la bande M en ℓ régions de largeur w. Un projecteur élémentaire agit au plus deux régions consécutives k et $k+1$. b) Zoom sur le ruban et définition des projecteurs $P_{k,k+1}$.

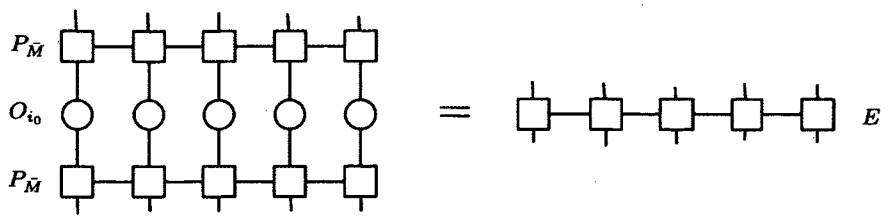


FIGURE 7.3 Caractère MPO de l'opérateur E

On construit a ainsi écrit le projecteur $P_{\bar{M}}$ sous la forme d'une interaction premier voisin

$$P_{\bar{M}} = \prod_{k=1}^{\ell-1} P_{k,k+1} \tag{7.8}$$

Or, en utilisant le fait que ces projecteurs commutent, on décomposer chaque paire de projecteur $[P_{k-1,k}; P_{k,k+1}] = 0$ sous la forme

$$P_{k-1,k}P_{k,k+1} = \sum_{\alpha_k} P_{k-1,k_L^{\alpha_k}} P_{k_R^{\alpha_k},k+1} \tag{7.9}$$

Cette décomposition est expliquée en détail en annexe dans la section D.2.1. En appliquant peut appliquer la décomposition 7.9 de façon répétée, on obtient

$$P_{\bar{M}} = \sum_{\alpha_2 \ldots \alpha_{\ell-1}} P_{12_L^{\alpha_2}} \otimes P_{2_R^{\alpha_2}3_L^{\alpha_3}} \otimes \ldots P_{\ell-1_R^{\alpha_{\ell-1}}\ell} \tag{7.10}$$

qui est un *matrix product operator* ou MPO, l'équivalent opérateur d'un MPS. Sa dimension de lien D est indépendante de la taille du système puisqu'elle ne dépend que de la commutation de projecteurs locaux. Rappelons que E est obtenu en conjuguant l'opérateur produit O_{i_o} par le MPO $P_{\bar{M}}$, cf. éq. (7.4) : il hérite donc lui aussi d'une structure MPO, comme représenté graphiquement sur la figure 7.3.

Il suffit maintenant de noter que U_τ, défini par l'éq. (7.5), peut être développer en série entière

$$U_\tau = \sum_{k=0}^{+\infty} \frac{(-i\tau)^k}{k!} E^k \tag{7.11}$$

et de remarquer que E^k est un MPO dont la dimension de lien est $D(E^k) \leq (D(E))^k$. Ainsi, pour tout k constant, la dimension de lien grandit mais reste constante. De plus, en tronquant la

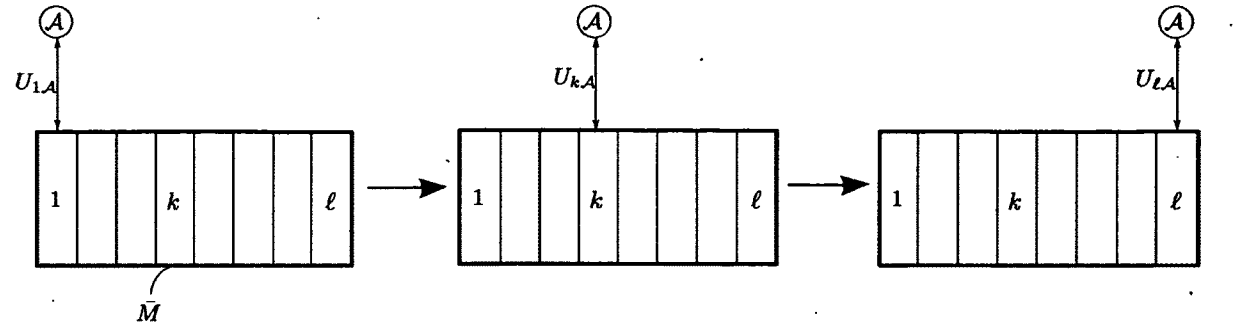


FIGURE 7.4 Application séquentielle de l'opérateur logique non-trivial.

somme à K, on obtient une très bonne approximation $U_\tau^{(K)} \equiv \sum_{k=0}^{K} \frac{(-i\tau)^k}{k!} E^k$ de U_τ en norme d'opérateur.

$$\left\| U_\tau - U_\tau^{(K)} \right\| \leq \sum_{k>K} \frac{\tau^k}{k!} \|E\|^k \tag{7.12}$$

Ainsi, on a approximé une transformation unitaire logique non-triviale par un MPO de dimension de lien indépendante de la taille du réseau.

L'espoir de Haah et Preskill était alors d'appliquer ce MPO séquentiellement en imaginant qu'un système ancillaire $\mathcal{A}$ de dimension constante interagisse successivement avec chacune des régions de la bande $\bar{M}$, comme sur la figure 7.4. Ainsi, on aurait $U_\tau^{(K)} \otimes \mathbb{I}_\mathcal{A} = \prod_{k=1}^{\ell} U_{k\mathcal{A}}$ et à la fin de la procédure , le système ancillaire est désintriqué de la région $\bar{M}$. Une telle procédure représenterait l'action locale de l'environnement et le système ancillaire correspondrait à des effets non-Markoviens de l'environnement.

Or, un résultat due à Lamata et al. [111] montre qu'il est impossible d'appliquer une transformation unitaire qui n'est pas un opérateur produit de cette façon. L'idée est de remarquer que si la première transformation unitaire $U_{1\mathcal{A}}$ crée de l'intrication, alors $\prod_{k=2}^{\ell} U_{k\mathcal{A}}$ est une transformation unitaire qui doit envoyer un espace vectoriel de dimension $2 \times d^{\ell-1}$ vers un espace de dimension $d^{\ell-1}$ où d est la dimension de l'espace de Hilbert de chaque région. Haah et Preskill concluent donc que cette technique ne peut pas fonctionner.

L'approche de notre article contourne ce résultat d'impossibilité sur l'application séquentielle d'une transformation unitaire car elle va faire apparaître des mesures projectives. Toutefois, afin que ces mesures donnent le résultat désiré, nous devrons imposer une condition supplémentaire aux codes CPC. Il s'agit de la condition de cohérence locale, celle-là même qui a été rajoutée à l'ordre topologique afin de garantir la stabilité du spectre.

7.3 Article : « Local topological order inhibits thermal stability in 2D »

Nous allons maintenant présenter la contribution majeure de la deuxième partie de cette thèse

Local topological order inhibits thermal stability in 2D
Olivier Landon-Cardinal et David Poulin.
Physical Review Letters, **110**, 090502 (2013)

7.3.1 Genèse et contribution

La genèse de cet article a été longue. L'idée originale a été suggérée par mon directeur, David Poulin, au début de ma thèse : il s'agit d'appliquer l'opérateur logique non-trivial séquentiellement sur chaque région de la bande $\bar{M}$ en utilisant des transformations unitaires aléatoires et des projections. Plus précisément, pour chaque région k, une transformation unitaire choisie au hasard est appliquée sur chaque région k et la mesure $P_{k-1,k}$ est effectuée. Ce processus est répété jusqu'à ce que la contrainte locale $P_{k-1,k}$ soit satisfaite et on passe alors à la région $k+1$.

Après plusieurs tentatives infructueuses, nous avons réalisé l'existence de cul-de-sac pour cette procédure, *i.e.* de cas où aucune transformation unitaire sur la région k ne pouvait satisfaire la contrainte locale $P_{k-1,k}$. Nos tentatives afin de contourner ce problème sont restées vaines. Finalement, la condition de cohérence locale, découlant du travail sur la stabilité spectrale est venue fournir la clé afin que notre procédure fonctionne à tout coup.

Ma contribution a été principalement de formaliser les raisonnements et de démontrer rigoureusement les résultats mathématiques nécessaires au résultat. J'ai rédigé les premières versions de l'article avec l'aide attentive de mon directeur.

7.3.2 Article

Local topological order inhibits thermal stability in 2D

Olivier Landon-Cardinal* and David Poulin†
Département de Physique, Université de Sherbrooke, Québec, J1K 2R1, Canada

We study the robustness of quantum information stored in the degenerate ground space of a local, frustration-free Hamiltonian with commuting terms on a 2D spin lattice. On one hand, a macroscopic energy barrier separating the distinct ground states under local transformations would protect the information from thermal fluctuations. On the other hand, local topological order would shield the ground space from static perturbations. Here we demonstrate that local topological order implies a constant energy barrier, thus inhibiting thermal stability.

PACS numbers: 03.67.Pp, 03.65.Ud, 03.67.Ac

A self-correcting quantum memory [1] is a physical system whose quantum state can be preserved over a long period of time *without* the need for any external intervention. The archetypical self-correcting classical memory is the two-dimensional (2D) Ising ferromagnet. The ground state of this system is two-fold degenerate, so it can store one bit of information. If the memory is put into contact with a heat bath after being initialized in one of these ground states, thermal fluctuations will lead to the creation of small error droplets of inverted spins. The boundary of such droplets are domain walls, i.e., one-dimensional excitations whose energy is proportional to the droplet perimeter. If the temperature is below the critical Curie temperature, the Boltzman factor will prevent the creation of macroscopic error droplets. Thus, when the system is cooled down (either physically or algorithmically) after some macroscopic storage time, it will very likely return to its original ground state: the memory is thermally stable.

This behaviour contrasts with the 1D Ising ferromagnet whose domain walls are point-like excitations. The creation of a domain wall costs some constant amount of energy (the gap), but once created they can freely diffuse on the chain at no extra energy cost. As a consequence, arbitrarily large error droplets can form at a constant energy cost, so this 1D memory is thermally unstable.

While the 2D Ising ferromagnet features thermal stability, it is vulnerable to static, local perturbations. Indeed, an arbitrarily weak magnetic field breaks the ground state degeneracy and favours one ground state over the other. When this perturbed system is subject to thermal fluctuations, the bulk contribution of the magnetic field overwhelms the boundary tension of the domain wall, so once error droplets reach a critical size, they rapidly expand to corrupt the memory. This type of instability plagues any system with a local order parameter, so they cannot be robust quantum memories. Indeed, distinct ground states give different values of this order parameter, so a local field coupling to the order parameter lifts degeneracy.

In 2D and higher, there exists quantum systems with no local order parameter and whose spectrum is stable under weak, local perturbations. These systems have a degenerate ground state separated from the other energy levels by a constant energy gap, and perturbations only alter these features by an exponentially vanishing amount as a function of the system size. Kitaev's toric code [2], a $\mathbb{Z}_2$ spin liquid, is the best known example of this type. However, excitations in Kitaev's code are point-like objects—as for the 1D Ising model—so it does not offer a macroscopic energy barrier protection to thermal fluctuations [1, 3–6].

In this Letter, we study the possibility of combining the thermal stability of the 2D Ising model with the spectral stability of Kitaev's code to obtain a robust quantum memory in 2D. We consider d-level spins located at the vertices V of a 2D lattice $\Lambda = (V, E)$, with Hamiltonian

$$H = -\sum_{X \subset V} P_X, \text{ with } [P_X, P_Y] = 0 \text{ and } \|P_X\| \leq 1. \quad (1)$$

We denote the number of spins $N \equiv |V|$. The term P_X is supported on the subset X of the spins, i.e., it acts trivially on the complement $\overline{X} = V - X$ of X. The Hamiltonian is local in the sense that $P_X = 0$ whenever X has radius larger than some constant w. Since we are only interested in the ground state and scaling of the energy gap, we can assume without loss of generality that each P_X is a projector. We also assume that H is frustration-free, meaning that the ground states minimize the energy of each term of the Hamiltonian separately, i.e., $P_X|\psi_0\rangle = |\psi_0\rangle$. Then, the ground space $\mathcal{C}$ is the image of the *code projector* $P = \prod_X P_X$ (henceforth, the $P_X = 0$ are not included in such products).

This family of lattice models, called *local commuting projector code* (LCPC) includes most known models of topological order including quantum doubles [2], Levin-Wen [7], and Turaev-Viro [8] models. It has been proved that LCPC have a stable spectrum [9–11, 29] if they obey the following *local topological order* condition.

Definition 1 (Local topological order). For any topologically trivial region A, let $P_A = \prod_{X: X \cap A \neq \emptyset} P_X$ be the product of projectors that intersect region A. For a system with *local topological order*, $\rho^A \equiv \mathrm{Tr}_{\bar{A}} P$ has the same

kernel as $\rho_A^{\text{loc}} \equiv \text{Tr}_{\bar{A}} P_A$ and moreover $\text{Tr}_{\bar{A}} \psi \propto \rho^A$ for any ground state $|\psi\rangle$.

Our main result is that any system with local topological order has only a constant energy barrier between ground states. This result can be understood intuitively when the low-energy excitations of the system are localized deconfined anyons. In that case, it is possible to modify the ground state by creating an anyon pair, dragging one of them along a topologically non-trivial loop, and annihilating it with its partner. This process clearly requires only a constant amount of energy—the mass gap—since the anyons are deconfined. The obstacle in formalizing this heuristic picture is to prove that the low-energy excitations are indeed localized deconfined anyons. Toy models of topological order [2, 7, 8] usually take such an anyon model as a starting point to construct a local Hamiltonian. Here, we take the opposite path: our starting point is any Hamiltonian with local topological order and we want to characterize its low-energy excitations. This is an active area of research [12]. It is unclear whether excitations are always point-like in these models and, as noted by Haah and Preskill [13], whether creating and moving them can always be realized by a sequence of local unitary transformations. Thus, our result is a step towards a general understanding of low-energy excitations in models with local topological order, which should be of independent interest. Indeed, our proof will essentially confirm the above heuristic picture, but circumvents the delicate unitarity issue.

Background— Characterizing the thermal stability of a memory requires detailed knowledge of its thermalization process. Since we seek to address a broad class of systems, our analysis cannot be model specific. We thus retain only two essential features common to all thermalization processes: (i) the bath interacts locally with the system, and (ii) high-energy states are penalized. As we now explain, we can combine these features to obtain a sufficient condition for thermal stability.

We will say that a memory with Hamiltonian Eq. (1) has an energy barrier at most Δ if there exists a ground state ψ_0 and a sequence of $T \in \text{poly}(N)$ CPTP maps $\mathcal{E}_k$, each acting locally on the system and a finite-dimensional ancilla A, such that (i) starting from ψ_0, the sequence returns the system to the ground space, (ii) in a state that differs from the initial state ψ_0, and (iii) the energy of any intermediate state is at most Δ above the ground state energy E_0. More formally, these conditions are

$$\text{Tr}\left[P \cdot \mathcal{E}_T \dots \mathcal{E}_2 \mathcal{E}_1(\psi_0 \otimes \rho_A)\right] \geq \frac{2}{3} \quad \text{(2a)}$$

$$\text{Tr}\left[\psi_0 \cdot \mathcal{E}_T \dots \mathcal{E}_2 \mathcal{E}_1(\psi_0 \otimes \rho_A)\right] \leq \frac{1}{3} \quad \text{(2b)}$$

$$\max_{k \in 1,2,\dots,T} \text{Tr}\left[H \cdot \mathcal{E}_k \dots \mathcal{E}_2 \mathcal{E}_1(\psi_0 \otimes \rho_A)\right] - E_0 = \Delta \quad \text{(2c)}$$

where P is the code projector and the factors $\frac{2}{3}$ and $\frac{1}{3}$ are arbitrarily chosen constants. The additional ancillary system, initially in state ρ_A, is used to model some finite non-Markovian effects of the bath, so each map $\mathcal{E}_k$ has complete access to it. The energy barrier of a memory is taken to be the smallest value of Δ over all such sequences of maps. If a memory has a macroscopic energy barrier $\Delta \geq N^\alpha$ for some constant $\alpha > 0$, then any short sequence of local transformations that returns the system to an altered ground state must visit a high energy state, and is therefore thermally stable. Our main result is obtained by exhibiting a sequence of maps $\mathcal{E}_k$ with an energy barrier Δ that is a constant, independent of the system size N for any LCPC. In addition, we show that the length T of this sequence is proportional to the linear size of the lattice when the system has local topological order.

We call *logical operator* an operator L that maps the ground space to itself, i.e. $[L, P] = 0$. In particular, we are interested in logical operators that act non-trivially on the code space, i.e., $LP \neq P$, as they can alter the encoded information. In a series of paper [13–16], it was shown that any 2D LCPCs admits at least one non-trivial logical operator supported only on a 1D (constant width) strip of the lattice. However, it was unclear how to apply it through a sequence of local transformations.

An important subclass of LCPCs are stabilizer codes [17], for which $P_X = \frac{1}{2}(I + S_X)$ with S_X tensor-product of Pauli matrices $\sigma_{0,1,2,3}$ (with σ_0 being the identity I). Because of this particular structure, the non-trivial string-like logical operator described above is also a tensor product of Pauli matrices, $L = \bigotimes_k^\ell \sigma_{j_k}^k$ where k labels the ℓ sites along the strip in some natural way, from left to right, say. Then, applying the error sequence $\{\sigma_{j_k}^k\}$ will build up to the operator L, and will only visit intermediate states with a constant energy above the ground state. Indeed, at an intermediate stage n, $0 < n < \ell$, only a segment $\bigotimes_k^n \sigma_{j_k}^k$ of the logical operator L has been applied. This segment commutes with all terms P_X except the ones within distance $\mathcal{O}(w)$ from site k, so only these terms contribute to the energy: the excitations are point-like objects located in the vicinity on the end of the string segment, so the memory is unstable [14, 16, 18, 19]. This simple argument fails for more general LCPCs because logical operators do not have a tensor product structure.

Noise model— In this Section, we present an error sequence $\{\mathcal{E}_k\}$ that achieves a constant energy barrier. To simplify the presentation, we coarse-grain the lattice—i.e., we partition the lattice into balls of radius w and view each ball as a site occupied by a single D-level spin with $D = d^{\mathcal{O}(w^2)}$—so we can assume without lost of generality that: Λ is a regular $\ell \times \ell$ square lattice; the non-zero terms P_X in Eq. (1) act only on 2×2 cells; and there exists a non-trivial logical operator supported on a single line $\mathcal{L}$ of the lattice. Projectors whose support intersect $\mathcal{L}$ define the strip projector $P_{\mathcal{L}} = \prod_{X \cap \mathcal{L} \neq \emptyset} P_X$ supported on the extended strip $\mathcal{L}'$, see Fig. 1. Similarly, projectors whose support intersect sites $k-1$ and k on

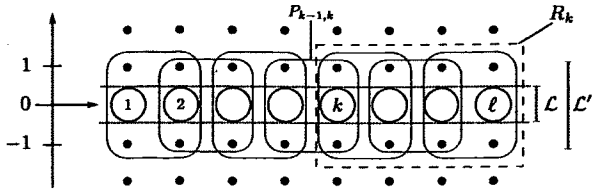


FIG. 1: The strip $\mathcal{L}$ contains ℓ sites (large circles) whose Cartesian coordinates are $\{(k,0)\ 1 \le k \le \ell\}$. Local constraints $P_{k-1,k}$ act on nearest-neighbours sites $k-1$ and k and on particles in the extended strip $\mathcal{L}' = \{(i,j)\ |j| \le 1\}$.

$\mathcal{L}$ define *local constraints* $P_{k-1,k} = \prod_{X \cap \{k-1,k\} \neq \emptyset} P_X$.

The *sequential noise model* is a sequence of individual *iterations* for every site $k \in \mathcal{L}$. Each iteration consists of several *trials*. Trial m of iteration k corresponds to (i) applying a *trial unitary* $U_k^{(m)}$ on site k, chosen at random from the Haar measure, and (ii) measuring the local constraint $P_{k-1,k}$. Trials are repeated until a *successful trial* in which the $+1$ outcome of $P_{k-1,k}$ is obtained and the next iteration begins. Given the state on the strip, a unitary is *eligible* if it leads to a successful trial with non-zero probability. The initial iteration $k = 1$ differs since the constraint is not measured. Physically, the whole procedure corresponds to creating a random excitation at iteration 1, and moving it along the strip across to the opposite edge by subsequent iterations.

The sequential noise model only creates intermediate states of constant energy. The reason is the same as for stabilizer codes: the excitations are point-like objects. Indeed, during iteration k, the state is almost everywhere indistinguishable from a ground state because it obeys all constraints $P_{i-1,i}$, except $P_{k-1,k}$ and $P_{k,k+1}$ since only those potentially do not commute with $U_k^{(m)}$. Furthermore, a failed trial during iteration k does not affect the outcome of previous iterations since local constraints commute. Next, we prove that each trial has a constant success probability and that the sequential noise model has a non-trivial effect on the ground space.

Expected number of trials— The sequential model would run into a dead-end, an iteration requiring an infinite number of trials, if the state of the strip admits no eligible unitary at the k^{th} iteration [30]. Such a dead-end occurs in the Ising-like toric code introduced in [9], where the plaquette operators B_p of the toric code are replaced by Ising-like interaction $B_p B_q$ whose symmetry is broken by a single defect plaquette B_{p^*} to recover the toric code ground space. The sequential model could start preparing the $B_p = -1$ sector and reach a dead-end when it encounters the $B_{p^*} \doteq +1$ constraint. However, that code does not have local topological order, and its spectrum is unstable (B_{p^*} is a local order parameter). We now show that dead-ends do not occur with local topological order.

Proposition 2. *Local topological order implies that, at any iteration k, there exists an eligible unitary U_k.*

Proof. We will prove the contrapositive. Let k be the first iteration where no U_k is eligible and let ψ be the state during this iteration. We have

$$P_{k-1,k} U_k |\psi\rangle = 0 \quad \forall U_k. \tag{3}$$

Mathematically, applying a Haar-random unitary operator is on average equivalent to applying the maximally depolarizing channel $\mathcal{D}_k$,

$$\mathcal{D}_k[\psi] \equiv \mathrm{Tr}_k[\psi] \otimes I_k/D = \int U_k |\psi\rangle\langle\psi| U_k^\dagger dU_k. \tag{4}$$

Thus, the average of Eq. (3) over the Haar measure, is

$$P_{k-1,k}(\mathrm{Tr}_k[\psi] \otimes I_k/D) = 0. \tag{5}$$

Tracing out the region $R_k = \{(i,j) : i \ge k\ |j| \le 1\} \subset \mathcal{L}'$ of the extended strip located at the right of site k, c.f. Fig. 1, Eq. (5) yields

$$\mathrm{Tr}_k[P_{k-1,k}] \mathrm{Tr}_{R_k}[\psi] = 0. \tag{6}$$

Thus, there exists a state $|\xi\rangle$ in the support of $\mathrm{Tr}_{R_k}[\psi]$ which is in the image of $P_{i-1,i}$ for $i < k$ but also is in the kernel of $\mathrm{Tr}_k[P_{k-1,k}]$. This entails violation of local topological order on site $k-2$ since $\mathrm{Tr}_{\overline{k-2}}[\xi]$ is in the kernel of $\rho_{k-2} = \mathrm{Tr}_{\overline{k-2}} P$ but in the image of $\rho_{k-2}^{\mathrm{loc}} = \mathrm{Tr}_{\overline{k-2}}[P_{k-3,k-2} P_{k-2,k-1}]$. □

Proposition 3. *When the system has local topological order, the expected number of trials A_k at iteration k is finite and independent of the system size.*

Proof. We introduce two maps

$$\mathcal{P}_{k-1,k}[\rho] = P_{k-1,k} \rho P_{k-1,k} \tag{7}$$

$$\mathcal{Q}_{k-1,k}[\rho] = (I - P_{k-1,k}) \rho (I - P_{k-1,k}) \tag{8}$$

which represent a successful and failed measurement of the local constraint $P_{k-1,k}$. In a failed trial, the map $\mathcal{Q}_{k-1,k}$ is always immediately preceded and followed by a depolarization of site k. This sequence can be rewritten in an equivalent form

$$\mathcal{D}_k \mathcal{Q}_{k-1,k} \mathcal{D}_k = \mathcal{B}_{k-1} \otimes \mathcal{D}_k \tag{9}$$

which defines a *biasing map* $\mathcal{B}_{k-1}$. This map is not trace-preserving since the trace of its unnormalized output state is the average probability of a failed trial.

The sequence of m failed trials followed by a successful trial produces the map

$$\mathcal{P}_{k-1,k} \mathcal{D}_k (\mathcal{Q}_{k-1,k} \mathcal{D}_k)^m = \mathcal{P}_{k-1,k} (\mathcal{B}_{k-1}^m \otimes \mathcal{D}_k) \tag{10}$$

where we have used Eq. (9) and $\mathcal{D}_k^2 = \mathcal{D}_k$. Thus, given a state ψ, the average probability $p_k^{(m)}(\psi)$ of a success after m failures is $p_k^{(m)}(\psi) = \mathrm{Tr}\left[\mathcal{P}_{k-1,k}(\mathcal{B}_{k-1}^m \otimes \mathcal{D}_k)[\psi]\right]$.

Therefore, the expected number of trials

$$A_k(\psi) = \sum_{m=1}^{\infty} (m+1)\mathrm{Tr}\left[\mathcal{P}_{k-1,k}\left(\mathcal{B}_{k-1}^m \otimes \mathcal{D}_k\right)[\psi]\right] \quad (11)$$

$$= \mathrm{Tr}\left[\mathcal{P}_{k-1,k}\left(\left(\mathcal{I}_{k-1} - \mathcal{B}_{k-1}\right)^{-2} \otimes \mathcal{D}_k\right)[\psi]\right] \quad (12)$$

is bounded by the norm of the superoperator inside the trace and thus only depends on the microscopic details of H, not on system size. Note that the geometric sum of Eq. (12) converges since $\mathcal{B}_{k-1}$ cannot have $+1$ eigenvectors in the groundspace for a local topological ordered system, since those would contradict Proposition 2. [31] □

Non-trivial average effect—We now prove that the sequential noise model corrupts the encoded information. The effect of a single iteration k averaged over all possible trials amounts to the map

$$\mathcal{E}_k = \sum_{m=0}^{\infty} \mathcal{P}_{k-1,k}\left(\mathcal{B}_{k-1}^m \otimes \mathcal{D}_k\right) \quad (13)$$

$$= \mathcal{P}_{k-1,k}\left(\left(\mathcal{I} - \mathcal{B}_{k-1}\right)^{-1} \otimes \mathcal{D}_k\right), \quad (14)$$

and the average total effect of the sequential noise model $\mathcal{E} \equiv \prod_{k=1}^{\ell} \mathcal{E}_k$ is

$$\mathcal{E} = \prod_{k=2}^{\ell} \mathcal{P}_{k-1,k}\left(\left(\mathcal{I} - \mathcal{B}_{k-1}\right)^{-1} \otimes \mathcal{D}_k\right)\mathcal{D}_1. \quad (15)$$

Terms with non-overlapping support trivially commute. We thus move all depolarizing channels to act first, which globally depolarizes the strip. To move the biasing operators past the projectors, it suffices to prove that $C \equiv [\mathcal{B}_k, \mathcal{P}_{k-1,k}]\mathcal{D}_{k+1}$ is zero. Because of their non-overlapping supports, $\mathcal{D}_{k+1}$ commutes with $\mathcal{P}_{k-1,k}$ and $C = [\mathcal{B}_k\mathcal{D}_{k+1}, \mathcal{P}_{k-1,k}] = [\mathcal{D}_{k+1}\mathcal{Q}_{k,k+1}\mathcal{D}_{k+1}, \mathcal{P}_{k-1,k}] = 0$ since $\mathcal{Q}_{k,k+1}$ and $\mathcal{P}_{k-1,k}$ commute. Hence, the terms of Eq. (15) can be reordered into

$$\mathcal{E} = \prod_{k=2}^{\ell} \mathcal{P}_{k-1,k} \prod_{k=1}^{\ell} \left(\mathcal{I} - \mathcal{B}_k\right)^{-1} \prod_{k=1}^{\ell} \mathcal{D}_k. \quad (16)$$

Thus, the average effect of the sequential noise model is equivalent to three consecutive transformations: First, all particles on the strip are removed and replaced by particles in random states. At this point, it is clear that the memory has been irreversibly corrupted. Then, an arbitrary transformation is applied on the strip before returning the system to its ground space. Hence, the error sequence $\{\mathcal{E}_k\}_{k=1}^{\ell}$ satisfies conditions Eq. (2).

Discussion— It is known that 2D LCPC have a unique Gibbs state [20], so they cannot store information in thermal equilibrium. Thus, the question of self-correction is fundamentally about the thermalization time. The noise process we presented corrupts the memory after a time that grows proportionally to the system size, which can be interpreted as a (macroscopic) upper bound to the storage time. However, there are good reasons to believe that the actual storage time is in fact independent of the system size. At nonzero temperature we expect a finite density of defects, so the noise process we described could be happening in parallel all over the lattice. As pairs of defects meet, they can fuse to the vacuum with some probability to create longer error strings. The memory time is then related to the percolation of these error chains, which should be independent of the system size.

Our result does not completely close the door to the existence of a robust quantum memory in 2D. First, local topological order is a sufficient, but perhaps not necessary condition for spectral stability. Second, we have restricted the form of the Hamiltonian. In realistc physical systems, the terms P_X need not to commute, the ground space can be frustrated, and interaction can decay algebraically with the distance between sites. Third, a macroscopic energy barrier is one mechanism that leads to thermal stability, but there may exist other mechanisms. In particular, a system in contact with a heat bath tends to minimize its free energy $F = E - TS$. Thus, we could imagine a system with a large entropy barrier: among all possible local noise sequences, only a vanishingly small fraction will induce a change of sector, while the overwhelming majority lead to dead-ends as described above. Such topological spin-glasses [21] could offer an enhanced quantum memory lifetime. This proposal is distinct from existing studies showing that disorder induces an exponential localization of anyons [22–25], as those only address zero-temperature storage.

Conclusion— Our main result hints at a general trade-off in 2D between a quantum memory's ability to suppress thermal and quantum fluctuations. Recent discoveries [26–28] indicate that this tradeoff is not necessary in 3D. Our result extends prior findings [14, 16] derived for stabilizer codes to a broader, widely studied class of models that includes quantum doubles [2], Levin-Wen [7], and Turaev-Viro [8] models. It also generalizes straightforwardly to higher dimensions for systems that have quasi-one dimensional non-trivial logical operators.

Acknowledgements— We thank Jeonwang Haah and John Preskill for insightful discussions. OLC is funded by NSERC through a Vanier Scholarship. This work was supported by Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract D11PC20167. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

* Electronic address: olivier.landon-cardinal@usherbrooke.ca
† Electronic address: david.poulin@usherbrooke.ca

[1] D. Bacon, Phys. Rev. A, **73**, 012340 (2006).
[2] A. Kitaev, Ann. Phys., **303**, 2 (2003).
[3] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys., **43**, 4452 (2002).
[4] R. Alicki, M. Fannes, and M. Horodecki, J. Phys. A: Math. and Theo., **40**, 6451 (2007).
[5] R. Alicki, M. Fannes, and M. Horodecki, J. Phys. A: Math. and Theo., **42**, 065303 (2009).
[6] Z. Nussinov and G. Ortiz, Phys. Rev. B, **77**, 064302 (2008).
[7] M. A. Levin and X.-G. Wen, Phys. Rev. B, **71**, 045110 (2005).
[8] R. Koenig, G. Kuperberg, and B. W. Reichardt, Ann. Phys., **325**, 2707 (2010), ISSN 0003-4916.
[9] S. Bravyi, M. Hastings, and S. Michalakis, J. Math. Phys., **51** (2010).
[10] S. Bravyi and M. B. Hastings, "A short proof of stability of topological order under local perturbations," ArXiv:1001.4363.
[11] S. Michalakis and J. Pytel, "Stability of frustration-free hamiltonians," arXiv:1109.1588 .
[12] L. Cincio and G. Vidal, "Characterizing topological order by studying the ground states of an infinite cylinder," (2012), arXiv:1208.2623 .
[13] J. Haah and J. Preskill, Phys. Rev. A, **86**, 032308 (2012).
[14] S. Bravyi and B. Terhal, New J. Phys., **11**, 043029 (2009).
[15] S. Bravyi, D. Poulin, and B. Terhal, Phys. Rev. Lett., **104**, 050503 (2010).
[16] A. Kay and R. Colbeck, "Quantum self-correcting stabilizer codes," (2008), arXiv:0810.3557.
[17] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, Caltech (1997).
[18] B. Yoshida, Ann. Phys., **326**, 2566 (2011).
[19] H. Bombin, G. Duclos-Cianci, and D. Poulin, New J. Phys., **14**, 073048 (2012).
[20] M. B. Hastings, Phys. Rev. Lett., **107**, 210501 (2011).
[21] C. Castelnovo and C. Chamon, Philo. Mag., **92**, 304 (2012).
[22] D. I. Tsomokos, T. J. Osborne, and C. Castelnovo, Phys. Rev. B, **83**, 075124 (2011).
[23] J. R. Wootton and J. K. Pachos, Phys. Rev. Lett., **107**, 030503 (2011).
[24] C. Stark, L. Pollet, A. m. c. Imamoğlu, and R. Renner, Phys. Rev. Lett., **107**, 030504 (2011).
[25] B. Röthlisberger, J. R. Wootton, R. M. Heath, J. K. Pachos, and D. Loss, Phys. Rev. A, **85**, 022313 (2012).
[26] J. Haah, Phys. Rev. A, **83**, 042330 (2011).
[27] S. Bravyi and J. Haah, Phys. Rev. Lett., **107**, 150504 (2011).
[28] K. Michnicki, "3-d quantum stabilizer codes with a power law energy barrier," arXiv:1208.3496 .
[29] I. Klich, Ann. Phys., **325**, 2120 (2010).
[30] Note that while the eligibility of a unitary depends on the state in general, it cannot change if the state has only support on the kernel of $P_{k-1,k}$.
[31] For $\mathcal{L} = \mathcal{D}_k \mathcal{Q}_{k-1,k} \mathcal{D}_k$, let $\mathcal{L}X_j = \lambda_j X_j$ denote its spectral decomposition. Note that $\lambda_j \geq 0$ and $\mathrm{Tr} X_j X_i^\dagger = \delta_{ij}$ since $\mathcal{L}$ is Hermitian and CP. Decompose the state at the current iteration in this basis $\psi = \sum_j \alpha_j X_j$, and let $\sigma = \lim_{n\to\infty} \frac{1}{n} \sum_{\eta=1}^{n} \mathcal{L}^\eta(\rho) = \sum_{j:\lambda_j=1} \alpha_j X_j$. Note that $\sigma \geq 0$ since $\mathcal{L}$ is CP, and that $\mathrm{Tr}\sigma = 0$ by Proposition 2, which implies that $\alpha_j = 0$ when $\lambda_j = 1$.

7.4 Discussion

Notre résultat impose des contraintes fortes sur l'existence d'une mémoire topologique autocorrectrice en 2D. Nous allons brièvement mentionner en 7.4.1 que la situation est différente lorsque le nombre de dimensions spatiales augmente, en particulier en 3D et 4D[3], avant de discuter des mécanismes de protection alternatifs qui sont envisageables en 2D, en 7.4.2. Cela nous amènera à discuter de l'idée de mémoire à protection entropique, en 7.4.2.2. Finalement, nous reviendrons sur la question soulevée dans l'article sur l'existence d'un modèle d'anyons pour tout hamiltonien CPC topologique, en 7.4.3.

7.4.1 Mémoires quantiques en dimension supérieure

En 3D, Haah a exhibé un code stabilisateur local (avec des interactions à 8 corps), appelé *code cubique*, dont la barrière d'énergie grandit logarithmiquement avec la taille du système [98]. De plus, Bravyi et Haah [112] ont montré analytiquement et numériquement que l'information quantique peut être préservée durant un temps $\tau_{mem} \geq L^{c\beta}$ où c est une constante et β est la température inverse. Toutefois, ce comportement n'est maintenu que pour une taille inférieure à une taille critique $L^* \sim e^{\beta/3}$. Le travail de Bravyi et Haah est particulièrement intéressant car il propose une modélisation Linbladienne de l'interaction mémoire-environnement et considère un décodeur explicite pour récupérer l'information après la période de stockage. Ce comportement est rendu possible grâce à la structure fractale des opérateurs logiques qui ne sont donc pas des opérateurs en ruban. En l'absence d'opérateurs en ruban, la barrière d'énergie doit au moins être logarithmique [113] et la dépendance logarithmique semble être typique d'un code stabilisateur local et invariant sous translation en 3D [114]. Si on lève la contrainte d'invariance sous translation, il est possible d'obtenir une plus grande barrière en énergie. Kamil Michnicki a exhibé un code stabilisateur local [115] dont la barrière d'énergie grandit $\Delta^* = L^{2/3}$. Ce code résulte de la fusion de plusieurs codes obtenus grâce à un procédé dit de soudure de code.

En 4D, la version 4D du code torique [116] possède un temps de stockage qui grandit exponentiellement avec la taille du système [117].

3. Un hamiltonien local 4D fait apparaître des interactions locales entre particules voisines sur un réseau en quatre dimensions spatiales. Expérimentalement, un tel hamiltonien pourrait décrire la dynamique de particules sur un réseau, mais des interactions à longue portée très sélective Il est possible d'utiliser un réseau avec un plus p

Or, la construction d'une mémoire en 3D, voire 4D, semble au-delà des capacités expérimentales actuelles. En effet, construire une mémoire 3D demande des interactions faisant intervenir un grand nombre de corps, par exemple des interactions à 8 qubits pour le code cubique. Pire, enchâsser une géométrie 4D en 3D demande d'avoir des interactions à longue portée très contrôlée. De plus, la plupart des procédés de nanofabrication actuelle sont optimisés pour des géométries 2D. Ainsi, il serait préférable d'avoir une mémoire auto-correctrice en 2D pour arriver à un dispositif expérimental. De plus, il est possible qu'une phase auto-correctrice 2D soit réalisée naturellement dans des matériaux. Par exemple, l'hébertsmithite ($ZnCu_3(OH)_6Cl_2$) est un cristal naturel dont on pense que les plans de Cu ont un état de liquide de spin 2D topologiquement ordonné [118]. Or, notre article montre que la protection grâce à une barrière d'énergie n'existe pas en 2D, au moins pour des codes CPC obéissant à la condition de cohérence locale. Nous allons donc explorer les mécanismes alternatifs en 2D.

7.4.2 Mécanismes alternatifs de protection en 2D

7.4.2.1 Mécanismes proposés dans la littérature

L'intuition de l'instabilité thermique est liée à la présence d'anyons qui se propagent sans coût énergétique. Une parade est alors de trouver un mécanisme qui entrave la propagation des anyons. Ainsi, plusieurs propositions existent afin de faire apparaître une interaction effective à longue portée entre les anyons [119, 120, 121]. Une autre façon d'empêcher la propagation des anyons est d'introduire du désordre dans le hamiltonien afin d'entraîner une localisation des défauts à température nulle [122, 123]. Finalement, une autre approche est de considérer l'interaction entre la mémoire et son environnement et d'y faire apparaître des termes qui imitent la correction d'erreur active [124] afin que la dissipation amène l'état du système vers l'espace fondamental, plutôt que vers un état de Gibbs. Nous allons maintenant proposer une autre approche, celle d'une protection basée sur des considérations entropique.

7.4.2.2 Mémoire à protection entropique ?

Importance de l'entropie Revenons sur l'exemple du modèle d'Ising classique 2D qui est thermiquement stable en raison d'une transition de phase en 2D. Ignorons pour cette discussion le fait que le spectre du modèle quantique soit instable pour ne se concentrer que sur les raisons de sa stabilité thermique. Ce modèle présente une barrière d'énergie $\Delta^* \sim L$ puisque l'énergie

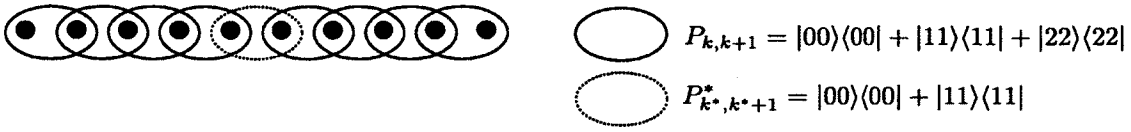


FIGURE 7.5 Modèle jouet pour une impasse.

d'une région macroscopique de spins basculés grandit avec le périmètre de la région. Toutefois, en remarquant que la densité de défauts est constante à température finie, l'énergie disponible dans le système grandit comme L^2. Ainsi, la barrière d'énergie ne suffit pas à expliquer pourquoi l'état du système garde la mémoire du fondamental initial. Or, à température non-nulle, le système a tendance à minimiser l'énergie libre $F = E - TS$. Ainsi, à basse température, il y a un compromis entre minimiser l'énergie et maximiser l'entropie. Pour le modèle d'Ising, maximiser l'entropie revient à favoriser le grand nombre de configurations microscopiques où une multitude de petites régions où les spins ont basculé plutôt que le petit nombre de configurations où une région macroscopique a basculé.

On comprend donc sur l'exemple du modèle d'Ising que la prise en compte de l'entropie est essentielle afin de comprendre la stabilité thermique. On peut alors se demander s'il est possible d'avoir un mécanisme de protection entropique afin de protéger un code 2D.

Apparition d'impasse en absence de cohérence locale Un premier indice vers un potentiel mécanisme de protection entropique apparaît dans le raisonnement de l'article par l'existence d'impasses si la condition de cohérence locale n'est pas respectée. Une impasse est une situation où l'état de la mémoire est dans un état qui respecte toutes les contraintes $P_{k,k+1}$ pour $k < m-1$ et $k > m$, mais pour lequel aucune transformation unitaire sur le site m ne permet de satisfaire la contrainte $P_{m-1,m}$. Nous allons maintenant voir un modèle simple où une telle impasse apparaît.

Soit une bande $\mathcal{L}'$ de ℓ sites dont la dimension quantique locale est 3. Il s'agit donc d'une chaîne de qutrits dont l'état appartient à vec $\{|0\rangle, |1\rangle, |2\rangle\}$. Les contraintes locales sont $P_{k,k+1} = |00\rangle\langle 00| + |11\rangle\langle 11| + |22\rangle\langle 22|$ sauf pour l'interaction entre les qutrits k^* et k^*+1 qui est $P^*_{k^*,k^*+1} = |00\rangle\langle 00| + |11\rangle\langle 11|$, cf. Fig. 7.5. Le projecteur sur la bande est alors $P_{\mathcal{L}'} = |0\rangle\langle 0|^{\otimes \ell} + |1\rangle\langle 1|^{\otimes \ell}$. Supposons que l'état initial soit $|0\rangle^{\otimes \ell}$. Suivant la prescription de l'article, l'environnement pourrait préparer l'état $|2\rangle^{\otimes i}|0\rangle^{\otimes \ell - i}$ qui respecte les contraintes locales. Toutefois, ce processus va rencontrer une impasse au niveau du site k^*. En effet, l'état sera alors $|2\rangle^{\otimes k^*}|0\rangle^{\otimes \ell - k^*}$ et aucune transformation unitaire sur $k^* + 1$ ne pourra faire en sorte que la contrainte $P^*_{k^*,k^*+1}$ soit respectée. Comme

démontré dans l'article, cela n'arrive que parce que la condition de cohérence est violée. En effet, pour n'importe quel site k loin du défaut k^*, on a $\rho_k \equiv \mathrm{Tr}_{\Lambda\setminus k} P_{\mathcal{C}} = |0\rangle\langle 0| + |1\rangle\langle 1|$ alors que la version locale est $\rho_k^{\mathrm{loc}} = \mathrm{Tr}_{\Lambda\setminus k} P_{k-1,k} P_{k,k+1} = |0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2|$ qui ont des noyaux différents.

On peut alors imaginer favoriser l'apparition de telles impasses afin que l'environnement reste coincé pour une longue durée de temps, sans perdre l'information sur son état initial. Il s'agirait alors d'une mémoire à protection entropique. Toutefois, il est difficile de favoriser l'apparition d'impasse. Idéalement, on aimerait disposer d'un outil graphique afin de les représenter.

A priori, toute l'information sur ces impasses se retrouvent dans les projecteurs élémentaires qui définissent le code. Toutefois, il ne suffit pas de regarder le projecteur sur la bande où vit un opérateur logique non-trivial puisque l'information peut provenir d'ailleurs dans le réseau. En effet, dans l'exemple d'impasse, le défaut était situé dans la bande, mais il aurait pu être situé à l'extérieur de la bande étendue $\mathcal{L}'$ et sa contrainte peut être amenée par des interactions locales jusqu'à la bande $\mathcal{L}$. On pourrait alors être tenté de définir des projecteurs $P'_{k,k+1} = \mathrm{Tr}_{\Lambda\setminus\{k,k+1\}}[P]$ qui tiennent compte de toutes les contraintes dans le code. Toutefois, il n'est pas clair que ces nouveaux projecteurs commutent. Ainsi, pour des codes sans cohérence locale, il faut comprendre la structure du projecteur global et en particulier de l'algèbre logique. Un pas intéressant dans cette direction est la travail de Hastings sur la caractérisation de l'ordre topologique en terme d'éléments centraux d'une algèbre d'interaction [125].

7.4.3 Excitations de basse énergie d'un hamiltonien topologique

L'intuition de notre résultat repose sur la capacité pour l'environnement de créer une excitation localisée à une frontière du réseau, la déplacer sur une distance macroscopique jusqu'à atteindre une autre frontière, puis retourner au vide en fusionnant cette excitation à cette autre frontière. Or, rien dans notre raisonnement formel ne fait apparaître de modèle d'anyons. En effet, le point de départ de notre travail est un hamiltonien CPC défini par ses projecteurs élémentaires auquel on impose les conditions d'ordre topologique et de cohérence locale. On ne sait pas si cela suffit à garantir l'existence d'un modèle anyonique pour décrire ses excitations de basse énergie. Cette question constitue l'inverse de l'approche courante où le hamiltonien topologique est construit à partir d'un modèle anyonique, comme pour les quantum double de Kitaev [80] ou les modèles de Levin-Wen [103]. Soulignons que même s'il existe, déterminer les caractéristiques du modèle anyonique correspondant à un hamiltonien microscopique est un problème non-trivial pour lequel

des procédures numériques ont été proposées récemment [126, 127].

Intuitivement, la présence d'un opérateur logique non-trivial en ruban laisse supposer qu'il existe des excitations anyoniques correspondant à l'application partielle de cet opérateur. Une façon de formaliser l'application partielle jusqu'au site k est de considérer la décomposition de Schmidt $L = \sum_{\alpha_{k+1}} L_{\alpha_{k+1}}^{[1;k]} \otimes L_{\alpha_{k+1}}^{[k+1;\ell]}$. L'application partielle de l'opérateur logique correspond à $\sum_{\alpha_{k+1}} L_{\alpha_{k+1}}^{[1;k]}$ où l'indice α_{k+1} énumère soit les charges topologiques qui peuvent apparaître sur la région $k+1$ ou encore l'historique des canaux de fusion. Ainsi, α_{k+1} pourrait prendre plusieurs valeurs dans le cas d'anyons non-abéliens dont la fusion n'est pas déterministe.

Notre résultat peut-être vu comme un premier pas vers la formalisation rigoureuse de cette intuition. En effet, la séquence d'erreur logique est compatible avec la vision en terme d'anyons. La transformation unitaire sur le site 1 crée de façon probabiliste une excitation locale qui n'est détectée que par la contrainte $P_{1,2}$. Puis, une transformation unitaire sur la région 2 qui parvient à satisfaire la contrainte $P_{1,2}$ correspond à bouger cette excitation sur le support de la contrainte $P_{2,3}$. Ainsi, chaque itération consiste à déplacer l'excitation avant de finalement fusionner cette excitation à une extrémité afin de revenir au vide en ayant probablement appliquer une transformation non-triviale sur l'espace fondamental. Un résultat intéressant serait d'éliminer le caractère aléatoire du choix des transformations unitaires afin de rendre la procédure déterministe.

Conclusion

Caractérisation pratique des systèmes quantiques

La première partie de cette thèse s'est intéressée à la caractérisation pratique des systèmes quantiques. Nous avons d'abord montré au chapitre 2 que la technique utilisée jusqu'à présent, dite de tomographie, demande une quantité de ressources expérimentales et numériques rhédibitoire pour des systèmes de plus d'une dizaine de particules. Nous avons expliqué que le coût exponentiel de la tomographie est lié à la généralité de la tâche qui consiste à reconstruire la fonction d'onde globale dans un espace de Hilbert dont la taille grandit exponentiellement avec le nombre de particules. Cela nous a amené à suggérer des tâches réduites qui visent à acquérir moins d'information et qui pourraient donc être accomplies avec des ressources moindres. Cette idée, bien qu'évoquée dans des travaux antérieurs, est un apport novateur de cette thèse.

Dans le chapitre 3, nous avons proposé la tâche de certification qui estime la distance entre l'état expérimental et l'état cible que l'expérimentateur voulait préparer. Ce protocole est très efficace pour plusieurs classes d'états importantes pour l'informatique quantique et demande systématiquement moins de ressources expérimentales et numériques que la tomographie. De plus, ce protocole est très flexible et peut être adapté afin de répondre aux contraintes d'un dispositif expérimental particulier. Ainsi, un projet mené conjointement avec un groupe expérimental serait la continuité logique de ce travail théorique. Rappelons à ce sujet que le protocole de certification a été utilisé expérimentalement [16].

Dans le chapitre 4, nous avons proposé le paradigme de la tomographie variationnelle. Celui-ci consiste à ne considérer qu'une classe variationnelle d'états lors de la reconstruction de l'état expérimental. L'espoir est alors de trouver des protocoles qui permettent d'identifier le petit nombre de paramètres variationnels grâce à un petit nombre de mesures expérimentales. Nous avons

exhibé de tels protocoles pour les classes variationnelles des états à produits matriciels ou *matrix product states* (MPS) et celle de l'ansatz pour intrication multi-échelle ou *multi-scale entanglement renormalization ansatz* (MERA). Ces deux classes variationnelles sont très utilisées numériquement, p.ex. afin de déterminer les états fondamentaux de hamiltoniens locaux. Ainsi, la tomographie variationnelle pourrait servir à confirmer expérimentalement des prédictions numériques. Des systèmes prometteurs pour cette comparaison théorie/expérience sont les simulateurs quantiques : des systèmes expérimentaux hautement contrôlables qui peuvent reproduire la dynamique d'autres systèmes quantiques. Un travail conjoint avec des expérimentateurs pour mettre en place expérimentalement ces techniques de tomographie variationnelle serait prometteur. Par ailleurs, notre travail ne fait qu'initier l'idée de la tomographie variationnelle : il serait intéressant de considérer d'autres classes variationnelles plus générales, p.ex. les états à produits de paires intriquées ou *projected entangled pair states* (PEPS). De plus, la question de l'élaboration une théorie complète de la tomographie variationnelle reste ouverte : des liens intriguants existent entre les classes d'états à description efficace et ceux qui peuvent être identifiés expérimentalement de façon efficace.

Mémoires quantiques auto-correctrices 2D

La seconde partie de la thèse s'intéresse à l'existence d'un système auto-correcteur constitué de particules disposées sur un réseau bidimensionnel (2D). L'information y serait encodée dans la superposition arbitraire d'états fondamentaux. Afin de préserver la cohérence de la superposition, il convient de trouver un hamiltonien dont le fondamental est dégénéré et qui ne possède pas de paramètre d'ordre local. Ainsi, les modèles à symétrie brisée doivent être évités et cela nous a amené à nous intéresser à des systèmes topologiques dans le chapitre 5. Après avoir introduit le modèle des liquides de spin pour des raisons pédagogiques, nous avons exploré en grand détail le modèle canonique d'ordre topologique : le code torique. Ce modèle fournit une intuition utile pour toutes les notions introduites ultérieurement.

Préserver la cohérence d'une superposition arbitraire est étroitement lié à la stabilité du spectre de basse énergie du hamiltonien. Le chapitre 6 introduit la classe des hamiltoniens locaux, non-frustrés, commutatifs pour lesquels il existe un critère qui garantit la stabilité spectrale. Ces modèles à spectre stable sont précisément les modèles topologiques. Toutefois, la stabilité du spectre ne garantit pas la robustesse de l'information face à un environnement thermique. Nous quantifions cette robustesse en évaluant la barrière d'énergie que doit surmonter toute séquence

de transformations locales faisant passer un état fondamental à un autre. Nous justifions qu'un système dont la barrière d'énergie grandit avec sa taille serait auto-correcteur. La formulation formelle de ce critère est un apport novateur de cette thèse, mais n'est qu'une première tentative afin de déterminer la robustesse d'un système face à la thermalisation. La question devient alors : parmi les modèles topologiques, en existe-t-il un qui soit auto-correcteur ? Le chapitre 7, et plus précisément notre article, répond à cette question par la négative.

L'intuition est que les systèmes topologiques admettent des excitations de basse-énergie qui sont des quasi-particules anyoniques. Dans ce cas, l'environnement peut créer facilement de telles quasi-particules. Deux états fondamentaux distincts vont alors aboutir dans le même état de basse énergie et l'information encodée dans l'état initial aura été perdue. Or, on ne sait pas démontrer que tout hamiltonien topologique (au sens du critère de stabilité spectrale) admet des excitations anyoniques. Il s'agit en fait d'une question ouverte importante. Toutefois, la littérature montre qu'il est possible de passer d'un état fondamental à un autre grâce à un opérateur agissant sur un ruban, une propriété remarquable pour un système 2D. Notre résultat propose une transformation stochastique qui déplace des paquets d'énergie afin de passer d'un état fondamental à un autre. Cela suffit à démontrer qu'aucun système 2D topologique ne peut avoir une barrière d'énergie grandissant avec la taille du système.

Ce résultat impose de très grandes contraintes sur l'existence éventuelle d'un système auto-correcteur 2D. Toutefois, on peut imaginer des mécanismes plus subtils, qui vont au-delà des considérations énergétiques. Nous suggérons l'exploration de mécanismes d'origine entropique afin de protéger l'information. De plus, notre résultat est un premier pas afin de démontrer rigoureusement l'existence d'anyons pour les systèmes topologiques. Établir ce lien rigoureusement serait une avancée majeure dans notre compréhension des systèmes topologiques.

Annexes

Annexe A

Éléments techniques sur les MPS

A.1 MPS comme états peu intriqués

A.1.1 Décomposition de Schmidt

Un outil très employé en informatique quantique est la décomposition de Schmidt. Elle affirme que tout état pur biparti $|\psi^{AB}\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ admet une décomposition biorthonormale de la forme

$$|\psi^{AB}\rangle = \sum_{i=1}^{\chi} \sqrt{p_i} |\phi_i^A\rangle |\varphi_i^B\rangle \tag{A.1}$$

où les familles $\left\{|\phi_i^A\rangle\right\}_i$ et $\left\{|\varphi_i^B\rangle\right\}_i$ sont orthonormales et les $p_i \in \mathbb{R}^{+*}$ vérifient $\sum_i p_i = 1$.

Une façon de comprendre la décomposition de Schmidt est de réaliser qu'elle est une reformulation de la décomposition en valeur singulière. Une autre approche, plus intuitive physiquement est de considérer la matrice densité réduite $\rho^A \equiv \mathrm{Tr}_B \left[|\psi^{AB}\rangle\langle\psi^{AB}|\right]$ qu'on diagonalise dans une base orthonormale (car elle est hermitienne)

$$\rho^A = \sum_{i=1}^{\chi} p_i |\phi^A\rangle\langle\phi^A|. \tag{A.2}$$

L'état pur se réécrit alors

$$|\psi^{AB}\rangle = \sum_{ik} c_{ik}|\phi_i^A\rangle|k_j^B\rangle = \sum_i |\phi_i^A\rangle|\tilde{\varphi}_i^B\rangle \tag{A.3}$$

où $\{|k_j^B\rangle\}$ est une base quelconque de $\mathcal{H}^B$ et nous avons défini $|\tilde{\varphi}_j^B\rangle \equiv \sum_k c_{ik}|k_k^B\rangle$. La matrice densité réduite sur A s'écrit maintenant

$$\rho^A = \sum_{ij} \langle\tilde{\varphi}_j^B|\tilde{\varphi}_i^B\rangle|\phi_i^A\rangle\langle\phi_j^A| \tag{A.4}$$

d'où on déduit que $\langle\tilde{\varphi}_j^B|\tilde{\varphi}_i^B\rangle = p_i\delta_{ij}$. Il suffit de normaliser en écrivant $|\tilde{\varphi}_i^B\rangle = \sqrt{p_i}|\varphi_i^B\rangle$ pour obtenir la forme désirée. Notons que les $\{|\phi_i^A\rangle\}_i$ sont les vecteurs propres de ρ^A et que les $\{|\varphi_i^B\rangle\}_i$ sont les vecteurs propres de ρ^B.

Le nombre de termes non-nuls dans la décomposition de Schmidt, appelé rang de Schmidt χ, quantifie (de façon très grossière [1]) l'intrication. En effet, un état est intriqué si et seulement s'il y au moins deux termes dans la décomposition de Schmidt.

A.1.2 Décompositions de Schmidt répétées

Une représentation des MPS, motivée par la simulation classique de calcul quantique, est obtenue grâce à des décompositions de Schmidt répétées [128]. Considérons un état $|\Psi\rangle$ sur n qubits. Nous penserons à ces qubits comme un système 1D : ils seront donc disposés sur une ligne et nous les numéroterons de 1 à n.

La décomposition de Schmidt sur la bipartition $1|2\ldots n$ donne

$$|\Psi\rangle = \sum_{\alpha_1=1}^{\chi_1} \lambda_{\alpha_1}^{[1]}|\phi_{\alpha_1}^{[1]}\rangle|\phi_{\alpha_1}^{[2\ldots n]}\rangle \tag{A.5}$$

Les vecteurs propres $\left\{|\phi_{\alpha_1}^{[1]}\rangle\right\}_{\alpha_1=1}^{\chi_1}$ peuvent être exprimés dans une base de référence, p. ex. la base

1. Considérons $|\psi^{AB}\rangle = \sqrt{1-\epsilon}|\phi^A\rangle|\varphi^B\rangle + \sqrt{\epsilon/d}\sum_i |ii\rangle$. La matrice densité réduite sur A est $\rho^A = (1-\epsilon)|\phi\rangle\langle\phi| + \epsilon\mathbb{I}/d$. Son entropie est très petite, de l'ordre de $\epsilon \log 1/\epsilon$ alors que le rang de Schmidt de $|\psi^{AB}\rangle$ est maximal $\chi = \sqrt{d}$.

de calcul $\{|i_1\rangle\}_{i_1=0}^{1}$, pour obtenir

$$|\Psi\rangle = \sum_{\alpha_1=1}^{\chi_1} \sum_{i_1} \lambda_{\alpha_1}^{[1]} \Gamma_{\alpha_1}^{[1]i_1} |i_1\rangle \otimes |\phi_{\alpha_1}^{[2\ldots n]}\rangle \tag{A.6}$$

où apparaissent les tenseurs à deux indices

$$\Gamma_{\alpha_1}^{[1]i_1} \equiv \langle i_1 | \phi_{\alpha_1}^{[1]} \rangle. \tag{A.7}$$

On veut maintenant exprimer les vecteurs $|\phi_{\alpha_1}^{[2\ldots n]}\rangle$ à l'aide des vecteurs propres $\left\{|\phi_{\alpha_2}^{[3\ldots n]}\rangle\right\}$ de $\rho^{[3\ldots n]}$. Plus généralement, on peut effecteur cette décomposition pour le qubit ℓ à l'aide des décompositions de Schmidt pour les bipartitions $1\ldots\ell-1|\ell\ldots n$ et $1\ldots\ell|\ell+1\ldots n$

$$|\Psi\rangle = \sum_{\alpha_{\ell-1}} \lambda_{\alpha_{\ell-1}}^{[\ell-1]} |\phi_{\alpha_{\ell-1}}^{[1\ldots\ell-1]}\rangle |\phi_{\alpha_{\ell-1}}^{[\ell\ldots n]}\rangle \tag{A.8}$$

$$= \sum_{\alpha_\ell} \lambda_{\alpha_\ell}^{[\ell]} |\phi_{\alpha_\ell}^{[1\ldots\ell]}\rangle |\phi_{\alpha_\ell}^{[\ell+1\ldots n]}\rangle. \tag{A.9}$$

En effet, on peut décomposer $|\phi_{\alpha_{\ell-1}}^{[\ell\ldots n]}\rangle$ en fonction de $|\phi_{\alpha_\ell}^{[\ell+1\ldots n]}\rangle$ de la façon suivante

$$|\phi_{\alpha_{\ell-1}}^{[\ell\ldots n]}\rangle = \sum_{\alpha_\ell} \frac{\lambda_{\alpha_\ell}^{[\ell]}}{\lambda_{\alpha_{\ell-1}}^{[\ell-1]}} \langle \phi_{\alpha_{\ell-1}}^{[1\ldots\ell-1]} | \phi_{\alpha_\ell}^{[1\ldots\ell]} \rangle |\phi_{\alpha_\ell}^{[\ell+1\ldots n]}\rangle \tag{A.10}$$

$$= \sum_{\alpha_\ell} \sum_{i_\ell} \lambda_{\alpha_\ell}^{[\ell]} \Gamma_{\alpha_{\ell-1}\alpha_\ell}^{[\ell]i_\ell} |i_\ell\rangle |\phi_{\alpha_\ell}^{[\ell+1\ldots n]}\rangle \tag{A.11}$$

où nous avons introduit le tenseur à trois indices

$$\Gamma_{\alpha_{\ell-1}\alpha_\ell}^{[\ell]i_\ell} \equiv \frac{\langle \phi_{\alpha_{\ell-1}}^{[1\ldots\ell-1]} \langle i_\ell | \phi_{\alpha_\ell}^{[1\ldots\ell]} \rangle}{\lambda_{\alpha_{\ell-1}}^{[\ell-1]}}. \tag{A.12}$$

Ainsi, en utilisant récursivement la formule (A.12) à partir de l'équation initiale (A.7), on écrit

$$|\Psi\rangle = \sum_{\alpha_1,\ldots,\alpha_{n-1}} \sum_{i_1,\ldots,i_n} \Gamma_{\alpha_1}^{[1]i_1} \lambda_{\alpha_1}^{[1]} \Gamma_{\alpha_1\alpha_2}^{[2]i_2} \lambda_{\alpha_2}^{[2]} \ldots \lambda_{i_{n-1}}^{[n-1]} \Gamma_{\alpha_{n-1}}^{[n]i_n} |i_1 \ldots i_n\rangle \tag{A.13}$$

Jusqu'ici, cette décomposition est très générale. Elle devient toutefois particulièrement intéressante, si le rang de Schmidt de chaque bipartition du système est bornée par une constante, *i.e.* si $\forall \ell\ \chi_\ell \leq D$, ce qu'on attend pour tous les états 1D obéissant à une loi d'aire, en particulier tous les fondamentaux de hamiltonien locaux gappés [67]. Dans ce cas, la décomposition (A.13) ne fait apparaître qu'un nombre linéaire de coefficients, grosso modo $2nD^2$. Ainsi, cette décomposition permet de reconstruire un nombre exponentiel de coefficients, les $\langle i_1 \dots i_n|\Psi\rangle$, à l'aide d'un nombre linéaire de coefficients, les tenseurs $\Gamma^{[\ell]i_\ell}_{\alpha_{\ell-1}\alpha_\ell}$ et les scalaires $\lambda^{[\ell]}_{\alpha_\ell}$.

A.2 MPS par projection de paires intriquées

Une approche alternative pour construire les MPS est de tous les construire à partir d'un état ressource très intriqué sur des particules virtuelles puis à « projeter » ces particules virtuelles sur des particules physiques localement afin d'obtenir le MPS désiré.

L'état ressource contient $2n$ quDits où D est la dimension de lien où chaque paire de quDits est dans un état maximalement intriqué

$$|\omega_D\rangle = \frac{1}{D}\sum_{i=1}^{D} |i\rangle \otimes |i\rangle. \tag{A.14}$$

L'état ressource s'écrit donc

$$|\Omega_{D,n}\rangle = \bigotimes_{k=1}^{n} |\omega_{2k-1,2k}\rangle \tag{A.15}$$

Ces quDits sont des degrés de liberté virtuels qui seront par la suite projetés sur l'espace physique. Pour ce faire, un opérateur $\mathcal{A}$ est défini sur des paires de quDits appartenant à des états maximalement intriqués, *i.e.* sur les quDits d'indice $(2k, 2k+1)$. Cet opérateur prend donc deux quDits et les « projète » sur un qudit. Il s'agit donc d'un tenseur de rang 3 $\mathcal{A}^i_{\alpha\beta}$. Ce tenseur, qu'on peut écrire $\mathcal{A} = \sum_i \sum_{\alpha\beta} A^i_{\alpha\beta}|i\rangle\langle\alpha\beta|$, n'est rien d'autre que la famille de matrices $\{A^i\}_i$ dans la définition matricielle des MPS. Schématiquement, la construction d'un MPS par projection de paires intriquées est représentée sur la figure A.1.

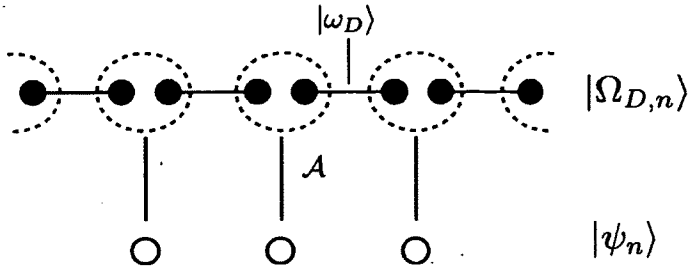


FIGURE A.1 MPS $|\psi_n\rangle$ sur n qudits, préparés à partir d'un état ressource $|\Omega_{D,n}\rangle$ sur $2n$ quDits.

État AKLT L'exemple historique de cette approche est l'état AKLT. Il est l'unique fondamental d'une chaîne de spin-1 dont le hamiltonien est

$$H_{AKLT} = \sum_i \frac{1}{2}\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{6}(\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 + \frac{1}{3}\mathbb{I} = \sum_i P_{i,i+1} \tag{A.16}$$

où $P_{i,i+1}$ est le projecteur sur les représentations de spin-2 d'une paire de spin-1. En effet, de la même façon que deux spin-1/2 donnent naissance à un espace singulet de dimension 1 (spin-0) et un espace triplet de dimension 3 (spin-1), l'addition des moments angulaires de deux spin-1 donnent naissance à trois espaces

$$1 \otimes 1 = 0 \oplus 1 \oplus 2 \tag{A.17}$$

de dimension 1, 3 et 5. L'opérateur $P_{i,i+1}$ est le projecteur sur les représentations de spin-2. Ainsi, le hamiltonien pénalise les spin-2. Pour construire le fondamental du modèle AKLT, l'idée est partir d'une chaîne de $2n$ spin-1/2 virtuels préparés dans un état singulet $|\psi^-\rangle \propto |01\rangle - |10\rangle$, *i.e.* dans un état de spin-0. L'état ressource virtuel est donc

$$\bigotimes_{k=1}^{n} |\psi^-_{2k-1,2k}\rangle. \tag{A.18}$$

Afin d'obtenir les particules physique de spin-1, on va projeter une paire de spin-1/2 virtuels, dont chacun appartient à un état singulet différent, vers un spin-1 grâce à l'opérateur $\mathcal{A}_{2k,2k+1}$ qui envoie l'espace triplet vers un spin-1

$$\mathcal{A}_{2k,2k+1} = |+1\rangle\langle 00| + |0\rangle\frac{\langle 01| + \langle 01|}{\sqrt{2}} + |-1\rangle\langle 11| \tag{A.19}$$

où les états $\{|+1\rangle, |0\rangle, |-1\rangle\}$ sont les états propres de l'opérateur S_z et forment une base de l'espace de Hilbert d'un spin-1. On obtient donc l'état

$$|AKLT_n\rangle = \bigotimes_{k=1}^{n} \mathcal{A}_{2k,2k+1} \bigotimes_{k=1}^{n} |\psi^{-}_{2k-1,2k}\rangle. \tag{A.20}$$

À partir des opérateurs de « projection », il est possible de reconstruire les tenseurs MPS de l'état AKLT, cf. section 4.2.1.2.

État AKLT ($d = 3$, $D = 2$)

$$A^{+} = \sqrt{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad A^{0} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad A^{-} = \sqrt{2}\begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \tag{A.21}$$

A.3 MPS et hamiltonien parent

Numériquement, on utilise souvent les MPS comme une classe variationnelle afin de déterminer le fondamental d'un hamiltonien. Il est intéressant de se poser le problème inverse : étant donné un MPS, est-il possible de construire un hamiltonien local dont il soit le fondamental, voire l'unique fondamental ? Nous allons qu'il est toujours possible de construire un tel hamiltonien [129], mais que l'unicité n'est obtenue que pour des MPS injectifs, que nous définissons maintenant.

La construction de ce hamiltonien est utilisée dans l'apprentissage des MPS, détaillé dans l'article en section 4.3.

A.3.1 Injectivité

Nous donnerons la définition d'injectivité en suivant l'approche de [129]. Pour simplifier, on considère un MPS dont la dimension de lien D ne dépend pas du site, *i.e.* $\forall k\ D_k = D$.

Une première notion important est celle de regroupement. Un MPS est défini par une famille de matrices $\left\{A^{[k]i_k}\right\}$ pour chaque particule k. Il peut être intéressant de regrouper plusieurs

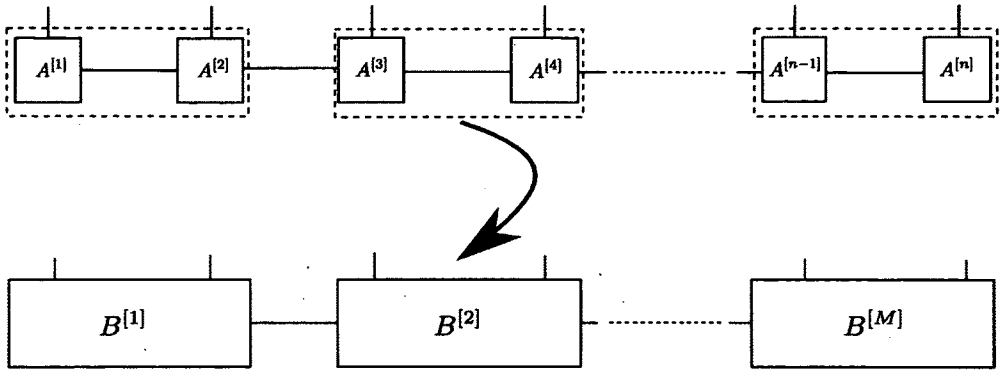


FIGURE A.2 Regroupement de tenseurs : chaque tenseur du MPS regroupé $\{B^{[m]}\}_{m=1}^{M}$ correspond à deux tenseurs du MPS original $\{A^{[k]}\}_{k=1}^{n}$.

particules, typiquement un nombre fini d'entres elles, pour obtenir des matrices pour M blocs de L particules

$$B^{[m_k]i_k\ldots i_{k+L-1}} = \prod_{\ell=k}^{k+L-1} A^{[\ell]i_\ell} \tag{A.22}$$

Schématiquement, l'opération de regroupement est représentée sur la figure A.2.

Pour un MPS générique, on s'attend à ce que la dimension de l'espace engendré par les matrices $\left\{B^{[k\ldots k+L-1]i_k\ldots i_{k+L-1}}\right\}$ grandissent comme d^L et donc génère l'espace des matrices $D \times D$ pour $L \sim 2\log_d D$. On peut formaliser cette idée intuitive de généricité par la condition suivante, dite d'injectivité.

Définition 7. Un MPS est injectif si, pour tout m_k, il existe une taille de regroupement $L \in \mathbb{N}^*$ telle que la fonction

$$\begin{array}{rccc} \Gamma_L^{[m_k]} : & \mathbb{C}^{D^2} & \longrightarrow & \mathcal{H}_d^{\otimes L} \\ & X & \mapsto & \sum_{i_1\ldots i_L} \mathrm{Tr}\left[X \prod_{\ell=k}^{k+L-1} A^{[\ell]i_\ell}\right] |i_k \ldots i_{k+L-1}\rangle \end{array} \tag{A.23}$$

est injective.

Intuitivement, $\Gamma_L^{[m_k]}(X)$ est l'état de L qudits obtenus lorsque l'on habille le tenseur $B^{[m_k]}$ par le tenseur X (appelé environnement), voir Fig. A.3. L'injectivité stipule que des environnements différents donnent naissance à des états différents. Puisque $\Gamma_L^{[m_k]}$ est une application linéaire,

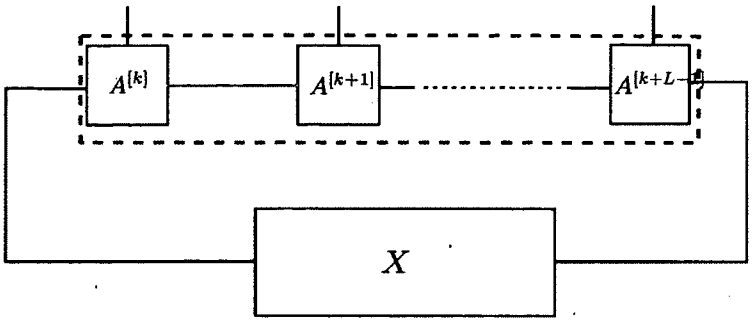


FIGURE A.3 Fonction définissant le critère d'injectivité des MPS

l'injectivité se ramène à $\Gamma_L^{[m_k]}(X) = 0 \Rightarrow X = 0$, *i.e.* que

$$\forall(i_k \dots i_{k+L-1}) \ \mathrm{Tr}\left[X \prod_{\ell=k}^{k+L-1} A^{[\ell]i_\ell}\right] = 0 \Rightarrow X = 0. \tag{A.24}$$

Autrement dit que les matrices définies par l'éq. (A.22) génèrent l'espace des matrices $D \times D$.

A.3.2 Hamiltonien parent

Soit $|\psi\rangle$ un MPS-OBC injectif sur n qudits. Afin de définir un hamiltonien H dont $|\psi\rangle$ est un fondamental, commençons par regrouper les qudits en M blocs de taille L telles que la condition d'injectivité soit satisfaite sur chaque bloc. Définissons la fonction $\Gamma_{j,j+1}$ agissant sur des blocs premiers voisins

$$\begin{array}{rccc} \Gamma_{j,j+1} : & \mathbb{C}^{D^2} & \longrightarrow & \mathcal{H}_d^{\otimes 2L} \\ & X & \mapsto & \sum_{i_1 \dots i_L} \mathrm{Tr}\left[X B^{[j]i_j} B^{[j+1]i_{j+1}}\right] |i_j i_{j+1}\rangle \end{array} \tag{A.25}$$

L'image de $\Gamma_{j,j+1}$ est un sous-espace strict de $\mathcal{H}_d^{\otimes 2L}$. Soit $P_{j,j+1}$ le projecteur sur l'image de $\Gamma_{j,j+1}$ et $h_{j,j+1} \equiv \mathbb{I} - P_{j,j+1}$ le projecteur sur le complément orthogonal de $\mathrm{Im}\Gamma_{j,j+1}$. On peut montrer que $|\psi\rangle$ est un fondamental du hamiltonien non-frustré

$$H = \sum_j h_{j,j+1} \quad \text{ou} \quad \tilde{H} = -\sum_j P_{j,j+1} \tag{A.26}$$

autrement dit que

$$\forall j \; h_{j,j+1}|\psi\rangle = 0 \quad \text{ou} \quad \forall j \; P_{j,j+1}|\psi\rangle = +|\psi\rangle. \tag{A.27}$$

En effet, $|\psi\rangle$ est dans l'image de chaque fonction $\Gamma_{j,j+1}$ pour $X = B^{[1]} \dots B^{[j-1]} B^{[j+1]} \dots B^{[M]}$.

L'unicité demande une démonstration technique, disponible dans [129]. Elle demeure vraie pour des MPS-TI-PBC injectif. Toutefois, pour des MPS non-injectifs, on perd cette unicité. L'exemple le plus simple est l'état $|GHZ\rangle$ qui est un fondamental du modèle d'Ising ferromagnétique dont le fondamental est généré par les états $|0\rangle^{\otimes n}$ et $|1\rangle^{\otimes n}$.

Annexe B

MERA en tant qu'ansatz pour les systèmes critiques

B.1 Renormalisation dans l'espace réel : isométries

Un protocole de renormalisation [130, 131, 132] dans l'espace réel vise à transformer un système $\mathcal{L}$ sur n particules en un système $\mathcal{L}'$ sur $n' < n$ particules, tout en préservant l'information physique, c.à.d. les valeurs moyennes d'observables. Par exemple, le fondamental $|\Psi_{gs}\rangle$ d'un hamiltonien H agissant sur $\mathcal{L}$ sera transformé en un fondamental $|\Psi'_{gs}\rangle$ d'un hamiltonien H' agissant sur $\mathcal{L}'$ et

$$\langle\Psi_{gs}|\bigotimes_{i=1}^{n} o_i|\Psi_{gs}\rangle = \langle\Psi'_{gs}|\bigotimes_{i=1}^{n'} o'_i|\Psi'_{gs}\rangle \tag{B.1}$$

où o'_i est une observable renormalisée obtenue à partir de o_i.

Plus précisément, l'idée sera d'encoder l'information de k particules dans une particule effective à l'aide d'une isométrie, *i.e.*, une transformation qui préserve les distances, mais ne préserve pas la dimension des espaces de Hilbert de départ et d'arrivée. Formellement, une isométrie $w^\dagger$ enverra k qudits vers 1 qudit effectif (cf. Fig B.1)

$$w : \mathcal{H}_{d'} \to \mathcal{H}_d^{\otimes k} \quad w^\dagger w = \mathbb{I}_{\mathcal{H}_d} \quad ww^\dagger = P \tag{B.2}$$

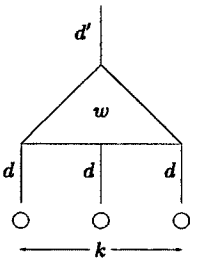


FIGURE B.1 Isométrie qui transforme k qudits vers 1 qudit effectif (ici, $k = 3$).

où P est un projecteur sur un sous-espace de $\mathcal{H}_d^{\otimes k}$ de dimension d'. En appliquant une isométrie $w^\dagger$ sur chaque bloc de k qudits, on obtient une isométrie globale $W \equiv w^{\otimes n/k}$ qui transforme n qudits en $n' = n/k$ qudits renormalisés. Formellement, l'état fondamental d'un hamiltonien H agissant sur $\mathcal{L}$ devient l'état fondamental de H' sur $\mathcal{L}'$ où

$$H' = W^\dagger H W. \tag{B.3}$$

Comment choisir les isométries w afin de satisfaire la propriété (B.1) ? Pour ce faire, il suffit que le projecteur P soit en fait le projecteur sur le support de l'état des k qudits à renormaliser. Soit $\rho = \mathrm{Tr}_{\bar{k}}\left[|\Psi_{gs}\rangle\langle\Psi_{gs}|\right]$ la matrice densité réduite de ces k qudits. Sa diagonalisation

$$\rho = \sum_{\alpha=1}^{\chi} p_\alpha |\phi_\alpha\rangle\langle\phi_\alpha| \tag{B.4}$$

montre que son support est engendré par ses vecteurs propres $\{|\phi_\alpha\rangle\}_{\alpha=1}^{\chi}$ associées à une valeur propre non-nulle. Ainsi, il faut que

$$ww^\dagger = \sum_{\alpha} |\phi_\alpha\rangle\langle\phi_\alpha|. \tag{B.5}$$

Dans ce cas, on obtient que l'isométrie globale $W \equiv w^{\otimes n/k}$ préserve l'état $|\Psi_{gs}\rangle$, *i.e.*,

$$WW^\dagger|\Psi_{gs}\rangle = |\Psi_{gs}\rangle \tag{B.6}$$

et il s'ensuit que

$$\langle \Psi_{gs} | \bigotimes_{i=1}^{n} o_i | \Psi_{gs} \rangle = \langle \Psi_{gs} | WW^\dagger \bigotimes_{i=1}^{n} o_i WW^\dagger | \Psi_{gs} \rangle \quad \text{(B.7)}$$

$$= \langle \Psi'_{gs} | W^\dagger \bigotimes_{i=1}^{n} o_i W | \Psi'_{gs} \rangle \quad \text{(B.8)}$$

$$= \langle \Psi'_{gs} | \bigotimes_{i=1}^{n'} o'_i | \Psi'_{gs} \rangle \quad \text{(B.9)}$$

où on a défini les observables renormalisées $o'_i = w^\dagger \left(\bigotimes_{i \in \mathcal{B}} o_i \right) w$.

Une question importante est la dimension d' des qudits renormalisés. Or, on vient de montrer que $d' = \chi$ où χ est le rang de Schmidt de l'état ρ d'un bloc à k qudits (éq. B.5). L'objectif d'un schéma de renormalisation est d'appliquer récursivement la procédure pour créer une séquence de modèles

$$(\mathcal{L}^{(0)}, H^{(0)}) \overset{w^{(1)}}{\mapsto} (\mathcal{L}^{(1)}, H^{(1)}) \overset{w^{(2)}}{\mapsto} (\mathcal{L}^{(2)}, H^{(2)}) \mapsto \ldots \quad \text{(B.10)}$$

où la dimension des espaces de Hilbert des systèmes $\mathcal{L}^{(\tau)}$ est décroissante. Or, pour ce faire, il faut que la dimension $\chi^{(\tau)}$ des qudits de $\mathcal{L}^{(\tau)}$ obéisse à

$$\log \chi^{(\tau+1)} < k \log \chi^{(\tau)} \quad \text{(B.11)}$$

On va donc s'intéresser au comportement de la suite $\left\{ \log \chi^{(\tau)} \right\}_{\tau \in \mathbb{N}}$.

B.2 Accumulation d'intrication

Comme mentionné précédemment en 4.2.1, le rang de Schmidt est lié à l'entropie. En particulier, on a la borne

$$\log \chi \geq S \quad \text{(B.12)}$$

Pour simplifier la discussion, on va considérer que $S \sim \log \chi$ et on va s'intéresser à la séquence des entropies $\left\{ S^{(\tau)} \right\}_{\tau \in \mathbb{N}}$.

Pour un système non-critique, on sait que pour l'état $|\Psi_{gs}\rangle$, l'entropie varie avec le nombre ℓ de particules dans le bloc avant de saturer, *i.e.*

$$S(\ell) \leq S_{\max} \in \mathcal{O}(1) \tag{B.13}$$

Au contraire, pour des systèmes critiques, cette entropie diverge logarithmiquement pour des systèmes 1D [133, 134]

$$S(\ell) \simeq \frac{c}{6} \log \ell \tag{B.14}$$

où c est la charge centrale de la théorie conforme décrivant le point critique. Or, k sites du réseau $\mathcal{L}^{(\tau)}$ correspondent à un bloc de taille $\ell = k^\tau$ sites sur $\mathcal{L}^{(0)}$. Donc, on a

$$\log \chi^{(\tau)} \sim S^{(\tau)} \in \mathcal{O}(1) \qquad \text{système 1D non-critique} \tag{B.15}$$

$$\log \chi^{(\tau)} \sim S^{(\tau)} = \frac{c}{6} \tau \log k \qquad \text{système 1D critique} \tag{B.16}$$

En particulier, on voit que pour les systèmes 1D critiques, la dimension des qudits renormalisés va grandir exponentiellement avec le nombre d'étapes de renormalisation. On peut montrer que ce phénomène est du à de l'intrication inter-bloc qui s'accumule lors du protocole de renormalisation. C'est justement à ce problème dit « accumulation d'intrication » que le MERA propose une solution.

B.3 Renormalisation d'intrication : désintricateur

Afin d'éviter l'accumulation d'intrication, l'approche MERA consiste à retirer de l'intrication entre blocs premiers voisins avant d'effectuer leur renormalisation avec des isométries. Pour ce faire, des transformations unitaires u, appelés « désintricateurs », vont agir sur le dernier qudit d'un bloc et le premier qudit d'un bloc suivant (cf. Fig B.2).

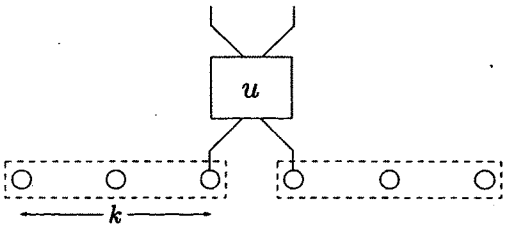


Figure B.2 Désintricateur qui retire de l'intrication entre deux blocs de k qudits.

Ainsi, les désintricateurs sont des transformations unitaires

$$u : \mathcal{H}_d^{\otimes 2} \to \mathcal{H}_d^{\otimes 2} \qquad uu^\dagger = u^\dagger u = \mathbb{I}_{\mathcal{H}_d^{\otimes 2}} \tag{B.17}$$

L'ensemble des désintricateurs correspond à une transformation unitaire $U = \bigotimes_m u^{[m]}$. La séquence de renormalisation est modifiée et on a maintenant (à comparer à l'éq. B.3)

$$H' = (UW)^\dagger H (UW) \tag{B.18}$$

et les isométries w doivent maintenant préserver le support de $\rho(u) = \mathrm{Tr}\left[U^\dagger |\Psi_{gs}\rangle\langle\Psi_{gs}| U\right]$.

Numériquement, ce protocole de renormalisation est très efficace, voir p.ex. [135] et permet d'éviter l'accumulation d'intrication. L'idée est que le MERA sera constituée d'une séquence de niveaux dont les désintricateurs puis les isométries permettront de progressivement retirer l'intrication à différentes échelles, voir la figure B.3.

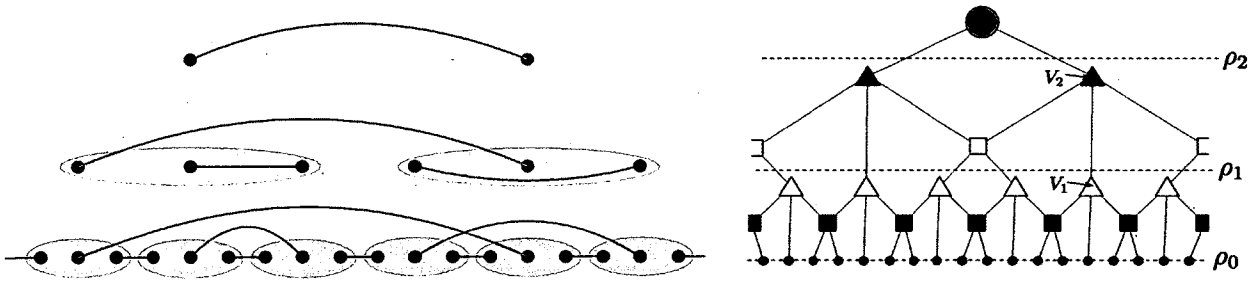


Figure B.3 État intriqué à plusieurs échelles. Chaque niveau du circuit MERA permet de gérer l'intrication (symbolisée par des liens colorés) à une échelle différente.

B.4 Circuit quantique correspondant au MERA

Le MERA peut donc être vu comme une procédure qui transforme un état $|\psi\rangle$ sur n particules vers un état $|\psi'\rangle$ sur un petit nombre de particules après plusieurs étapes de renormalisation. Chaque étape de renormalisation fait intervenir des désintricateurs, qui sont des transformations unitaires, et des isométries. Il suffit de transformer les isométries en transformation unitaire afin d'obtenir un circuit quantique dont les portes sont les désintricateurs et la version unitaire des isométries.

Notons que les isométries peuvent simplement être promues en transformations unitaires : en effet, une isométrie est définie (éq. (B.2)) comme une transformation d'un espace de Hilbert de dimension d' vers un espace de Hilbert de dimension $d^k > d'$. Pour simplifier, supposons que tous les qudits ont même dimension, *i.e.* $d' = d$. Une isométrie est alors une transformation de 1 qudit vers k qudits. Pour obtenir une unitaire, il faut que les espaces de Hilbert de départ et d'arrivée aient même dimension. L'idée est de voir l'isométrie comme une unitaire $\tilde{w}$ qui agit sur k qudits, mais que sur des états de type $|0\rangle^{\otimes k-1}|\phi\rangle$ où $|\phi\rangle$ est un état arbitraire sur un qudit (cf. Fig B.4).

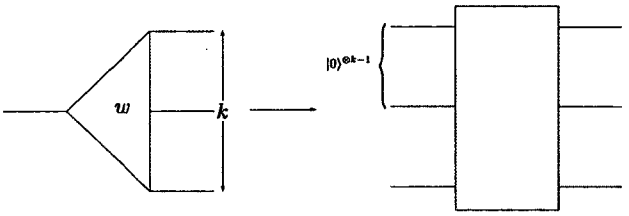


Figure B.4 Isométrie comme transformation unitaire

Il suffit alors de compléter sa définition sur le reste de l'espace afin d'obtenir une vraie transformation unitaire. Le circuit correspondant est représenté sur la figure 4.8. Ce circuit prend en entrée l'état $|\psi\rangle$ et le transforme en un état $|\psi'\rangle$ sur quelques particules. On peut même transformer l'état ainsi $|\psi'\rangle$ vers un état de référence, par exemple, l'état $|0\rangle^{\otimes m}$ grâce à une transformation unitaire agissant sur les m particules de l'état $|\psi'\rangle$. En échangeant l'entrée et la sortie du circuit quantique, on obtient alors un circuit qui construit un état $|\psi\rangle$ à partir d'un état produit $|0\rangle^{\otimes n}$. C'est cette représentation que l'on utilise dans notre article afin de faire l'apprentissage des états MERA.

Annexe C

Modèle anyonique du code torique

Nous allons maintenant montrer que les trois types de quasi-particules du code torique (cf. section 5.3.3.2), notés e pour électrique, m pour magnétique et ϵ pour la particule composite exhibent des statistiques mutuelles anyoniques. Considérons un état correspondant à une paire d'excitations électriques et une paire d'excitations magnétiques obtenues à partir du vide, *i.e.*, un état fondamental $|\Omega\rangle$, comme sur la figure C.1. Ainsi, l'état initial est

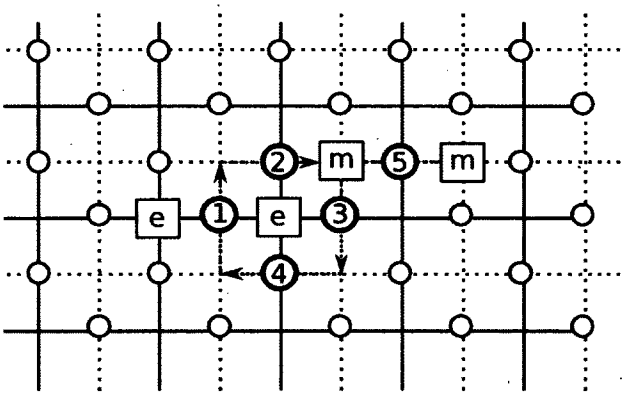


Figure C.1 Double échange d'une quasi-particule magnétique (m) avec une quasi-particule électriques (e).

$$|\psi_i\rangle = Z_1 X_5 |\Omega\rangle \tag{C.1}$$

Afin de déterminer les statistiques mutuelles entre e et m, on aimerait les échanger et voir comment la fonction d'onde est modifiée. Or, ces particules vivent l'une sur le réseau et l'autre sur le réseau

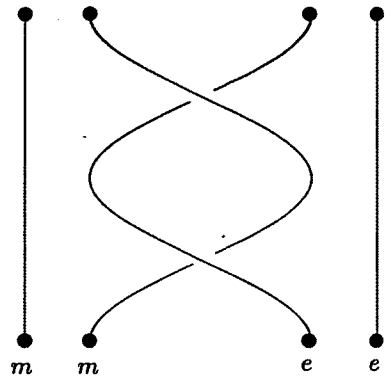


FIGURE C.2 Double-échange de particules électrique et magnétique.

réciproque. Il est donc impossible de les échanger. On va donc considérer ce qui se passe lorsque les particules sont échangées deux fois. En 3D, cette opération est triviale et la fonction d'onde ne change pas de signe. Notons que dans le référentiel d'une des particules, ce double-échange correspond à voir l'autre particule tourner autour d'elle. C'est précisément ce que nous allons faire en déplaçant une excitation m autour de l'excitation e, cf figure C.2. Ce faisant, on obtient une configuration finale $|\psi_f\rangle$ donnée par

$$
\begin{aligned}
|\psi_f\rangle &= X_2X_1X_4X_3|\psi_i\rangle && \text{(C.2)}\\
&= X_2X_1X_4X_3Z_1X_5|\Omega\rangle && \text{(C.3)}\\
&= -Z_1X_5\underbrace{X_2X_1X_4X_3}_{\epsilon\mathcal{S}}|\Omega\rangle && \text{(C.4)}\\
&= -|\psi_i\rangle && \text{(C.5)}
\end{aligned}
$$

Ainsi, la fonction d'onde prend un signe $-$ sous un double échange. Il s'agit d'une signature d'une statistique mutuelle anyonique pour les excitations e et m. Par ailleurs, le double-échange de e et m correspond aussi à une torsade de 2π pour la quasi-particule composite ϵ qui correspond donc à une phase de -1.

Qu'en est-il des statistiques propres de ces quasi-particules ? L'échange de deux quasi-particules magnétiques n'a aucun effet sur la fonction d'onde car cela revient à modifier la chaîne d'erreur qui les créent. De même pour les quasi-particules électriques. Les particules m et e sont donc des bosons. Il est possible de déduire la statistique propre de la quasi-particule composite $\epsilon = e \times m$

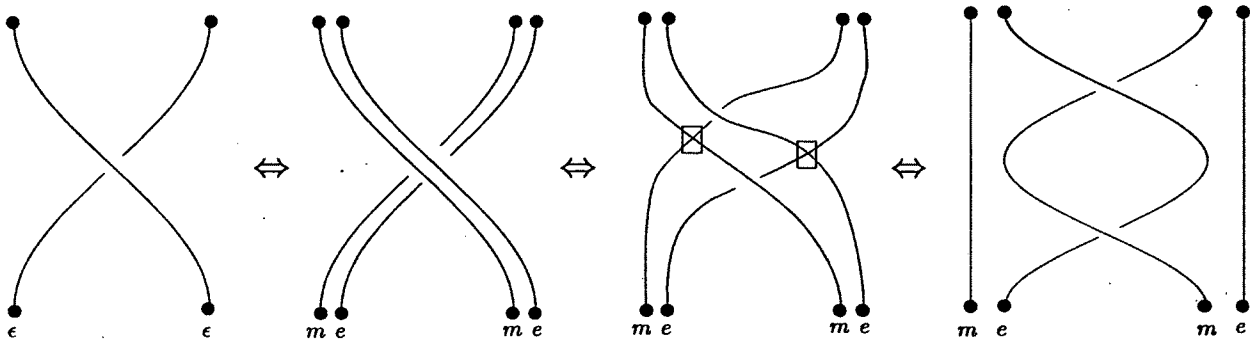


Figure C.3 Nature fermionique de la particule composite du code torique.

à partir des statistiques propres de e et m ainsi que leur statistique mutuelle. Le graphe de la figure C.3 montre qu'un échange de deux particules ϵ entraînent l'apparition d'un signe -1 sur la fonction d'onde : il s'agit donc d'un fermion.

Annexe D

Existence d'une bande non-corrigible en 2D

Dans cette annexe, nous allons montrer qu'il existe toujours un opérateur logique non-trivial en ruban dans un code CPC. Nous allons d'abord montrer son existence pour les codes stabilisateurs, ce qui ce démontre grâce au lemme de nettoyage, cf. D.1.1. Nous verrons ensuite comment ce lemme de nettoyage peut être étendu au cas plus général des codes CPC pour devenir le lemme de désintrication, cf. D.2.2. Ceci permettra de démontrer l'existence d'opérateurs ruban dans ce cas plus général.

D.1 Codes stabilisateurs

Soit un code stabilisateur défini par ses générateurs $\mathcal{S} = \langle g_1 \dots g_m \rangle$ sur un réseau Λ, tel que défini en 5.3.1. Pour un sous-groupe h des opérateurs de Pauli $\mathcal{P}$, on notera $\mathcal{C}(h)$ l'ensemble des opérateurs de Pauli qui commutent avec tous les éléments de h

$$\mathcal{C}(h) = \{P \in \mathcal{P} | \, \forall Q \in h \quad [P,Q] = 0\} \tag{D.1}$$

Il s'agit d'un sous-groupe des opérateurs de Pauli, appelé centralisateur de h. En particulier, $\mathcal{C}(\mathcal{S})$ est le sous-groupe des opérateurs logiques et $\mathcal{C}(\mathcal{S})\backslash\mathcal{S}$ est l'ensemble des opérateurs logiques non-triviaux.

D.1.1 Lemme de nettoyage

On considère une région quelconque $M \subset \Lambda$. Le lemme de nettoyage, prouvé par Bravyi et Terhal [106], montre que deux possibilités, mutuellement exclusives, existent pour cette région M.

Lemme 8 (Nettoyage). *Soit un code stabilisateur $\mathcal{S} = \langle g_1 \dots g_m \rangle$.* Pour toute région M,

1. *Soit il existe un opérateur logique non-trivial $P \in \mathcal{C}(\mathcal{S}) \backslash \mathcal{S}$ dont le support est contenu dans M*
2. *Soit, pour tout opérateur logique $P \in \mathcal{C}(S)$, il est possible de nettoyer la région M en définissant un opérateur logique équivalent $P' = PS$ qui agit trivialement sur M. De plus, l'opérateur de nettoyage $S \in \mathcal{S}$ est un produit de générateurs $S = \prod g_k$ dont le support touche M, i.e., $supp(g_k) \cap M \neq \emptyset$.*

Le cas 1 se présente par exemple dans le code d'Ising défini par les générateurs $Z_i Z_{i+1}$ où Z_k est un opérateur logique non-trivial. Le cas 2 apparaît pour toute région M topologiquement triviale du code torique où il est possible de déformer un opérateur logique en le multipliant par des générateurs dont le support touche M afin qu'il évite la région M. Ces deux cas sont représentés sur la figure D.1.

Démonstration. Cette démonstration personnelle suit les grandes lignes de [106].

Définissons deux sous-ensemble d'opérateur de Pauli qui n'agissent que sur M :
- $\mathcal{S}_M$ est le groupe généré par la restriction des stabilisateurs à la région M

$$\mathcal{S}_M = \{P \in \mathcal{P}(M) \; \exists Q \in \mathcal{P}(\Lambda \backslash M) \; PQ \in \mathcal{S}\} \tag{D.2}$$

- $\mathcal{S}(M)$ est le sous-groupe des stabilisateurs dont le support est inclus dans M

$$\mathcal{S}(M) = \{P \in \mathcal{S} \; \text{supp}(P) \subset M\} \tag{D.3}$$

Notons que ces sous-groupes sont généralement distincts. Par exemple, pour le code torique, $\mathcal{S}(M)$ sera constitué de tous les opérateurs boucles dont les boucles sont contenues dans M alors que $\mathcal{S}_M$ contiendra en plus des chaînes qui sont la restriction d'opérateurs boucles dont les boucles traversent la région M.

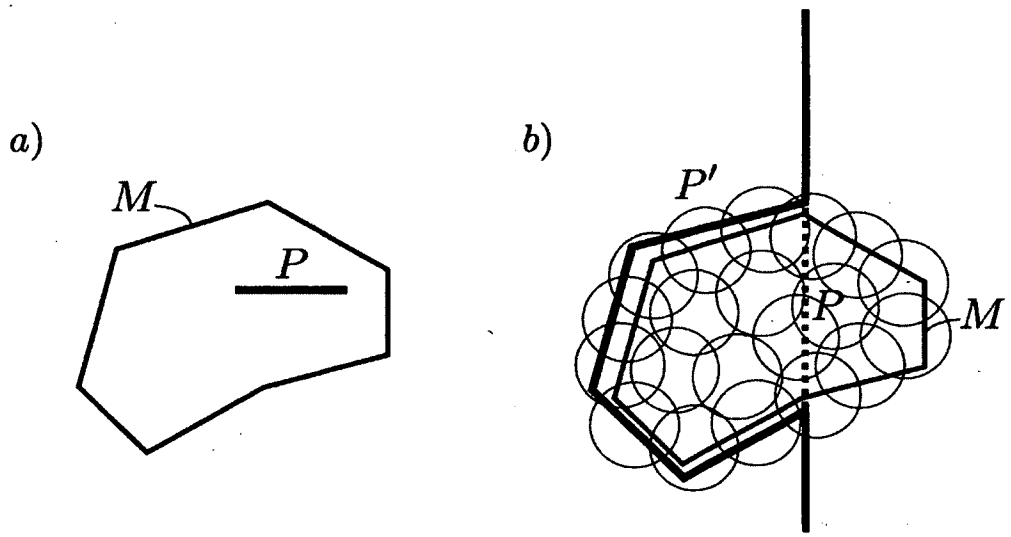


FIGURE D.1 Deux cas du le lemme de nettoyage.
En a), le cas 1 où un opérateur logique non-trivial est supporté dans M.
En b), le cas 2 où un opérateur logique P est transformé en un opérateur logique équivalent P' à l'extérieur de M en le multipliant par les générateurs dont le support intersecte M.

Par construction, les éléments de $\mathcal{S}(M)$ commutent avec ceux de $\mathcal{S}_M$. En effet, soient $x \in \mathcal{S}(M)$ et $y \in \mathcal{S}_M$. Par définition de $\mathcal{S}_M$, il existe $z \in \mathcal{P}(\Lambda \backslash M)$ tel que $yz \in \mathcal{S}$. On a alors

$$\begin{aligned} xyz &= (yz)x \quad \text{car } \mathcal{S} \text{ est abélien} \\ &= yxz \quad \text{car } x \text{ et } y \text{ ont des supports disjoints} \end{aligned}$$

L'élément z étant inversible, on a $xy = yx$ et on a montré que

$$\langle i\mathbb{I} \rangle \mathcal{S}(M) \subseteq \mathcal{C}(\mathcal{S}_M) \cap \mathcal{P}(M) \tag{D.4}$$

où $\langle i\mathbb{I} \rangle \mathcal{S}(M)$ est la version de $\mathcal{S}(M)$ où on a oublié la phase (cf. 2.2.2.2). On va maintenant distinguer le cas où l'inclusion est stricte du cas d'égalité.

Si l'inclusion est stricte, il existe un opérateur de Pauli $P \in \mathcal{C}(\mathcal{S}_M) \cap \mathcal{P}(M) \backslash \mathcal{S}(M)$. Il commute avec toute restriction de stabilisateur dans M et puisque son support est contenu dans M, il commute donc avec tous les stabilisateurs. De plus, P n'est pas un stabilisateur car sinon il appartiendrait à $\mathcal{S}(M)$. D'où, P est un opérateur logique non-trivial de support inclus dans M.

Si l'inclusion est une égalité, on peut considérer l'ensemble des opérateurs qui commutent avec $\mathcal{S}(M) = \mathcal{C}(\mathcal{S}_M) \cap \mathcal{P}(M)$ qui s'écrit

$$\mathcal{C}\left(\mathcal{C}(\mathcal{S}_M)\right) \cap \mathcal{P}(M) = \mathcal{C}\left(\mathcal{S}(M)\right) \cap \mathcal{P}(M) \tag{D.5}$$

Or, $\mathcal{C}\left(\mathcal{C}(\mathcal{S}_M)\right) = \mathcal{S}_M$: l'inclusion $\mathcal{S}_M \subset \mathcal{C}\left(\mathcal{C}(\mathcal{S}_M)\right)$ est triviale et sa réciproque est vraie (à la phase près) pour un sous-groupe des opérateurs de Pauli (il s'agit d'un cas particulier du théorème du double centralisateur). Ainsi, on a

$$\mathcal{S}_M = \mathcal{C}\left(\mathcal{S}(M)\right) \cap \mathcal{P}(M) \tag{D.6}$$

Considérons maintenant un opérateur logique $P \in \mathcal{C}(\mathcal{S})$ et en particulier sa restriction P_M sur M, *i.e.* $P = P_M R_{\bar{M}}$ car P est un opérateur de Pauli. P commute avec tout stabilisateur donc commute en particulier avec ceux supportés sur M, *i.e.* appartenant à $\mathcal{S}(M)$, donc sa restriction P_M à M commute avec tous les opérateurs de $\mathcal{S}(M)$. De plus, P_M est un opérateur de Pauli contenu dans M. Ainsi, $P_M \in \mathcal{C}\left(\mathcal{S}(M)\right) \cap \mathcal{P}(M)$ et selon l'éq. (D.6), on en déduit que $P_M \in \mathcal{S}_M$, *i.e.*, il est la restriction d'un stabilisateur à M.

Il existe donc $R_{\bar{M}}$ supporté à l'extérieur de M tel que $P_M R_{\bar{M}} = S \in \mathcal{S}$. On a alors

$$PS = P_M Q_{\bar{M}} P_M R_{\bar{M}} = \mathbb{I}_M \otimes Q_{\bar{M}} R_{\bar{M}} \tag{D.7}$$

De plus, on peut décomposer S sur les générateurs du groupe stabilisateur en distinguant les générateurs contenus dans $\bar{M}$ de ceux qui intersectent M,

$$S = \underbrace{\prod_{\text{supp}(g_k)\cap M \neq \emptyset} g_k}_{S'} \prod_{\text{supp}(g_i)\cap M = \emptyset} g_{\bar{M}} \tag{D.8}$$

Plutôt que de nettoyer P avec S, on peut se contenter de le nettoyer avec S' pour obtenir

$$PS' = \mathbb{I}_M \otimes Q_{\bar{M}} R_{\bar{M}} \prod g_{\bar{M}} \tag{D.9}$$

qui agit trivialement sur la région M. □

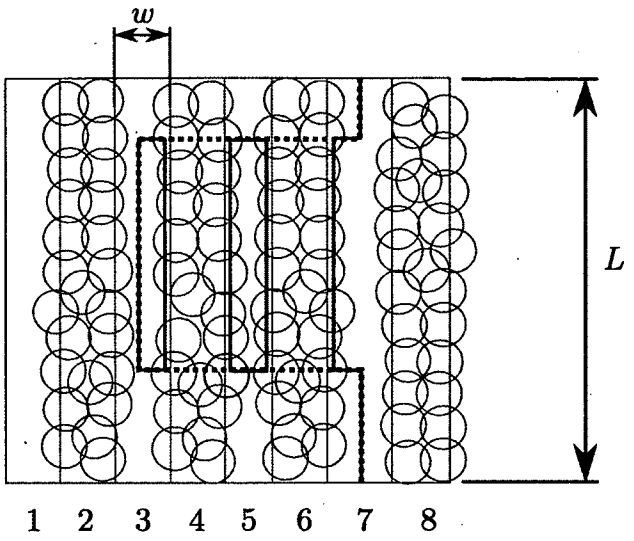


FIGURE D.2 Découpage du réseau en bandes de largeur w.
Les cercles représentent les supports des générateurs utilisés afin de nettoyer les bandes d'indice pair. Le support de l'opérateur logique non-trivial P est en trait plein alors que les traits pointillés représentent les supports des opérateurs triviaux P'_3 et P'_5 et de l'opérateur logique non-trivial P'_7.

D.1.2 Existence d'un opérateur logique

Armé du lemme de nettoyage, il est maintenant possible de montrer l'existence d'un opérateur logique non-trivial en ruban, *i.e.* supporté sur une bande de largeur $w \times L$ où w est le diamètre maximal d'un générateur du groupe stabilisateur.

Découpons le réseau Λ en bandes verticales de largeur w que nous numéroterons de 1 à m, comme sur la figure D.2 . La largeur des bandes est choisie afin que chaque générateur agisse au plus sur deux bandes. On va distinguer les bandes d'indice pair de celles d'indice impair.

Si un opérateur logique non-trivial est supporté sur une bande d'indice pair, le résultat est prouvé. Supposons que ce ne soit pas le cas, *i.e.* aucun opérateur logique non-trivial n'est supporté sur une bande d'indice pair et considérons un opérateur de Pauli logique non-trivial P (qui existe si le code est de dimension au moins 2).

En appliquant le lemme de nettoyage à chaque bande d'indice pair, comme représenté sur la Fig. D.2, on obtient un nouvel opérateur logique non-trivial P' qui est un produit d'opérateurs de Pauli sur les bandes d'indice impair (car aucun générateur n'agit non-trivialement sur deux bandes

d'indice pair)

$$P' = \bigotimes P'_{2i+1} \tag{D.10}$$

Or, chaque générateur n'agit au plus que sur une bande d'indice impair. Ainsi, puisque P' commute avec tous les générateurs, chaque opérateur P'_{2i+1} commute avec tous les générateurs qui agissent sur sa bande, mais aussi avec les autres générateurs qui n'ont pas de support commun avec lui. Donc, chaque P'_{2i+1} est un opérateur logique. De plus, puisque P' n'est pas un stabilisateur, au moins un des P'_{2i+1} n'est pas un stabilisateur. Ainsi, au moins un des P'_{2i+1} est un opérateur logique non-trivial.

D.2 Codes à projecteurs commutatifs

Nous allons maintenant montrer que les codes à projecteurs commutatifs 2D admettent toujours un opérateur logique non-trivial en ruban. La propriété clé pour montrer ce résultat est la commutativité des projecteurs. En effet, nous allons voir qu'il existe une caractérisation simple de cette propriété qui fait apparaître une décomposition de l'espace de Hilbert où les projecteurs agissent tous les deux non-trivialement. Muni de cette décomposition, il sera alors possible d'étendre les idées du lemme de nettoyage afin d'obtenir le lemme de désintrication holographique, en D.2.2, puis de montrer l'existence d'opérateurs en ruban, en D.2.

D.2.1 Décomposition de deux projecteurs qui commutent

Considérons deux projecteurs P_{AB} et P_{BC} qui commutent. La région de l'espace où ils agissent tous deux non-trivialement est notée B et elle est associée à un espace de Hilbert $\mathcal{H}_B$.

Un cas simple où les projecteurs commutent est celui où les particules de la région B peuvent être séparés en deux groupes disjoints B_L et B_R de telle façon que $P_{AB} = P_{AB_L} \otimes \mathbb{I}_{B_R}$ agit trivialement sur les particules dans B_R et $P_{BC} = \mathbb{I}_{B_L} \otimes P_{B_RC}$ agit trivialement sur les particules dans B_L. En effet, puisque les supports géométriques sont disjoints, les projecteurs commutent. L'espace de Hilbert a donc été décomposé sous la forme $\mathcal{H}_B = \mathcal{H}_{B_L} \otimes \mathcal{H}_{B_R}$.

En fait, ce cas est très général. En effet, dès que deux projecteurs commutent, il est possible de découper l'espace $\mathcal{H}_B$ où ils agissent conjointement en une somme directe de sous-espaces $\mathcal{H}_B = \bigoplus_j \mathcal{H}_{B^j}$ et d'effectuer la séparation en produit tensoriel sur chacun de ces sous-espaces

$$\mathcal{H}_B = \bigoplus_j \mathcal{H}_{B_L^j} \otimes \mathcal{H}_{B_R^j} \tag{D.11}$$

de telle façon que P_{AB} (resp. P_{BC}) agissent trivialement sur $\mathcal{H}_{B_R^j}$ (resp. $\mathcal{H}_{B_L^j}$). Ainsi, les projecteurs se décomposent sous la forme

$$P_{AB} = \sum_j (\mathbb{I}_A \otimes \Pi_j) P_{AB} (\mathbb{I}_A \otimes \Pi_j) = \sum_j P_{AB_L^j} \otimes \mathbb{I}_{B_R^j} \tag{D.12}$$

$$P_{BC} = \sum_j (\Pi_j \otimes \mathbb{I}_C) P_{BC} (\Pi_j \otimes \mathbb{I}_C) = \sum_j \mathbb{I}_{B_L^j} \otimes P_{B_R^j C} \tag{D.13}$$

$$P_{AB} P_{BC} = \sum_j P_{AB_L^j} \otimes P_{B_R^j C} \tag{D.14}$$

Notons que certains des projecteurs qui apparaissent dans la somme peuvent être nuls. Notons que cette décomposition est un fait bien connu des algèbres de matrice [136, 137] qui est très utilisé en informatique quantique, voir p.ex. [138].

D.2.1.1 Partition du réseau

Considérons un code défini par son projecteur global $P = \prod_X P_X$ qui est le produit de tous les projecteurs élémentaires P_X. Il est souvent pratique de découper le réseau en trois parties $\Lambda = ABC$ telles qu'aucun projecteur élémentaire n'agisse à la fois sur A et C. On définit alors les projecteurs $P_{AB} = \prod_{X \cap C = \emptyset} P_X$ et $P_{BC} \equiv \prod_{X \cap C \neq \emptyset} P_X$; autrement dit, chaque projecteur élémentaire dont le support intersecte A est attribué à P_{AB}, chaque projecteur dont le support intersecte C est attribué à P_{BC} et les projecteurs dont le support est contenu dans B sont, arbitrairement, attribués à P_{AB}. Ces deux projecteurs commutent et n'agissent tous deux non-trivialement que sur B. Ainsi, ils peuvent être décomposés sous la forme

$$P = P_{AB} P_{BC} = \sum_j P_{AB_L^j} \otimes P_{B_R^j C}. \tag{D.15}$$

Dans le cas où B est corrigible, cette décomposition deviendra encore plus simple.

D.2.1.2 Application à une région corrigible

Par définition, une région$M$ est corrigible s'il est possible de récupérer l'information contenue dans un état code $\rho \in \mathcal{C}$ même quand toutes les particules de la région M ont été perdues :

Définition 9 (Correctabilité). Une région M est corrigible s'il existe une transformation CPTP de correction $\mathcal{R}$ qui restaure tout état code $\rho \in \mathcal{C}$ après la perte de l'information sur la région M, *i.e.*,

$$\exists \mathcal{R}\ \forall \rho \in \mathcal{C}\ \mathcal{R}\mathrm{Tr}_M\left[\rho\right] = \rho \tag{D.16}$$

Si B est corrigible, la décomposition (D.15) prend une forme simple puisqu'elle ne peut contenir qu'un seul terme. Ainsi, la décomposition D.11 se réduit à un seul produit tensoriel $\mathcal{H}_B = \mathcal{H}_{B_L} \otimes \mathcal{H}_{B_R}$. Cette structure de produit tensoriel apparaît sur des degrés de liberté qui peuvent être délocalisés sur toute la région B. Il suffit alors d'appliquer une transformation unitaire V_B sur B pour faire apparaître le produit tensoriel au niveau physique.

Lemme 10 (ABC). *Si B est corrigible, alors* $\exists V_B \in \mathcal{L}(\mathcal{H}_B)\ V_B P V_B^\dagger = P_{AB_L} \otimes P_{B_R C}$ *où* $B_L, B_R \subset \Lambda$.

Démonstration. En effet, supposons par l'absurde qu'il y ait au moins deux termes dans la décomposition

$$P = P_{AB_L^1} \otimes P_{B_R^1 C} + P_{AB_L^2} \otimes P_{B_R^2 C} \tag{D.17}$$

et considérons deux états codes $|\Omega^{1,2}\rangle$tels que

$$P_{AB_L^i} \otimes P_{B_R^i C} |\Omega^i\rangle = +|\Omega^i\rangle \tag{D.18}$$

On sait que la décomposition est une somme directe. Ainsi, ces deux états codes sont orthogonaux, *i.e.* $\langle \Omega^1 | \Omega^2 \rangle = 0$. Or, il serait possible de distinguer ces états localement grâce à une observable $O = \sum_j \lambda_j \Pi_j$. Autrement dit, une superposition arbitraire de ces deux états $|\Omega'\rangle = \alpha|\Omega^1\rangle + \beta|\Omega^2\rangle$ ne pourrait pas être corrigée car

$$\mathcal{R}\mathrm{Tr}_B\left[\Omega'\right] = |\alpha|^2\, \mathcal{R}\mathrm{Tr}_B\left[\Omega^1\right] + |\beta|^2\, \mathcal{R}\mathrm{Tr}_B\left[\Omega^1\right] = |\alpha|^2\, \Omega^1 + |\beta|^2\, \Omega^2 \tag{D.19}$$

ce qui contredit que B soit corrigible. □

Nous allons voir que cette simple remarque appliquée à des régions judicieusement choisies permet de démontrer l'équivalent pour les codes CPC du lemme de nettoyage.

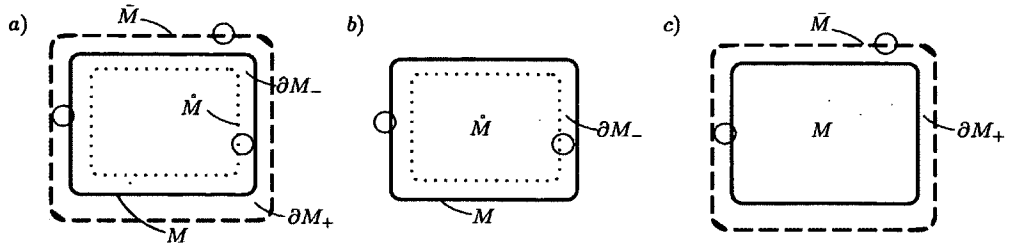


FIGURE D.3 Frontières d'une région $M \subset \Lambda$.
Les cercles représentent le support des projecteurs élémentaires. La région $\mathring{M}$ est délimitée par un trait pointillé, M par un trait plein et $\bar{M}$ par des tirets. Les sous-figures b) et c) correspondent aux partitions utilisées dans la démonstration du lemme de désintrication holographique.

D.2.2 Lemme de désintrication holographique

Pour formuler le lemme de désintrication holographique, nous allons définir les frontières intérieures et extérieures d'une région M. Cela nous permet d'appliquer le lemme 10, à deux partitions du réseau.

D.2.2.1 Définitions des frontières intérieures et extérieures

Étant donné une région M, nous allons définir ses frontières en considérant les projecteurs élémentaires qui définissent le code. Rappelons que ces projecteurs sont locaux. On appellera l'intérieur de M, noté $\mathring{M}$, l'union de tous les supports de projecteurs élémentaires contenus dans M

$$\mathring{M} = \bigcup_{\text{supp}(P_i) \subset M} \text{supp}(P_i) \tag{D.20}$$

De façon similaire, la fermeture de M, notée $\bar{M}$, sera l'union de tous les supports de projecteurs élémentaires qui touchent la région M

$$\bar{M} = \bigcup_{\text{supp}(P_i) \cap M \neq \emptyset} \text{supp}(P_i) \tag{D.21}$$

La frontière de M, notée ∂M, sera la différence entre son intérieur et sa fermeture $\partial M \equiv \bar{M} \backslash \mathring{M}$. Elle se décompose en la frontière interne $\partial M_- \equiv M \backslash \mathring{M}$ et la frontière externe $\partial M_+ \equiv \bar{M} \backslash M$. Notons que pour toute région $A \subset \Lambda$, sa région complémentaire est notée $A^c \equiv \Lambda \backslash A$. Les différentes frontières sont représentées sur la figure D.3.

D.2.2.2 Lemme

Le lemme de désintrication holographique, démontré dans [107], montre que si une région M est corrigible et que sa frontière ∂M l'est aussi, alors il est possible d'agir sur la frontière pour que l'état à l'intérieur de la région M soit indépendant de l'état global du code.

Lemme 11 (Désintrication holographique). *Soient $M \subset \Lambda$ une région corrigible dont la frontière ∂M est aussi corrigible. Il est alors possible d'appliquer une transformation unitaire agissant sur la frontière $U_{\partial M}$ sur tout état code $|\Omega\rangle$ afin de désintriquer M du reste du réseau.*

$$\exists U_{\partial M}\ \exists|\phi_M\rangle\ \forall|\Omega\rangle \in \mathcal{C} \quad U_{\partial M}|\Omega\rangle = |\phi_M\rangle \otimes |\Omega'_{M^c}\rangle \tag{D.22}$$

Démonstration. Nous allons appliquer le lemme 10 pour différents choix de région A, B et C.

- Décomposition 1 : $A = \mathring{M}$, $B = \partial M_-$, $C = M^c$, cf. Fig. D.3 b).

∂M_- est corrigible car $\partial M_- \subset M$ et M est corrigible. Donc, il existe $V_{\partial M_-}$ tel que

$$V_{\partial M_-} P V_{\partial M_-}^\dagger = P_{\mathring{M}\partial M_-^1} \otimes P_{\partial M_-^2 M^c} \tag{D.23}$$

- Décomposition 2 : $A = M$, $B = \partial M_+$, $C = \bar{M}^c$, cf. Fig. D.3 c).

∂M_+ est corrigible car $\partial M_+ \subset \partial M$ et ∂M est corrigible. Donc, il existe $V_{\partial M_+}$ tel que

$$V_{\partial M_+} P V_{\partial M_+}^\dagger = P_{M\partial M_+^1} \otimes P_{\partial M_+^2 \bar{M}^c} \tag{D.24}$$

En combinant les deux résultats intermédiaires (D.23) et (D.24), on en déduit que la transformation $V_{\partial M} = V_{\partial M_-} \otimes V_{\partial M_+}$ transforme le projecteur sur le code en

$$V_{\partial M} P V_{\partial M}^\dagger = P_{\mathring{M}\partial M_-^1} \otimes P_{\partial M_-^2 \partial M_+^1} \otimes P_{\partial M_+^2 \bar{M}^c} \tag{D.25}$$

De plus, $P_{\mathring{M}\partial M_-^1}$ doit être de rang 1, puisque $\mathring{M}\partial M_-^1 \subset M$ est corrigible. Donc, $P_{\mathring{M}\partial M_-^1} = |\mu\rangle\langle\mu|_{\mathring{M}\partial M_-^1}$. De même, $\partial M_-^2 \partial M_+^1 \subset \partial M$ est corrigible donc $P_{\partial M_-^2 \partial M_+^1} = |\nu\rangle\langle\nu|_{\partial M_-^2 \partial M_+^1}$. Schématiquement, la situation est représentée sur la Fig. D.4.

Considérons maintenant n'importe quelle transformation unitaire $W_{\partial M}$ qui agit sur la frontière ∂M et qui désintrique l'état $|\nu\rangle$ vers un état $|\alpha\rangle_{\partial M_-^2} \otimes |\beta\rangle_{\partial M_+^1}$. En posant $U_{\partial M} = W_{\partial M} V_{\partial M}$, on a

$$U_{\partial M} P U_{\partial M}^\dagger = |\mu\rangle\langle\mu|_{\mathring{M}\partial M_-^1} \otimes |\alpha\rangle\langle\alpha|_{\partial M_-^2} \otimes |\beta\rangle\langle\beta|_{\partial M_+^1} \otimes P_{\partial M_+^2 \bar{M}^c} \tag{D.26}$$

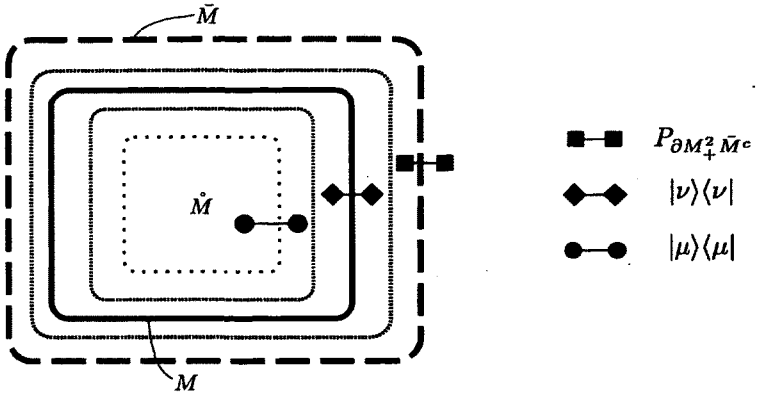


FIGURE D.4 Représentation schématique de la démonstration du lemme de désintrication holographique.
Chaque liaison représente un état potentiellement intriqué sur les régions correspondant aux extrémités de la liaison.

En particulier, n'importe quel état $|\Omega\rangle$du code sera donc transformé comme

$$U_{\partial M}|\Omega\rangle = \underbrace{|\mu\rangle_{\mathring{M}\partial M^1_-} \otimes |\alpha\rangle_{\partial M^2_-}}_{|\phi_M\rangle} \otimes \underbrace{|\beta\rangle_{\partial M^1_+} \otimes P_{\partial M^2_+ \bar{M}^c}|\Omega\rangle}_{|\Omega'_{M^c}\rangle} \tag{D.27}$$

ce qui est précisément la forme attendue. □

Nous allons maintenant prouver un corollaire de ce lemme, qui permet d'agrandir une région corrigible si sa frontière l'est aussi.

Corollaire 12 (Extension). *Si une région M et sa frontière ∂M sont corrigibles, alors $\bar{M} = M \cup \partial M$ est corrigible.*

Démonstration. Si ∂M est corrigible, il existe une opération de récupération $\mathcal{R}$ tel que pour tout état code $\Omega \equiv |\Omega\rangle\langle\Omega|$

$$\mathcal{R}\mathrm{Tr}_{\partial M}[\Omega] = \Omega \tag{D.28}$$

Or, selon le lemme, il existe $U_{\partial M}$, tel que $U_{\partial M}|\Omega\rangle = |\phi_M\rangle \otimes |\Omega'_{M^c}\rangle$ donc

$$\mathrm{Tr}_{\partial M}[\Omega] = \mathrm{Tr}_{\partial M}[\phi_M] \otimes |\Omega'\rangle\langle\Omega'|_{M^c} \tag{D.29}$$

autrement dit, toute l'information sur $|\Omega\rangle$ est à l'extérieur de $M \cup \partial M$. Ainsi, il existe une opération de récupération $\mathcal{R}'$ qui corrige $\mathrm{Tr}_{M\cup\partial M}[\Omega] = |\Omega'\rangle\langle\Omega'|_{M^c}$ qui agit en deux temps. Dans un premier

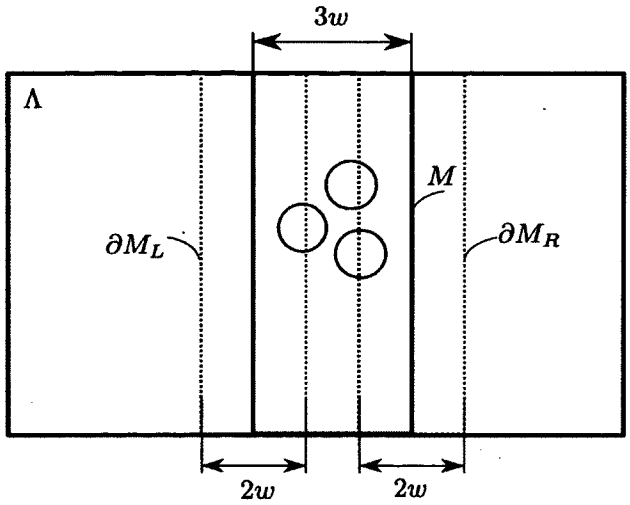


FIGURE D.5 Ruban M et ses frontières gauche ∂M_L et droite ∂M_R.

temps, $\mathcal{R}'$ introduit un système ancillaire dans l'état ϕ_M afin de reconstituer $|\phi_M\rangle \otimes |\Omega'_{M^c}\rangle$, puis dans un deuxième temps, $\mathcal{R}'$ applique la correction $\mathcal{R}$ sur l'état obtenu. Ainsi, $M \cup \partial M$ est corrigible. □

D.2.3 Preuve d'existence d'un ruban non-corrigible

Muni du lemme de désintrication et de son corollaire, nous sommes maintenant parés pour prouver l'existence de ruban non-corrigible pour les codes CPC. Nous reprenons le raisonnement de [108].

Soit M un ruban, comme représenté sur la Fig. D.5 de longueur L et de largeur $3w$. Si M n'est pas corrigible, nous avons obtenu le résultat désiré.

Sinon, dans le cas où M est corrigible, intéressons-nous à sa frontière ∂M. Cette frontière se décompose en une partie à gauche, notée ∂M_L, et à droite, notée ∂M_R. Notons que ces deux bandes ont pour largeur $2w$ et crucialement qu'aucun projecteur élémentaire n'agit sur ∂M_L et ∂M_R. Si ∂M_L ou ∂M_R n'est pas corrigible, on obtient l'existence d'un ruban non-corrigible. Sinon, $\partial M = \partial M_L \cup \partial M_R$ est corrigible car aucun projecteur élémentaire n'agit sur les deux. De plus, par le corollaire 12, on a $\bar{M} = M \cup \partial M$ qui est corrigible.

On peut donc itérer cette construction tant que les frontières gauche et droite sont corrigibles. Il faut alors distinguer le cas d'un réseau aux conditions ouvertes de celui d'un réseau aux conditions

périodiques. Pour des conditions aux frontières ouverte, la bande sera de plus en plus large et finira par couvrir le réseau Λ. Pour des conditions aux frontières périodiques, il suffit de considérer deux larges bandes M_1 et M_2 qui recouvrent le réseau $\Lambda = M_1 \cup M_2$ et telle que $\partial M_1 \cup M_2$. Dans ce cas, une modification directe du corollaire montre que $M_1 \dot{\cup} M_2 = \Lambda$ est corrigible. Dans les deux cas, on aboutit à la conclusion absurde que tout le réseau Λ est corrigible. Ainsi, ce processus ne peut continuer indéfiniment et une frontière finira par ne pas être corrigible.

Annexe E

Article : « Towards efficient decoding of classical-quantum polar codes »

Dans cette annexe, nous reproduisons l'article

> Towards efficient decoding of classical-quantum polar codes
> Mark M. Wilde, Olivier Landon-Cardinal, Patrick Hayden
> *arXiv :1302.0398*

qui a été accepté à la conférence 8th Conference on the Theory of Quantum Computation, Communication, and Cryptography et apparaîtra dans les actes de cette conférence.

Ce travail a été mené en collaboration avec Mark Wilde, post-doctorant, et Patrick Hayden, professeur au département d'informatique de McGill, durant mon doctorat. Il concerne un domaine de l'informatique quantique très différent de ceux abordés dans le reste de cette thèse, ce qui justifie qu'il n'apparaisse qu'en annexe. Toutefois, des liens forts existent avec la question de la discrimination d'états, qui a été abordée en 4.4.2.

Towards efficient decoding of classical-quantum polar codes

Mark M. Wilde[1], Olivier Landon-Cardinal[2], and Patrick Hayden[1]

1 School of Computer Science, McGill University
3480 University Street, Montreal, Quebec H3A 2A7, Canada
mwilde@gmail.com; patrick@cs.mcgill.ca
2 Département de Physique, Université de Sherbrooke
Sherbrooke, Québec J1K 2R1, Canada
olivier.landon-cardinal@usherbrooke.ca

Abstract

Known strategies for sending bits at the capacity rate over a general channel with classical input and quantum output (a cq channel) require the decoder to implement impractically complicated collective measurements. Here, we show that a fully collective strategy is not necessary in order to recover all of the information bits. In fact, when coding for a large number N uses of a cq channel W, $N \cdot I(W_{\text{acc}})$ of the bits can be recovered by a non-collective strategy which amounts to coherent quantum processing of the results of product measurements, where $I(W_{\text{acc}})$ is the accessible information of the channel W. In order to decode the other $N(I(W) - I(W_{\text{acc}}))$ bits, where $I(W)$ is the Holevo rate, our conclusion is that the receiver should employ collective measurements. We also present two other results: 1) collective Fuchs-Caves measurements (quantum likelihood ratio measurements) can be used at the receiver to achieve the Holevo rate and 2) we give an explicit form of the Helstrom measurements used in small-size polar codes. The main approach used to demonstrate these results is a quantum extension of Arikan's polar codes.

1998 ACM Subject Classification H.1.1 Systems and Information Theory, E.4 Coding and Information Theory, Error control codes

Keywords and phrases classical-quantum channel, classical-quantum polar codes, quantum likelihood ratio, quantum successive cancellation decoder

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

One of the most impressive recent developments in coding theory is the theory of polar codes [1]. These codes are provably capacity achieving, and their encoding and decoding complexities are both $O(N \log N)$, where N is the number of channel uses. Polar codes are based on the channel polarization effect, in which a recursive encoding induces a set of N synthesized channels from N instances of the original channel, such that some of the synthesized channels are nearly perfect and the others are nearly useless. The fraction of synthesized channels that is nearly perfect is equal to the capacity of the channel, and thus the coding scheme is simple: send the information bits through the synthesized channels that are nearly perfect.

An essential component of the polar coding scheme is Arikan's successive cancellation decoding algorithm [1]. This algorithm is channel dependent and operates as its name suggests: it decodes the information bits one after another, using previously decoded information to aid in constructing a test for decoding each bit in succession. In particular, the test for decoding

each information bit is a likelihood ratio test. Due to the structure in the polar encoder, there is a great deal of structure in the decoding tests, so much so that each likelihood ratio can be recursively computed. The upshot is that the complexity of the decoding algorithm is $O(N \log N)$.

Recently, there has been some effort in extending the theory of polar coding to the problem of transmission over quantum channels [23, 18, 26, 25]. In particular, these works developed the theory of polar coding for transmitting classical data over an arbitrary quantum channel [23], private classical data over an arbitrary quantum channel [25], quantum data over a quantum Pauli or erasure channel [18], and quantum data over an arbitrary quantum channel [26]. To prove that the polar coding schemes in Refs. [23, 26, 25] achieve communication rates equal to well-known formulas from quantum information theory, the authors of these works constructed a quantum successive cancellation decoder as a sequence of quantum hypothesis tests (in the spirit of Arikan [1]) and employed Sen's non-commutative union bound [20] in the error analysis. The major question left open from this effort is whether there exists an efficient implementation for a quantum successive cancellation decoder.[1,2]

In this paper, we detail our progress towards finding an efficient quantum successive cancellation decoder. The decoder outlined here is useful for decoding classical information transmitted over a channel with classical inputs and quantum outputs (known as a "classical-quantum channel" or "cq channel" for short). Since the schemes for private classical communication [25] and quantum communication [26] rely on the quantum successive cancellation decoder from Ref. [23], our results here have implications for these polar coding schemes as well. Our main result can be stated succintly as follows:

▶ **Claim 1.** *In order to achieve the symmetric Holevo capacity $I(W)$ of an arbitrary cq channel W, at most $N(I(W) - I(W_{acc}))$ of the bits require a fully collective strategy in order for them to be decoded reliably, while the other $N \cdot I(W_{acc})$ bits can be decoded efficiently and reliably in time $O(N^2)$ on a quantum computer using a product strategy that amounts to coherent quantum processing of the outcomes of product measurements.*

Although the main result of this paper might be considered modest in light of reaching the full goal stated above, it still represents non-trivial progress beyond prior research and towards answering the efficient polar decoding question. Indeed, one might think that collective measurements would be necessary in order to recover any of the bits of a message when communicating at the Holevo capacity rate, as suggested by the original work of Holevo [15], Schumacher, and Westmoreland [19] and follow-up efforts on the pure-loss bosonic channel [6, 8]. Even the recent sequential decoding schemes suggest the same [7, 20] (see also [24] for the pure-loss bosonic case). As a side note, these sequential decoding schemes require a number of measurements exponential in the number of channel uses—thus, even though the physical realization of a single one of these measurements may be within experimental

[1] By efficient, we mean that the decoder should run in $O(N^2)$ time on a quantum computer (or even better $O(N \log N)$). In computational complexity theory, "efficient" is often regarded to mean that an algorithm runs in time polynomial in the input length. However, for the demanding application of channel coding where delay should be minimized, we will consider a decoding algorithm to be "efficient" if it has a near-linear running time.

[2] Note that the scheme from Ref. [18] *does* provide an efficient $O(N \log N)$ implementation of a quantum successive cancellation decoder, essentially because sending classical states (encoded in some orthonormal basis) through a Pauli or erasure channel induces an effectively classical channel at the output (such that the resulting output states are commuting). One can then exploit a coherent version of Arikan's successive cancellation decoder to decode quantum information. Although this advance is useful, we would like to have an efficient decoder for an *arbitrary* quantum channel.

reach [17], the fact that these schemes require an exponential number of measurements excludes them from ever being practical. The previous result in Ref. [23] suggests that only a linear number of collective measurements are required to achieve the Holevo rate, and our work here demonstrates that the number of collective measurements required is at most $N(I(W) - I(W_{\text{acc}}))$.

This paper contains other results of interest. First, we prove that collective Fuchs-Caves measurements (or quantum likelihood ratio measurements) [5] suffice for achieving the Holevo information rate with a cq polar coding scheme. It was already known from Ref. [23] that a sequence of Helstrom measurements suffices for achieving this rate, so this new result just adds to the ways in which one can achieve the Holevo rate of communication. We also plot the fraction of requisite collective measurements as a function of the mean photon number of the signaling states for the case of the pure-loss bosonic channel, in order to have a sense of the physical requirements necessary for high-rate communication over this channel. As one would expect, the fraction of collective measurements needed increases as the mean photon number of the signaling states decreases—we expect this to happen since the low photon-number regime is more quantum due to the non-orthogonality of the signaling states. Finally, we detail the explicit form of a polar decoder that uses Helstrom measurements—we do this for some simple two-, four-, and eight-bit polar codes. This final result should give a sense of how one can specify these tests for larger blocklength polar codes.

The paper is organized as follows. The next section reviews background material such as cq channels, the Holevo quantity, quantum fidelity, the accessible information, and the classical fidelity (Bhattacharya parameter). Section 3 reviews the Fuchs-Caves measurement from Ref. [5] and provides a useful upper bound on the error probability of a hypothesis test that employs this measurement as the decision rule. We review classical-quantum polar codes in Section 4.1. Our first simple observation is that collective Fuchs-Caves measurements suffice for achieving the Holevo rate of communication (Section 4.2). Our main result, a justification for Claim 1, appears in Section 4.3. In Section 5, we discuss the implications of Claim 1 for the pure-loss bosonic channel. Our last result on the explicit form of the Helstrom decoder for two-, four-, and eight-bit polar codes appears in Section 6. Finally, we conclude with a summary of our results and suggest that the Schur transform might be helpful in obtaining a general solution to the problem discussed in this paper.

2 Preliminaries

A classical-quantum channel (cq channel) has a classical input and a quantum output. In this work, we only consider cq channels with binary inputs, written as

$$W : x \to \rho_x, \tag{1}$$

where W labels the channel, the input $x \in \{0, 1\}$, and ρ_x is a density operator. The symmetric Holevo information of this channel is

$$I(W) \equiv H((\rho_0 + \rho_1)/2) - [H(\rho_0) + H(\rho_1)]/2, \tag{2}$$

where $H(\sigma) \equiv -\text{Tr}\{\sigma \log_2 \sigma\}$ is the von Neumann entropy. The symmetric Holevo information gives one way to characterize the quality of a cq channel for data transmission: it is equal to one if ρ_0 is orthogonal to ρ_1 and equal to zero if $\rho_0 = \rho_1$. The quantum fidelity $F(W)$ is another parameter that characterizes the quality of a cq channel:

$$F(W) \equiv F(\rho_0, \rho_1) \equiv \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1, \tag{3}$$

where the trace norm $\|A\|_1$ of an operator A is defined as $\|A\|_1 \equiv \mathrm{Tr}\{\sqrt{A^\dagger A}\}$ [22, 16].[3] The quantum fidelity $F(W)$ is equal to one if $\rho_0 = \rho_1$ and equal to zero if ρ_0 is orthogonal to ρ_1. We have the following relationships between the symmetric Holevo information and the quantum fidelity:

$$I(W) \approx 1 \Leftrightarrow F(W) \approx 0, \tag{4}$$

$$I(W) \approx 0 \Leftrightarrow F(W) \approx 1, \tag{5}$$

which are made precise in Proposition 1 of Ref. [23].

From any cq channel, it is possible to induce a purely classical channel $p_{Y|X}(y|x)$ by having the receiver perform a quantum measurement at its output:

$$p_{Y|X}(y|x) \equiv \mathrm{Tr}\{\Lambda_y \rho_x\}, \tag{6}$$

where $\Lambda \equiv \{\Lambda_y\}$ is a positive operator-valued measure (POVM), a set of operators satisfying $\Lambda_y \geq 0$ and $\sum_y \Lambda_y = I$. Letting X be a uniform Bernoulli random variable and letting Y be the random variable corresponding to the outcome of the measurement, we can define the symmetric mutual information of the induced channel as

$$I(W, \Lambda) \equiv I(X;Y) \equiv H(X) + H(Y) - H(XY), \tag{7}$$

where H is the Shannon entropy of these random variables. The classical Bhattarcharya parameter is the statistical overlap between the resulting distributions:

$$Z(W, \Lambda) \equiv \sum_y \sqrt{p_{Y|X}(y|0)\; p_{Y|X}(y|1)}. \tag{8}$$

If one were to encode the conditional distribution $p_{Y|X}(y|x)$ along the diagonal of a matrix (so that it becomes a density operator), then it is clear that the symmetric Holevo information and fidelity of the resulting "cq channel" are equal to the symmetric mutual information and classical Bhattacharya parameter, respectively.

The symmetric accessible information is equal to the optimized symmetric mutual information:

$$I(W_{\mathrm{acc}}) \equiv \max_{\{\Lambda_y\}} I(W, \Lambda), \tag{9}$$

where the optimization is with respect to all POVMs $\Lambda = \{\Lambda_y\}$. As a consequence of the well-known Holevo bound, the symmetric Holevo information is an upper bound to the symmetric accessible information [14]:

$$I(W_{\mathrm{acc}}) \leq I(W). \tag{10}$$

3. The Fuchs-Caves Measurement

Rather than choosing a measurement to optimize the symmetric mutual information, one could also choose a measurement in such a way that it minimizes the statistical overlap

[3] Note that the quantum fidelity sometimes is defined as $\left\|\sqrt{\rho_0}\sqrt{\rho_1}\right\|_1^2$ in order for it to have the interpretation as a probability. We choose to remove the square in this work (as is often done) in order for it to reduce to the classical Bhattacharya parameter when the states are just probability distributions.

between the resulting distributions $p_{Y|X}(y|0)$ and $p_{Y|X}(y|1)$ [5]. We call such a measurement a "Fuchs-Caves" measurement since these authors proved that the minimum statistical overlap is equal to the quantum fidelity:

$$\min_{\{\Lambda_y\}} Z(W, \Lambda) = F(W). \tag{11}$$

Furthermore, they gave an explicit form for the measurement that achieves the minimum and interpreted it as a kind of "quantum likelihood ratio." Indeed, the measurement that achieves the minimum in (11) corresponds to a measurement in the eigenbasis of the following Hermitian operator:

$$\rho_0 \# \rho_1^{-1} \equiv \rho_1^{-1/2} \sqrt{\rho_1^{1/2} \rho_0 \rho_1^{1/2}} \rho_1^{-1/2}. \tag{12}$$

Diagonalizing $\rho_0 \# \rho_1^{-1}$ as

$$\rho_0 \# \rho_1^{-1} = \sum_y \lambda_y |y\rangle\langle y|, \tag{13}$$

Fuchs and Caves observed that the eigenvalues of $\rho_0 \# \rho_1^{-1}$ take the following form:

$$\lambda_y = \left(\frac{\langle y|\rho_0|y\rangle}{\langle y|\rho_1|y\rangle} \right)^{1/2}, \tag{14}$$

furthermore suggesting that this measurement is a good quantum analog of a likelihood ratio. In addition, Fuchs and Caves also observed that the operator

$$\rho_1 \# \rho_0^{-1} \equiv \rho_0^{-1/2} \sqrt{\rho_0^{1/2} \rho_1 \rho_0^{1/2}} \rho_0^{-1/2} \tag{15}$$

commutes with and is the inverse of $\rho_0 \# \rho_1^{-1}$. Thus, the eigenvectors of $\rho_1 \# \rho_0^{-1}$ are the same as those of $\rho_0 \# \rho_1^{-1}$ and its eigenvalues are the reciprocals of those of $\rho_0 \# \rho_1^{-1}$.

▶ **Lemma 2.** *When using the Fuchs-Caves measurement to distinguish ρ_0 from ρ_1, we have following upper bound on the probability of error $p_e(W)$ in terms of the quantum fidelity $F(W)$:*

$$p_e(W) \leq \tfrac{1}{2} F(W). \tag{16}$$

Proof. After performing the measurement specified by (13), the decision rule is as follows:

$$\text{decide } \rho_0 \text{ if } \lambda_y \geq 1, \tag{17}$$

$$\text{decide } \rho_1 \text{ if } \lambda_y < 1, \tag{18}$$

which corresponds to the projectors

$$\Pi_0 \equiv \sum_{y \,:\, \lambda_y \geq 1} |y\rangle\langle y|, \tag{19}$$

$$\Pi_1 = \sum_{y \,:\, \lambda_y < 1} |y\rangle\langle y|. \tag{20}$$

It is then easy to prove the bound in (16):

$$
\begin{aligned}
2\,p_e(W) &= \mathrm{Tr}\{\Pi_0\rho_1\} + \mathrm{Tr}\{\Pi_1\rho_0\} && (21)\\
&= \sum_{y\,:\,\lambda_y\geq 1}\langle y|\rho_1|y\rangle + \sum_{y\,:\,\lambda_y<1}\langle y|\rho_0|y\rangle && (22)\\
&= \sum_{y\,:\,\lambda_y\geq 1}\langle y|\rho_1|y\rangle^{1/2}\langle y|\rho_1|y\rangle^{1/2} + \sum_{y\,:\,\lambda_y<1}\langle y|\rho_0|y\rangle^{1/2}\langle y|\rho_0|y\rangle^{1/2} && (23)\\
&\leq \sum_{y\,:\,\lambda_y\geq 1}\langle y|\rho_1|y\rangle^{1/2}\langle y|\rho_0|y\rangle^{1/2} + \sum_{y\,:\,\lambda_y<1}\langle y|\rho_0|y\rangle^{1/2}\langle y|\rho_1|y\rangle^{1/2} && (24)\\
&= \sum_{y}\langle y|\rho_1|y\rangle^{1/2}\langle y|\rho_0|y\rangle^{1/2} && (25)\\
&= F(\rho_0,\rho_1) && (26)
\end{aligned}
$$

where the last equality follows from (11). ◀

4 Decoding Classical-Quantum Polar Codes

4.1 Review

Ref. [23] demonstrated how to construct synthesized versions of W, by channel combining and splitting [1]. The synthesized channels $W_N^{(i)}$ are of the following form:

$$W_N^{(i)} : u_i \to \rho_{(i),u_i}^{U_1^{i-1}B^N}, \tag{27}$$

$$\rho_{(i),u_i}^{U_1^{i-1}B^N} \equiv \sum_{u_1^{i-1}} \frac{1}{2^{i-1}} |u_1^{i-1}\rangle\langle u_1^{i-1}|^{U_1^{i-1}} \otimes \overline{\rho}_{u_1^i}^{B^N}, \tag{28}$$

$$\overline{\rho}_{u_1^i}^{B^N} \equiv \sum_{u_{i+1}^N} \frac{1}{2^{N-i}} \rho_{u^N G_N}^{B^N}, \qquad \rho_{x^N}^{B^N} \equiv \rho_{x_1}^{B_1} \otimes \cdots \otimes \rho_{x_N}^{B_N}, \tag{29}$$

where G_N is Arikan's encoding circuit matrix built from classical CNOT and permutation gates. The registers labeled by U are classical registers containing the bits u_1 through u_{i-1}, and the registers labeled by B contain the channel outputs. If the channel is classical, then these states are diagonal in the computational basis, and the above states correspond to the distributions for the synthesized channels [1]. The interpretation of $W_N^{(i)}$ is that it is the channel "seen" by the input u_i if the previous bits u_1^{i-1} are available and if the future bits u_{i+1}^N are randomized. This motivates the development of a quantum successive cancellation decoder [23] that attempts to distinguish $u_i = 0$ from $u_i = 1$ by adaptively exploiting the results of previous measurements and quantum hypothesis tests for each bit decision.

The synthesized channels $W_N^{(i)}$ polarize, in the sense that some become nearly perfect for classical data transmission while others become nearly useless. To prove this result, one can model the channel splitting and combining process as a random birth process [1, 23], and then demonstrate that the induced random birth processes corresponding to the channel parameters $I(W_N^{(i)})$ and $F(W_N^{(i)})$ are martingales that converge almost surely to zero-one valued random variables in the limit of many recursions. The following theorem characterizes the rate with which the channel polarization effect takes hold [2, 23], and it is useful in proving statements about the performance of polar codes for cq channels:

▶ **Theorem 3.** *Given a binary input cq channel W and any $\beta < 1/2$, it holds that*

$$\lim_{n\to\infty} \Pr_J\{F(W_{2^n}^{(J)}) < 2^{-2^{n\beta}}\} = I(W), \tag{30}$$

where n indicates the level of recursion for the encoding, $W_{2^n}^{(J)}$ is a random variable characterizing the J^{th} split channel, and $F(W_{2^n}^{(J)})$ is the fidelity of that channel.

Assuming knowledge of the identities of the good and bad channels, one can then construct a coding scheme based on the channel polarization effect, by dividing the synthesized channels according to the following polar coding rule:

$$\mathcal{G}_N(W,\beta) \equiv \{i \in [N] : F(W_N^{(i)}) < 2^{-N^\beta}\}, \tag{31}$$

$$\mathcal{B}_N(W,\beta) \equiv [N] \setminus \mathcal{G}_N(W,\beta), \tag{32}$$

so that $\mathcal{G}_N(W,\beta)$ is the set of "good" channels and $\mathcal{B}_N(W,\beta)$ is the set of "bad" channels. The sender then transmits the information bits through the good channels and "frozen" bits through the bad ones. A helpful assumption for error analysis is that the frozen bits are chosen uniformly at random and known to both the sender and receiver.

One of the important advances in Ref. [23] was to establish that a quantum successive cancellation decoder performs well for polar coding over classical-quantum channels with equiprobable inputs. Corresponding to the split channels $W_N^{(i)}$ in (27) are the following projectors that attempt to decide whether the input of the i^{th} split channel is zero or one:

$$\Pi_{(i),0}^{U_1^{i-1}B^N} \equiv \left\{\rho_{(i),0}^{U_1^{i-1}B^N} - \rho_{(i),1}^{U_1^{i-1}B^N} \geq 0\right\}, \tag{33}$$

$$\Pi_{(i),1}^{U_1^{i-1}B^N} \equiv I - \Pi_{(i),0}^{U_1^{i-1}B^N}, \tag{34}$$

where $\{B \geq 0\}$ denotes the projector onto the positive eigenspace of a Hermitian operator B. After some calculations, one readily sees that

$$\Pi_{(i),0}^{U_1^{i-1}B^N} = \sum_{u_1^{i-1}} |u_1^{i-1}\rangle\langle u_1^{i-1}|^{U_1^{i-1}} \otimes \Pi_{(i),u_1^{i-1}0}^{B^N}, \tag{35}$$

where

$$\Pi_{(i),1}^{U_1^{i-1}B^N} = I - \Pi_{(i),0}^{U_1^{i-1}B^N}, \tag{36}$$

$$\Pi_{(i),u_1^{i-1}0}^{B^N} \equiv \{\overline{\rho}_{u_1^{i-1}0}^{B^N} - \overline{\rho}_{u_1^{i-1}1}^{B^N} \geq 0\}, \tag{37}$$

$$\Pi_{(i),u_1^{i-1}1}^{B^N} \equiv I - \Pi_{(i),u_1^{i-1}0}^{B^N}. \tag{38}$$

The observations above lead to a decoding rule for a successive cancellation decoder similar to Arikan's [1]:

$$\hat{u}_i = \begin{cases} u_i & \text{if } i \in \mathcal{A}^c \\ h(\hat{u}_1^{i-1}) & \text{if } i \in \mathcal{A} \end{cases}, \tag{39}$$

where $h(\hat{u}_1^{i-1})$ is the outcome of the i^{th} collective measurement:

$$\{\Pi_{(i),\hat{u}_1^{i-1}0}^{B^N}, \Pi_{(i),\hat{u}_1^{i-1}1}^{B^N}\} \tag{40}$$

on the codeword received at the channel output (after $i-1$ measurements have already been performed). The set $\mathcal{A}$ labels the information bits. The measurement device outputs "0" if the outcome $\Pi_{(i),\hat{u}_1^{i-1}0}^{B^N}$ occurs and it outputs "1" otherwise. (Note that we can set $\Pi_{(i),\hat{u}_1^{i-1}u_i}^{B^N} = I$ if the bit u_i is a frozen bit.) The above sequence of measurements for the

whole bit stream u^N corresponds to a positive operator-valued measure (POVM) $\{\Lambda_{u^N}\}$ where

$$\Lambda_{u^N} \equiv \Pi^{B^N}_{(1),u_1} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1}, \tag{41}$$

and $\sum_{u_{\mathcal{A}}} \Lambda_{u^N} = I^{B^N}$. The probability of error $P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c})$ for code length N, number K of information bits, set $\mathcal{A}$ of information bits, and choice $u_{\mathcal{A}^c}$ for the frozen bits is

$$P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}) = 1 - \frac{1}{2^K} \sum_{u_{\mathcal{A}}} \mathrm{Tr}\{\Lambda_{u^N} \rho_{u^N}\}, \tag{42}$$

where we are assuming a particular choice of the bits $u_{\mathcal{A}^c}$ in the sequence of projectors $\Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1}$ and setting $\Pi^{B^N}_{(i),u_1^{i-1}u_i} = I$ if u_i is a frozen bit. The formula also assumes that the sender transmits the information sequence $u_{\mathcal{A}}$ with uniform probability 2^{-K}. The probability of error averaged over all choices of the frozen bits is then

$$P_e(N, K, \mathcal{A}) = \frac{1}{2^{N-K}} \sum_{u_{\mathcal{A}^c}} P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}). \tag{43}$$

The following proposition from Ref. [23] determines an upper bound on the average ensemble performance of polar codes with a quantum successive cancellation decoder:

▶ **Proposition 4.** *For any classical-quantum channel W with binary inputs and quantum outputs and any choice of $(N, K, \mathcal{A})$, the following bound holds*

$$P_e(N, K, \mathcal{A}) \leq 2\sqrt{\sum_{i \in \mathcal{A}} \tfrac{1}{2} F(W_N^{(i)})}. \tag{44}$$

The proposition is proved by exploiting Sen's non-commutative union bound [20] and Lemma 3.2 of Ref. [11] (which upper bounds the probability of error in a binary quantum hypothesis test by the fidelity between the test states). The bound in (44) applies provided the sender chooses the information bits $U_{\mathcal{A}}$ from a uniform distribution. Thus, by choosing the channels over which the sender transmits the information bits to be in $\mathcal{A}$ and those over which she transmits agreed-upon frozen bits to be in $\mathcal{A}^c$, we obtain that the probability of decoding error satisfies $\Pr\{\widehat{U}_{\mathcal{A}} \neq U_{\mathcal{A}}\} = o(2^{-\frac{1}{2}N^{\beta}})$, as long as the code rate obeys $R = K/N < I(W)$.

A final point that will be useful is that Ref. [23] also proved that measurements consisting of the projections

$$\left\{ \sqrt{\rho^{U_1^{i-1}B^N}_{(i),0}} - \sqrt{\rho^{U_1^{i-1}B^N}_{(i),1}} \geq 0 \right\}, \tag{45}$$

rather than those in (33)-(34), also achieve the performance stated in Proposition 4.

4.2 Collective Fuchs-Caves Measurements Achieve the Holevo Rate

Our first observation is rather simple, just being that collective Fuchs-Caves measurements can also achieve the performance stated in Proposition 4. This result follows from Lemma 2's bound on the error probability of a Fuchs-Caves measurement and by performing an error analysis similar to that in the proof of Proposition 4 of Ref. [23] given in Section V of that paper. The explicit form of a Fuchs-Caves quantum successive cancellation decoder is given by projectors of the form in (35)-(38), with the Helstrom tests replaced by Fuchs-Caves projectors as given in (19)-(20).

This result also demonstrates that there are a variety of decoding measurements that one can exploit for achieving the Holevo information rate. However, the quantum successive cancellation decoder consisting of Helstrom measurements should outperform either the measurements in (45) or the Fuchs-Caves measurements when considering finite blocklength performance because the Helstrom measurement is the optimal test for distinguishing two quantum states.

4.3 Main Result

Our main observation is a bit more subtle than the above, but it is still elementary. Nevertheless, this observation has nontrivial consequences and represents a step beyond the insights in prior work regarding decoding of classical information sent over quantum channels [15, 19, 6, 8, 7, 20, 24, 23].

We begin by considering the "Fuchs-Caves" classical channel W_{FC} induced from W by performing the Fuchs-Caves measurement on every channel output:

$$W_{\text{FC}} : x \to p_{Y|X}(y|x) = \langle y|\rho_x|y\rangle, \tag{46}$$

where the orthonormal basis $\{|y\rangle\}$ is the same as that in (13). The specification of the polar code in the previous section specializes to this induced classical channel. The code consists of a set of "good" synthesized channels $\mathcal{G}_N(W_{\text{FC}}, \beta)$ and "bad" synthesized channels $\mathcal{B}_N(W_{\text{FC}}, \beta)$, where

$$\mathcal{G}_N(W_{\text{FC}}, \beta) \equiv \{i \in [N] : F(W^{(i)}_{\text{FC},N}) = Z(W^{(i)}_{\text{FC},N}) < 2^{-N^\beta}\}, \tag{47}$$

$$\mathcal{B}_N(W_{\text{FC}}, \beta) \equiv [N] \setminus \mathcal{G}_N(W_{\text{FC}}, \beta), \tag{48}$$

and the equality $F(W^{(i)}_{\text{FC},N}) = Z(W^{(i)}_{\text{FC},N})$ holds because the induced channels are classical. Furthermore, by Theorem 3, the number of good channels in the limit that N becomes large is as follows:

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{G}_N(W_{\text{FC}}, \beta)| = I(W_{\text{FC}}). \tag{49}$$

Finally, each bit of this classical polar code can be decoded in time $O(N)$ using a recursive calculation of likelihood ratios as given in (75)-(76) of Ref. [1].[4]

Now, our main observation is the following relationship between the good channels of W_{FC} and the good channels of W:

$$\mathcal{G}_N(W_{\text{FC}}, \beta) \subseteq \mathcal{G}_N(W, \beta). \tag{50}$$

This relationship holds because of the Fuchs-Caves formula from (11). For all i, we have that

$$F(W^{(i)}_N) = \min_{\{\Lambda_y\}} Z(W^{(i)}_N, \Lambda) \leq Z(W^{(i)}_{\text{FC},N}), \tag{51}$$

where the inequality follows because the tensor-product Fuchs-Caves measurement that induces the synthesized channel $W^{(i)}_{\text{FC},N}$ is a particular kind of measurement, and so its classical statistical overlap can only be larger than that realized by the optimal measurement

[4] Note that this is the "first decoding algorithm" of Arikan. A refinement implies that all of the bits can be decoded in time $O(N \log N)$, but the first decoding algorithm is what we will use in this work.

(which in general will be a collective measurement rather than a product measurement). Now, for all $i \in \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$, we have that

$$Z(W_{\mathrm{FC},N}^{(i)}) < 2^{-N^\beta}. \tag{52}$$

This in turn implies that $F(W_N^{(i)}) < 2^{-N^\beta}$ by (51), and so for this i, we have that $i \in \mathcal{G}_N(W, \beta)$ and can conclude (50).

This observation has non-trivial implications for the structure of the polar decoder. For all of the bits in $\mathcal{G}_N(W_{\mathrm{FC}}, \beta)$, the receiver can decode them with what amounts to an effectively "product" or "non-collective" strategy,[5] while for the bits in $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$, we still require collective measurements in order for the receiver to decode them with the error probability guarantee given by (31). However, when decoding the bits in $\mathcal{G}_N(W_{\mathrm{FC}}, \beta)$, the receiver should be careful to decode them in the least destructive way possible so that Sen's non-commutative union bound is still applicable and we obtain the overall error bound guaranteed by Proposition 4. In particular, the decoder should begin by performing an isometric extension of the Fuchs-Caves measurement on each channel output:

$$\sum_y |y\rangle\langle y| \otimes |\lambda_y\rangle, \tag{53}$$

where the orthonormal basis $\{|y\rangle\}$ is from the eigendecomposition in (13) and the basis $\{|\lambda_y\rangle\}$ encodes the eigenvalues to some finite precision. Such an operation coherently copies the likelihood ratios λ_y of the Fuchs-Caves measurement into an ancillary register. The receiver then performs a reversible implementation of Arikan's decoding algorithm to process these likelihood ratios according to (75)-(76) of Ref. [1]. Finally, the receiver coherently copies the value of a single decision qubit with a CNOT gate to an ancillary register, measures the decision qubit, and "uncomputes" these operations by performing the inverse of the Arikan circuit and the inverse of the operations in (53). Figure 1 depicts these operations. The effect of these operations is to implement a projection of the channel output onto a subspace spanned by eigenvectors $|y^N\rangle = |y_1\rangle \otimes \cdots \otimes |y_N\rangle$ of the Fuchs-Caves measurements such that

$$W_{\mathrm{FC},N}^{(i)}(y^N, u_1^{i-1}|0) \geq W_{\mathrm{FC},N}^{(i)}(y^N, u_1^{i-1}|1), \tag{54}$$

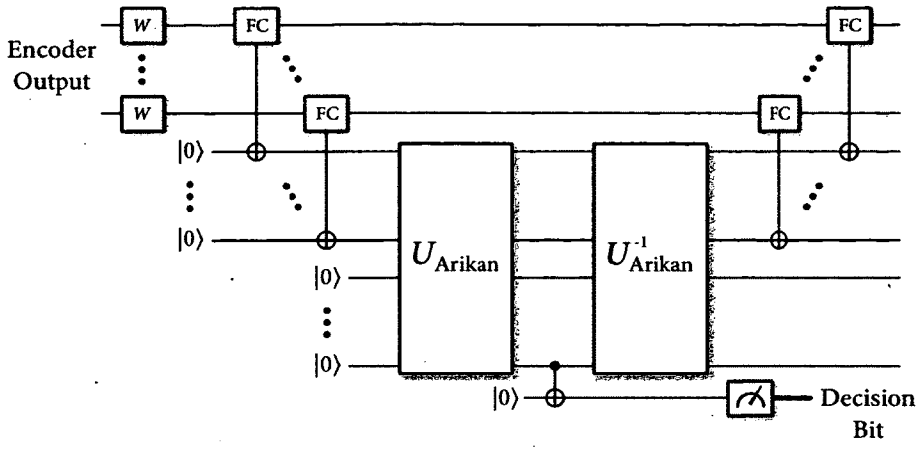
or onto the complementary subspace spanned by eigenvectors $|y^N\rangle$ such that

$$W_{\mathrm{FC},N}^{(i)}(y^N, u_1^{i-1}|0) < W_{\mathrm{FC},N}^{(i)}(y^N, u_1^{i-1}|1), \tag{55}$$

where y^N is the classical output of the Fuchs-Caves channel and u_1^{i-1} denotes the previously decoded bits. Thus, the fidelity bound from (52) is applicable and Sen's non-commutative union bound guarantees that the overall contribution of the error in decoding bit $i \in \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$ is no larger than 2^{-N^β}. The time that it takes to process each bit $i \in \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$ is $O(N)$, which is clear from the structure of the circuit and Arikan's "first decoding algorithm."

For all of the remaining bits $i \in \mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$, we still do not know whether there exists an efficient quantum algorithm for decoding them while having the error probability from Proposition 4. Thus, for now, we simply suggest for the receiver to use collective measurements to recover them.

[5] If a decoding strategy amounts to coherent implementations of product measurements followed by coherent processing of the outcomes, we still say that it is a product strategy rather than collective.



□ **Figure 1** The circuit for recovering an information bit in the set $\mathcal{G}_N(W_{\text{FC}}, \beta)$. The encoder output is fed into N instances of the channel W. The receiver acts with N of the unitaries in (53), labeled as "FC" boxes which coherently copy the likelihood ratios $\lambda_{y_1}, \ldots, \lambda_{y_N}$ into ancillary registers. The receiver then acts with a reversible implementation of Arikan's likelihood ratio computations, copies the decision bit into an ancillary register, and measures the decision bit to decode the i^{th} bit. The receiver finally performs the inverse of these operations to "clean up," i.e., to ensure that the next measurement can be performed, whether it be to decode a bit in the set $\mathcal{G}_N(W_{\text{FC}}, \beta)$ or the set $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\text{FC}}, \beta)$. The effect of this circuit is to perform the desired "gentle projection."

It should be clear from Proposition 3 and (49) that the size of the set $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\text{FC}}, \beta)$ in the limit is equal to

$$\lim_{N\to\infty} \frac{1}{N} |\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\text{FC}}, \beta)| = I(W) - I(W_{\text{FC}}). \tag{56}$$

This makes it clear that one does not require a collective strategy in order to recover all of the information bits, but a collective strategy is only required in order to bridge the gap between $I(W_{\text{FC}})$ and $I(W)$.

Observe also that similar reasoning applies to any product measurement, not just the Fuchs-Caves measurements (we focused on the Fuchs-Caves measurement due to its strong analogy with a likelihood ratio and because Arikan's decoding algorithm processes likelihood ratios). With this in mind, we could simply choose the product measurement to be the one that maximizes the accessible information, in order to maximize the number of bits that can be processed efficiently. Let W_{acc} be the classical channel induced by performing the measurement that maximizes the accessible information. One would then process the bits in $\mathcal{G}_N(W_{\text{acc}}, \beta)$ in a way very similar as described above. All of the observations above then justify Claim 1.

The reasoning also leads to a generalization of Lemma 2 that applies when using Fuchs-Caves measurements to distinguish a tensor-product state $\rho_0^{\otimes N}$ from $\rho_1^{\otimes N}$. The test consists of performing product measurements followed by classical post-processing. If one wishes to perform this test in the most delicate way possible, one could perform it as in Figure 1.

► **Lemma 5.** *When using product Fuchs-Caves measurements to distinguish $\rho_0^{\otimes N}$ from $\rho_1^{\otimes N}$, the probability of error p_e is bounded from above in terms of the quantum fidelity $F(\rho_0, \rho_1)$:*

$$p_e \leq \frac{1}{2}[F(\rho_0, \rho_1)]^N. \tag{57}$$

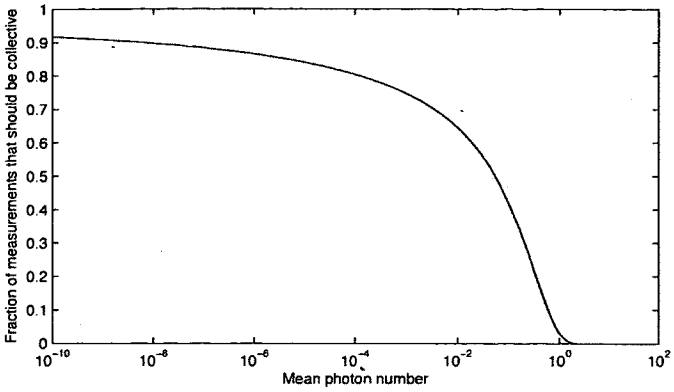


Figure 2 The fraction of collective measurements required for a polar decoder plotted as a function of the mean photon number E at the receiving end, when using a BPSK coding strategy.

Proof. The proof is very similar to the proof of Lemma 2. The test, though, consists of performing individual Fuchs-Caves measurements on the N systems, and these tests result in likelihood ratios $\lambda_{y_1}, \ldots, \lambda_{y_N}$. The decision rule is then as follows:

$$\text{decide } \rho_0^{\otimes N} \text{ if } \lambda_{y_1} \times \cdots \times \lambda_{y_N} \geq 1, \tag{58}$$

$$\text{decide } \rho_1^{\otimes N} \text{ if } \lambda_{y_1} \times \cdots \times \lambda_{y_N} < 1. \tag{59}$$

An analysis proceeding exactly as in (21)-(26) leads to the following bound:

$$\begin{aligned} 2\, p_e(W) &\leq \sum_{y_1,\ldots,y_N} \left[\langle y_1| \cdots \langle y_N|\rho_1^{\otimes N}|y_1\rangle \cdots |y_N\rangle\right]^{1/2} \left[\langle y_1| \cdots \langle y_N|\rho_0^{\otimes N}|y_1\rangle \cdots |y_N\rangle\right]^{1/2} \\ &= \sum_{y_1,\ldots,y_N} \langle y_1|\rho_1|y_1\rangle^{1/2} \cdots \langle y_N|\rho_1|y_N\rangle^{1/2} \langle y_1|\rho_0|y_1\rangle^{1/2} \cdots \langle y_N|\rho_0|y_N\rangle^{1/2} \qquad (60) \\ &= \sum_{y_1} \langle y_1|\rho_1|y_1\rangle^{1/2} \langle y_1|\rho_0|y_1\rangle^{1/2} \cdots \sum_{y_N} \langle y_N|\rho_1|y_N\rangle^{1/2} \langle y_N|\rho_0|y_N\rangle^{1/2} \qquad (61) \\ &= [F(\rho_0,\rho_1)]^N. \qquad (62) \end{aligned}$$

Furthermore, one can implement this test efficiently and non-destructively on a quantum computer as described in Figure 1. The result is to project onto two different subspaces: the one spanned by eigenvectors whose corresponding eigenvalues satisfy (58) and the other. ◀

5 Decoding the Pure-Loss Bosonic Channel

A channel of particular practical interest is the pure-loss bosonic channel. A simple physical model for this channel is a beamsplitter of transmissivity $\eta \in [0,1]$, where the sender has access to one input port, the environment injects the vacuum state into the other input port, the receiver has access to one output port, and the environment obtains the other output port. It is well known that the Holevo capacity of this channel is equal to $g(\eta N_S) \equiv (\eta N_S + 1)\log(\eta N_S + 1) - \eta N_S \log(\eta N_S)$ [6], where N_S is the mean input photon number. In the low-photon number regime, one can come very close to achieving the capacity by employing a binary phase-shift keying (BPSK) strategy (using coherent states

$|\alpha\rangle$ and $|-\alpha\rangle$ as the signaling states) [21]. The BPSK strategy induces a cq channel of the following form: $x \to |(-1)^x\alpha\rangle\langle(-1)^x\alpha|$. The symmetric Holevo rate for this channel is equal to $\chi(E) \equiv h_2\big([1+e^{-2E}]/2\big)$, where h_2 is the binary entropy and $E \equiv \eta N_S$. If the receiver performs a Helstrom measurement at every channel output, this induces a classical channel with symmetric mutual information equal to $I_{\text{Hel}}(E) \equiv 1 - h_2([1-\sqrt{1-e^{-4E}}]/2)$. (See Ref. [9], for example, for explicit calculations.) Our results in the previous section demonstrate that the fraction of information bits required to be decoded using a collective strategy is equal to $1 - I_{\text{Hel}}(E)/\chi(E)$. Figure 2 reveals that this fraction is rather small for mean photon number (MPN) larger than one, but then it rises sharply as we enter a quantum regime where the MPN is less than one. Even deep in the quantum regime at a MPN of 10^{-8}, however, roughly 10% of the bits do not require collective decoding.

6 Small Blocklength Polar Decoders

This section briefly discusses how the Helstrom measurements [12, 13] in the quantum successive cancellation decoder from Ref. [23] decompose for very small size polar codes.

6.1 Two-Bit Polar Decoder

We begin by considering the simple two-bit polar code. The channel is of the form $x \to \rho_x$, where $x \in \{0,1\}$ and ρ_x is some conditional density operator. The two-bit polar code performs the simple transformation on the input bits u_1 and u_2:

$$(u_1, u_2) \to (u_1 + u_2, u_2), \tag{63}$$

where addition is modulo 2.

The first step of the successive cancellation decoder is to recover u_1, assuming that bit u_2 is chosen uniformly at random. The optimal measurement is a Helstrom measurement, and in this case, it amounts to distinguishing between the following two states

$$\frac{1}{2}\sum_{u_2} \rho_{u_2} \otimes \rho_{u_2}, \qquad \frac{1}{2}\sum_{u_2} \rho_{u_2+1} \otimes \rho_{u_2}. \tag{64}$$

The Helstrom measurement is given by the projector onto the positive eigenspace of the difference of the two density operators above:

$$\left\{\frac{1}{2}\sum_{u_2} \rho_{u_2} \otimes \rho_{u_2} - \frac{1}{2}\sum_{u_2} \rho_{u_2+1} \otimes \rho_{u_2} \geq 0\right\} = \left\{\sum_{u_2} (\rho_{u_2} - \rho_{u_2+1}) \otimes \rho_{u_2} \geq 0\right\} \tag{65}$$

$$= \left\{\sum_{u_2} (-1)^{u_2} (\rho_0 - \rho_1) \otimes \rho_{u_2} \geq 0\right\} \tag{66}$$

$$= \left\{(\rho_0 - \rho_1) \otimes \sum_{u_2} (-1)^{u_2} \rho_{u_2} \geq 0\right\} \tag{67}$$

$$= \{(\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \geq 0\}. \tag{68}$$

Thus, this test factorizes into the parity of the individual quantum hypothesis tests $\{(\rho_0 - \rho_1) \geq 0\}$. That is, supposing that $\Pi_+ \equiv \{(\rho_0 - \rho_1) \geq 0\}$ and $\Pi_- \equiv \{(\rho_0 - \rho_1) < 0\}$, one can write the two-bit test as the product of two controlled gates

$$U_1 \equiv I_{B_1} \otimes (\Pi_+)_{B_2} \otimes I_A + I_{B_1} \otimes (\Pi_-)_{B_2} \otimes (\sigma_X)_A, \tag{69}$$

$$U_2 \equiv (\Pi_+)_{B_1} \otimes I_{B_2} \otimes I_A + (\Pi_-)_{B_1} \otimes I_{B_2} \otimes (\sigma_X)_A, \tag{70}$$

where B_1 is the first channel output, B_2 is the second channel output, and A is an ancillary system initialized to the state $|0\rangle$. The product of these two unitary gates is equal to

$$U_1 U_2 = ((\Pi_+)_{B_1} \otimes (\Pi_+)_{B_2} + (\Pi_-)_{B_1} \otimes (\Pi_-)_{B_2}) \otimes I_A + ((\Pi_-)_{B_1} \otimes (\Pi_+)_{B_2} + (\Pi_+)_{B_1} \otimes (\Pi_-)_{B_2}) \otimes (\sigma_X)_A. \quad (71)$$

The receiver would then measure the ancillary system A in order to make a decision about u_1.

Next, we determine the decoding of u_2, given that u_1 has already been decoded. By the definition of the polar encoder transformation in (63), the goal is to distinguish between the following two states:

$$\rho_{u_1} \otimes \rho_0, \qquad \rho_{u_1+1} \otimes \rho_1. \quad (72)$$

The optimal quantum hypothesis test is given by the following projector:

$$\{\rho_{u_1} \otimes \rho_0 - \rho_{u_1+1} \otimes \rho_1 \geq 0\}. \quad (73)$$

This optimal quantum hypothesis test is not factorizable into smaller tests, and indeed, it is necessary to perform a collective measurement in order to implement it. Nonetheless, Lemma 5 provides a simple implementation of the Fuchs-Caves measurement for distinguishing these two states.

6.2 Four-Bit Polar Decoder

We now consider the form of Helstrom measurements for a four-bit polar code. Recall that the input transformation for the four-bit polar code is as follows:

$$(u_1, u_2, u_3, u_4) \to (u_1 + u_2 + u_3 + u_4, u_3 + u_4, u_2 + u_4, u_4). \quad (74)$$

It is straightforward to find the form of the four different tests for decoding u_1 through u_4. (See the appendix for derivations.) The test for decoding u_1 is again a parity test:

$$\{(\rho_0 - \rho_1)^{\otimes 4} \geq 0\}. \quad (75)$$

The test for decoding u_2 given u_1 is

$$\left\{ \left(\sum_{u_3'} \rho_{u_1+u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{u_4} \otimes \rho_{u_4} \right) - \left(\sum_{u_3'} \rho_{u_1+1+u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{1+u_4} \otimes \rho_{u_4} \right) \geq 0 \right\}. \quad (76)$$

It remains unclear to us if there is a simple way to decompose the above test any further into non-collective actions (or even approximately using, e.g., the Fuchs-Caves measurement). The test for decoding u_3 given u_2 and u_1 is

$$\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes (\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}. \quad (77)$$

One could actually approximate this test "efficiently" by performing a product Fuchs-Caves measurement of the first two systems, a product Fuchs-Caves measurement of the last two, and then take the parity of the results of these two tests (of course implementing these tests coherently). The final Helstrom test for decoding u_4 given u_3, u_2, and u_1 is

$$\{\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0 - \rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1 \geq 0\}. \quad (78)$$

Clearly, it would be better to perform this last test by processing the likelihood ratios resulting from individual Fuchs-Caves measurements, rather than performing the optimal collective Helstrom measurement.

6.3 Polar Decoder for Larger Blocklengths

One can continue in the above fashion to determine the form of a quantum successive cancellation decoder that recovers each bit of an eight-bit polar code. We again try to simplify each Helstrom measurement and provide an expression for each one in Appendix B. A few tests simplify, in particular those used to recover the first bit u_1 (Eq. (101)), the fifth bit u_5 (Eq. (108), the seventh bit u_7 (Eq. (111)), and the last bit u_8 (Eq. (113)). However, for the other tests, it is unclear if they can be approximated by some combination of Helstrom and Fuchs-Caves measurements, followed by coherent post-processing.

From considering the eight-bit polar decoder, we can make several observations. For any blocklength, it is always possible to recover the first bit efficiently by calculating the parity of individual Helstrom measurements (though, this bit is always the "worst" bit, so the receiver would never actually be decoding it in practice). The receiver can always recover the last bit by performing a Fuchs-Caves measurement (this is always the "best" bit, so this should already be evident from the main observation in this paper). Furthermore, there are many bits that can be recovered by first performing Fuchs-Caves measurements, followed by the parity of these tests. Unfortunately, the fraction of these tests tends to zero in the limit of large blocklength. Thus, there still remains much to understand regarding the structure of a polar decoder.

7 Conclusion

The main result of this paper is an advance over previous schemes for decoding classical information transmitted over channels with classical inputs and quantum outputs. In particular, we have shown that $N \cdot I(W_{\text{acc}})$ of the information bits can be decoded reliably and efficiently on a quantum computer by a "non-collective" coherent decoding strategy, while closing the gap to the Holevo information rate (decoding the other $N(I(W) - I(W_{\text{acc}}))$ bits) should require a collective strategy. For the pure-loss bosonic channel, this implies that the majority of the bits transmitted can be decoded by a product strategy whenever the mean photon number is larger than one, while the fraction of collective measurements required increases sharply as the mean photon number decreases below one, marking the beginning of the quantum regime. Remarkably, even at mean photon numbers as low as 10^{-8}, roughly 10% of the bits do not require collective decoding, however. As another contribution, we have shown that a receiver can also employ collective Fuchs-Caves measurements when decoding a classical-quantum polar code. Finally, we gave the explicit form of the Helstrom measurements of a quantum successive cancellation decoder for two-, four-, and eight-bit polar codes. This should be helpful in determining the explicit form of tests for larger blocklength polar codes.

The main open question is still to determine whether all of the information bits can be efficiently decoded on a quantum computer. To answer this question, one might consider employing the Schur transform [3, 10, 4] and exploiting the structure inherent in polar codes. Unfortunately, it is not clear to us that this approach will lead to a quantum successive cancellation decoder with time complexity $O(N \log N)$ because the complexity of the Schur transform is higher than this.

We acknowledge helpful discussions with Frédéric Dupuis, Saikat Guha, Hari Krovi, David Poulin, and Joseph Renes. MMW acknowledges support from Montreal's Centre de Recherches Mathématiques. OLC aknowledges support from NSERC through a Vanier scholarship. PH acknowledges support from the Canada Research Chairs program, the Perimeter Institute, CIFAR, FQRNT's INTRIQ, NSERC, and ONR through grant N000140811249.

References

1 Erdal Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009.

2 Erdal Arikan and Emre Telatar. On the rate of channel polarization. In *Proceedings of the 2009 International Symposium on Information Theory*, pages 1493–1495, Seoul, Korea, June 2009. arXiv:0807.3806.

3 Robin Blume-Kohout, Sarah Croke, and Michael Zwolak. Ideal state discrimination with an O(1)-qubit quantum computer. arXiv:1201.6625.

4 Matthias Christandl. *The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography.* PhD thesis, University of Cambridge, April 2006. arXiv:quant-ph/0604183.

5 Christopher A. Fuchs and Carlton M. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995. arXiv:quant-ph/9604001.

6 Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, Jeffrey H. Shapiro, and Horace P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Physical Review Letters*, 92(2):027902, January 2004.

7 Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Achieving the Holevo bound via sequential measurements. *Physical Review A*, 85:012302, January 2012. arXiv:1012.0386.

8 Saikat Guha. Structured optical receivers to attain superadditive capacity and the holevo limit. *Physical Review Letters*, 106:240502, June 2011. arXiv:1101.1550.

9 Saikat Guha and Mark M. Wilde. Polar coding to achieve the holevo capacity of a pure-loss optical channel. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 546–550, Boston, Massachusetts, USA, 2012. arXiv:1202.0533.

10 Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory.* PhD thesis, Massachusetts Institute of Technology, September 2005. arXiv:quant-ph/0512255.

11 Masahito Hayashi. *Quantum Information: An Introduction.* Springer-Verlag, Berlin Heidelberg, 2006.

12 Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969.

13 Carl W. Helstrom. *Quantum Detection and Estimation Theory.* Academic, New York, 1976.

14 Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

15 Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998. arXiv:quant-ph/9611023.

16 Richard Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

17 Daniel K. L. Oi, Vaclav Potocek, and John Jeffers. Measuring nothing. July 2012. arXiv:1207.3011.

18 Joseph M. Renes, Frédéric Dupuis, and Renato Renner. Efficient polar coding of quantum information. *Physical Review Letters*, 109:050504, August 2012. arXiv:1109.3195.

19 Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

20 Pranab Sen. Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding. September 2011. arXiv:1109.0802.

21 Masaki Sohma and Osamu Hirota. Binary discretization for quantum continuous channels. *Physical Review A*, 62:052312, October 2000.

22 Armin Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

23 Mark M. Wilde and Saikat Guha. Polar codes for classical-quantum channels. *IEEE Transactions on Information Theory*, 59(2):1175–1187, February 2013. arXiv:1109.2591.

24 Mark M. Wilde, Saikat Guha, Si-Hui Tan, and Seth Lloyd. Explicit capacity-achieving receivers for optical communication and quantum reading. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 551–555, Boston, Massachusetts, USA, July 2012. arXiv:1202.0518.

25 Mark M. Wilde and Joseph M. Renes. Polar codes for private classical communication. In *Proceedings of the 2012 International Symposium on Information Theory and its Applications*, Honolulu, Hawaii, USA, October 2012. arXiv:1203.5794.

26 Mark M. Wilde and Joseph M. Renes. Quantum polar codes for arbitrary channels. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 334–338, Boston, Massachusetts, USA, July 2012. arXiv:1201.2906.

A Derivations for the Four-Bit Polar Decoder Measurements

The four-bit polar encoder amounts to the following transformation:

$$(u_1, u_2, u_3, u_4) \to (u_1 + u_2 + u_3 + u_4, u_3 + u_4, u_2 + u_4, u_4). \tag{79}$$

A.1 Recovering u_1

Let us first determine how the quantum successive cancellation decoder (QSCD) recovers the bit u_1, assuming that u_2, u_3, and u_4 are chosen uniformly at random. The test aims to distinguish between the following two states:

$$\frac{1}{2^3} \sum_{u_2,u_3,u_4} \rho_{u_2+u_3+u_4} \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4}, \tag{80}$$

$$\frac{1}{2^3} \sum_{u_2,u_3,u_4} \rho_{u_2+u_3+u_4+1} \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4}, \tag{81}$$

and it performs the following projection:

$$\left\{ \sum_{u_2,u_3,u_4} (\rho_{u_2+u_3+u_4} - \rho_{u_2+u_3+u_4+1}) \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\}$$

$$= \left\{ \sum_{u_2,u_3,u_4} (-1)^{u_2+u_3+u_4} (\rho_0 - \rho_1) \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\} \tag{82}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2,u_3,u_4} (-1)^{u_2+u_3+u_4} \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\} \tag{83}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2,u_3,u_4} (-1)^{u_3+u_4} \rho_{u_3+u_4} \otimes (-1)^{u_2+u_4} \rho_{u_2+u_4} \otimes (-1)^{u_4} \rho_{u_4} \geq 0 \right\} \tag{84}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2', u_3', u_4'} (-1)^{u_2'} \rho_{u_2'} \otimes (-1)^{u_3'} \rho_{u_3'} \otimes (-1)^{u_4'} \rho_{u_4'} \geq 0 \right\} \tag{85}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2'} (-1)^{u_2'} \rho_{u_2'} \otimes \sum_{u_3'} (-1)^{u_3'} \rho_{u_3'} \otimes \sum_{u_4'} (-1)^{u_4'} \rho_{u_4'} \geq 0 \right\} \tag{86}$$

$$= \{ (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \geq 0 \}. \tag{87}$$

Thus, this first test nicely factors as the parity of the four individual tests $\{(\rho_0 - \rho_1) \geq 0\}$.

A.2 Recovering u_2 given u_1

We now determine how the quantum successive cancellation decoder recovers u_2 given u_1, while randomizing over u_3 and u_4. The aim is to distinguish between the following two states:

$$\frac{1}{2^2} \sum_{u_3, u_4} \rho_{u_1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_4}, \tag{88}$$

$$\frac{1}{2^2} \sum_{u_3, u_4} \rho_{u_1 + 1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_4}, \tag{89}$$

which translates to a projection of the following form:

$$\left\{ \sum_{u_3, u_4} \rho_{u_1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_4} - \rho_{u_1 + 1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_4} \geq 0 \right\}. \tag{90}$$

Define $u_3' = u_3 + u_4$ and the above becomes

$$\left\{ \sum_{u_3', u_4} \rho_{u_1 + u_3'} \otimes \rho_{u_3'} \otimes \rho_{u_4} \otimes \rho_{u_4} - \rho_{u_1 + 1 + u_3'} \otimes \rho_{u_3'} \otimes \rho_{1 + u_4} \otimes \rho_{u_4} \geq 0 \right\}$$

$$= \left\{ \left(\sum_{u_3'} \rho_{u_1 + u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{u_4} \otimes \rho_{u_4} \right) - \left(\sum_{u_3'} \rho_{u_1 + 1 + u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{1 + u_4} \otimes \rho_{u_4} \right) \geq 0 \right\}. \tag{91}$$

A.3 Recovering u_3 given u_2 and u_1

Let us determine how the QSCD recovers u_3 given u_2 and u_1, while randomizing over u_4. The test distinguishes between the following two states:

$$\frac{1}{2} \sum_{u_4} \rho_{u_1 + u_2 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_2 + u_4} \otimes \rho_{u_4}, \tag{92}$$

$$\frac{1}{2} \sum_{u_4} \rho_{u_1 + u_2 + 1 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_2 + u_4} \otimes \rho_{u_4}, \tag{93}$$

and amounts to a projector of the following form:

$$\left\{\sum_{u_4} \rho_{u_1+u_2+u_4} \otimes \rho_{u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} - \sum_{u_4} \rho_{u_1+u_2+1+u_4} \otimes \rho_{1+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\}$$

$$= \left\{\sum_{u_4} (\rho_{u_1+u_2+u_4} \otimes \rho_{u_4} - \rho_{u_1+u_2+1+u_4} \otimes \rho_{1+u_4}) \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{94}$$

$$= \left\{\sum_{u_4} (-1)^{u_4} (\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{95}$$

$$= \left\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes \sum_{u_4} (-1)^{u_4} \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{96}$$

$$= \{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes (\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}. \tag{97}$$

Thus, this test nicely factorizes as the parity of two tests $\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \geq 0\}$ and $\{(\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}$.

A.4 Recovering u_4 given u_3, u_2, and u_1

Finally, we determine how the QSCD recovers u_4 given all of the previous bits. The test in this case just aims to distinguish the following states:

$$\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0, \tag{98}$$

$$\rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1, \tag{99}$$

and amounts to the following projection:

$$\{\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0 - \rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1 \geq 0\}. \tag{100}$$

B Measurements for the Eight-Bit Polar Decoder

Here, we provide the form of a quantum successive cancellation decoder that recovers each bit of an eight-bit polar code. Full derivations of the results in this section are available from the authors upon request.

B.1 Recovering u_1

The test to recover the first bit u_1 is simply the parity of eight individual Helstrom measurements:

$$\{(\rho_0 - \rho_1)^{\otimes 8} \geq 0\}. \tag{101}$$

B.2 Recovering u_2 given u_1

The test to recover bit u_2 given u_1 projects onto the positive eigenspace of the difference of

$$\left(\sum_{u_3', u_4', u_5'} \rho_{u_1+u_3'+u_4'+u_5'} \otimes \rho_{u_3'} \otimes \rho_{u_4'} \otimes \rho_{u_5'}\right) \otimes \left(\sum_{u_6', u_7', u_8'} \rho_{u_6'+u_7'+u_8'} \otimes \rho_{u_6'} \otimes \rho_{u_7'} \otimes \rho_{u_8'}\right) \tag{102}$$

and

$$\left(\sum_{u_3',u_4',u_5'} \rho_{u_1+u_3'+u_4'+u_5'} \otimes \rho_{u_3'} \otimes \rho_{u_4'} \otimes \rho_{u_5'}\right) \otimes \left(\sum_{u_6',u_7',u_8'} \rho_{u_6'+u_7'+u_8'} \otimes \rho_{u_6'} \otimes \rho_{u_7'} \otimes \rho_{u_8'}\right). \tag{103}$$

As such, it is not clear to us how one could approximate this test as some combination of Helstrom and Fuchs-Caves tests.

B.3 Recovering u_3 given u_2, and u_1

The test to recover bit u_3 given u_1 and u_2 is equal to the parity of the following two tests:

$$\left\{ \begin{array}{c} \left(\sum_{u_4'} \rho_{u_1+u_2+u_4'} \otimes \rho_{u_4'}\right) \otimes \left(\sum_{u_5'} \rho_{u_5'} \otimes \rho_{u_5'}\right) \\ -\left(\sum_{u_4'} \rho_{u_1+u_2+1+u_4'} \otimes \rho_{u_4'}\right) \otimes \left(\sum_{u_5'} \rho_{1+u_5'} \otimes \rho_{u_5'}\right) \geq 0 \end{array} \right\}, \tag{104}$$

$$\left\{ \begin{array}{c} \left(\sum_{u_6'} \rho_{u_2+u_6'} \otimes \rho_{u_6'}\right) \otimes \left(\sum_{u_8''} \rho_{u_8''} \otimes \rho_{u_8''}\right) \\ -\left(\sum_{u_6'} \rho_{u_2+u_6'+1} \otimes \rho_{u_6'}\right) \otimes \left(\sum_{u_8''} \rho_{1+u_8''} \otimes \rho_{u_8''}\right) \geq 0 \end{array} \right\}. \tag{105}$$

It is again unclear to us how to decompose this measurement further.

B.4 Recovering u_4 given u_3, u_2, and u_1

The test to recover bit u_4 given u_1, u_2, and u_3 projects onto the positive eigenspace of the difference of

$$\left(\sum_{u_5'} \rho_{u_1+u_2+u_3+u_5'} \otimes \rho_{u_5'}\right) \otimes \left(\sum_{u_6'} \rho_{u_3+u_6'} \otimes \rho_{u_6'}\right) \otimes \left(\sum_{u_7'} \rho_{u_2+u_7'} \otimes \rho_{u_7'}\right) \otimes \left(\sum_{u_8'} \rho_{u_8'} \otimes \rho_{u_8'}\right) \tag{106}$$

and

$$\left(\sum_{u_5'} \rho_{u_1+u_2+u_3+1+u_5'} \otimes \rho_{u_5'}\right) \otimes \left(\sum_{u_6'} \rho_{u_3+1+u_6'} \otimes \rho_{u_6'}\right) \otimes \left(\sum_{u_7'} \rho_{u_2+1+u_7'} \otimes \rho_{u_7'}\right) \otimes \left(\sum_{u_8'} \rho_{1+u_8'} \otimes \rho_{u_8'}\right) \tag{107}$$

Again, this one remains unclear how to decompose further.

B.5 Recovering u_5 given $u_4, \ldots, u_1$

The test to recover bit u_5 given u_1 through u_4 is equal to

$$\left\{ \begin{array}{c} (\rho_{u_1+u_2+u_3+u_4} \otimes \rho_0 - \rho_{u_1+u_2+u_3+u_4+1} \otimes \rho_1) \otimes (\rho_{u_3+u_4} \otimes \rho_0 - \rho_{u_3+u_4+1} \otimes \rho_1) \\ \otimes (\rho_{u_2+u_4} \otimes \rho_0 - \rho_{u_2+u_4+1} \otimes \rho_1) \otimes (\rho_{u_4} \otimes \rho_0 - \rho_{u_4+1} \otimes \rho_1) \geq 0 \end{array} \right\}. \tag{108}$$

It is easy to see that one could approximate this test by first performing four Fuchs-Caves measurements on adjacent pairs of channel outputs and taking the parity of these tests.

B.6 Recovering u_6 given $u_5, \ldots, u_1$

The test to recover bit u_6 given u_1 through u_5 is a projection onto the positive eigenspace of the difference of

$$\left(\sum_{u_7'} \rho_{u_1+\cdots+u_5+u_7'} \otimes \rho_{u_5+u_7'} \otimes \rho_{u_3+u_4+u_7'} \otimes \rho_{u_7'}\right) \otimes \left(\sum_{u_8'} \rho_{u_2+u_4+u_8'} \otimes \rho_{u_8'} \otimes \rho_{u_4+u_8'} \otimes \rho_{u_8'}\right) \tag{109}$$

and

$$\left(\sum_{u_7'} \rho_{u_1+\cdots+u_5+1+u_7'} \otimes \rho_{u_5+1+u_7'} \otimes \rho_{u_3+u_4+u_7'} \otimes \rho_{u_7'}\right) \otimes \left(\sum_{u_8'} \rho_{u_2+u_4+1+u_8'} \otimes \rho_{1+u_8'} \otimes \rho_{u_4+u_8'} \otimes \rho_{u_8'}\right). \tag{110}$$

A simple decomposition of this test remains unclear.

B.7 Recovering u_7 given $u_6, \ldots, u_1$

The test for recovering bit u_7 given the previous ones is

$$\left\{ \begin{array}{c} (\rho_{u_1+\cdots+u_6} \otimes \rho_{u_5+u_6} \otimes \rho_{u_3+u_4} \otimes \rho_0 - \rho_{u_1+\cdots+u_6+1} \otimes \rho_{u_5+u_6+1} \otimes \rho_{u_3+u_4+1} \otimes \rho_1) \otimes \\ (\rho_{u_2+u_4+u_6} \otimes \rho_{u_6} \otimes \rho_{u_4} \otimes \rho_0 - \rho_{u_2+u_4+u_6+1} \otimes \rho_{u_6+1} \otimes \rho_{u_4+1} \otimes \rho_1) \geq 0 \end{array} \right\}, \tag{111}$$

which is clearly implementable by performing a Fuchs-Caves measurement on the first four qubits and the last four, and than taking the parity of these two tests.

B.8 Recovering u_8 given $u_7, \ldots, u_1$

The final test for recovering the last bit u_8 given all others is a projection onto the positive eigenspace of the difference of

$$\rho_{u_1+\cdots+u_7} \otimes \rho_{u_5+u_6+u_7} \otimes \rho_{u_3+u_4+u_7} \otimes \rho_{u_7} \otimes \rho_{u_2+u_4+u_6} \otimes \rho_{u_6} \otimes \rho_{u_4} \otimes \rho_0, \tag{112}$$

and

$$\rho_{u_1+\cdots+u_7+1} \otimes \rho_{u_5+u_6+u_7+1} \otimes \rho_{u_3+u_4+u_7+1} \otimes \rho_{u_7+1} \otimes \rho_{u_2+u_4+u_6+1} \otimes \rho_{u_6+1} \otimes \rho_{u_4+1} \otimes \rho_1. \tag{113}$$

It is clear that we can approximate this test with a Fuchs-Caves measurement.

Bibliographie

[1] Marcus P. da Silva, Olivier Landon-Cardinal, et David Poulin. *Phys. Rev. Lett.* **107**, 210404 (2011).

[2] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, et Yi-Kai Liu. *Nature Communications* **1**, 149 (2010).

[3] Olivier Landon-Cardinal et David Poulin. *New J. Phys.* **14**(8), 085004 (2012).

[4] Olivier Landon-Cardinal et David Poulin. *Phys. Rev. Lett.* **110**, 090502 (2013).

[5] Mark M Wilde, Olivier Landon-Cardinal, et Patrick Hayden. *arXiv preprint arXiv :1302.0398* .

[6] John Preskill. *Introduction To Quantum Computation And Information*, chapitre Fault-Tolerant Quantum Computation. World Scientific (1998). arXiv :quant-ph/9712048.

[7] Michael A Nielsen et Isaac L Chuang. *Quantum computation and quantum information.* Cambridge : Cambridge University Press, (2000).

[8] N.D. Mermin. *Quantum Computer Science : An Introduction.* Cambridge University Press, (2007).

[9] Michel Le Bellac. *Physique quantique (2e édition).* Savoirs Actuels. EDP Sciences, (2012).

[10] Olivier Landon-Cardinal. Mémoire de Maîtrise, Université de Montréal, (2010).

[11] T D Ladd, F Jelezko, R Laflamme, Y Nakamura, C Monroe, et J L O'Brien. *Nature* **464**(7285), 45–53 (2010).

[12] David Poulin, Angie Qarry, Rolando Somma, et Frank Verstraete. *Phys. Rev. Lett.* **106**, 170501 (2011).

[13] Miguel Aguado et Guifré Vidal. *Phys. Rev. Lett.* **100**, 070404 (2008).

[14] Christopher A. Fuchs et Asher Peres. *Phys. Rev. A* **53**, 2038–2045 (1996).

[15] David P Divincenzo. *Fortschritte der Physik* **48**, 771–783 (2000).

[16] A Fedorov, L Steffen, M Baur, M P da Silva, et A Wallraff. *Nature* **481**(7380), 170–2 (2012).

[17] Jonathan Simon, Waseem S Bakr, Ruichao Ma, M Eric Tai, Philipp M Preiss, et Markus Greiner. *Nature* **472**(7343), 307–312 (2011).

[18] Thomas Monz, Philipp Schindler, Julio T. Barreiro, Michael Chwalla, Daniel Nigg, William A. Coish, Maximilian Harlander, Wolfgang Hänsel, Markus Hennrich, et Rainer Blatt. *Phys. Rev. Lett.* **106**, 130506 (2011).

[19] H Häffner, W Hänsel, C F Roos, J Benhelm, D Chek-al Kar, M Chwalla, T Körber, U D Rapol, M Riebe, P O Schmidt, C Becher, O Gühne, W Dür, et R Blatt. *Nature* **438**(7068), 643–6 (2005).

[20] Daniel Nigg, Julio T. Barreiro, Philipp Schindler, Masoud Mohseni, Thomas Monz, Michael Chwalla, Markus Hennrich, et Rainer Blatt. *Phys. Rev. Lett.* **110**, 060403 (2013).

[21] Hideharu Mikami, Yongmin Li, Kyosuke Fukuoka, et Takayoshi Kobayashi. *Phys. Rev. Lett.* **95**(15), 2–5 (2005).

[22] K. Resch, P. Walther, et a. Zeilinger. *Phys. Rev. Lett.* **94**(7) (2005).

[23] J. O'Brien, G. Pryde, a. Gilchrist, D. James, N. Langford, T. Ralph, et a. White. *Phys. Rev. Lett.* **93**(8), 080502 (2004).

[24] N. Boulant, E. Fortunato, M. Pravia, G. Teklemariam, D. Cory, et T. Havel. *Phys. Rev. A* **65**(2), 024302 (2002).

[25] Andrew Childs, Isaac Chuang, et Debbie Leung. *Phys. Rev. A* **64**(1), 012314 (2001).

[26] Matthias Steffen, M. Ansmann, R. McDermott, N. Katz, Radoslaw Bialczak, Erik Lucero, Matthew Neeley, E. Weig, a. Cleland, et John Martinis. *Phys. Rev. Lett.* **97**(5), 4–7 (2006).

[27] Yanwen Wu, Xiaoqin Li, L. Duan, D. Steel, et D. Gammon. *Phys. Rev. Lett.* **96**(8), 087402 (2006).

[28] G de Lange, Z H Wang, D Ristè, V V Dobrovitski, et R Hanson. *Science (New York, N.Y.)* **330**(6000), 60–3 (2010).

[29] P Neumann, N Mizuochi, F Rempp, P Hemmer, H Watanabe, S Yamasaki, V Jacques, T Gaebel, F Jelezko, et J Wrachtrup. *Science (New York, N.Y.)* **320**(5881), 1326–9 (2008).

[30] F. Jelezko, T. Gaebel, I. Popa, M. Domhan, A. Gruber, et J. Wrachtrup. *Phys. Rev. Lett.* **93**(13), 130501 (2004).

[31] CM Dawson et MA Nielsen. *Quantum Information and Computation* **6**(1), 81–95 (2006).

[32] A. Y. Kitaev. *Russ. Math. Surv.* **52 (6)**, 1191–1249 (1997).

[33] Robin Blume-Kohout. *New J. Phys.* **12**(4), 043034 (2010).

[34] Joseph B Altepeter, Daniel F V James, et Paul G Kwiat. Dans *Quantum State Estimation*, Matteo Paris et Jaroslav Reháček, volume 145, chapitre 4, 113–145. Springer Berlin Heidelberg (2004).

[35] Z Hradil. *Phys. Rev. A* **55**(3), 1561–1564 (1997).

[36] Daniel F. V. James, Paul G. Kwiat, William J. Munro, et Andrew G. White. *Phys. Rev. A* **64**(5), 052312 (2001).

[37] Robin Blume-Kohout, Jun O. S. Yin, et S. J. van Enk. *Phys. Rev. Lett.* **105**(17), 170501 (2010).

[38] Matthias Christandl et Renato Renner. *Phys. Rev. Lett.* **109**(12), 120403 (2012).

[39] Seth T Merkel, Jay M Gambetta, John A Smolin, S Poletto, A D Corcoles, B. R. Johnson, Colm A. Ryan, et Matthias Steffen. *arXiv* , 1–10 (2012).

[40] M. Mohseni, a. Rezakhani, et D. Lidar. *Phys. Rev. A* **77**(3), 1–15 (2008).

[41] Steven T. Flammia et Yi-Kai Liu. *Phys. Rev. Lett.* **106**(23), 230501 (2011).

[42] Scott Aaronson, David Chen, Daniel Gottesman, Vincent Liew, et Ashley Montanaro. in preparation, http ://www.physics.usyd.edu.au/quantum/Coogee2013/Presentations/.

[43] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, et Jens Eisert. *Phys. Rev. Lett.* **105**(15), 150401 (2010).

[44] EJ Candès et MB Wakin. *Signal Processing Magazine, IEEE* (March 2008), 21–30 (2008).

[45] Scott Aaronson. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences* **463**(2088), 3089–3114 (2007).

[46] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. Blakestad, J. Jost, C. Langer, R. Ozeri, S. Seidelin, et D. Wineland. *Phys. Rev. A* **77**(1), 012307 (2008).

[47] Daniel Gottesman. alk at InternationalConference on Group Theoretic Methods in Physics, (1998). arXiv :quant-ph/9807006.

[48] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G Cory, et Raymond Laflamme. *Science (New York, N.Y.)* **317**(5846), 1893–6 (2007).

[49] Easwar Magesan, J. M. Gambetta, et Joseph Emerson. *Phys. Rev. Lett.* **106**(18), 180504 (2011).

[50] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm a. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George a. Keefe, Mary B. Rothwell, Thomas a. Ohki, Mark B. Ketchen, et M. Steffen. *Phys. Rev. Lett.* **109**(8), 080505 (2012).

[51] Ariel Bendersky, Fernando Pastawski, et Juan Paz. *Phys. Rev. Lett.* **100**(19), 190403 (2008).

[52] Siddhartha Chib et Edward Greenberg. *The American Statistician* **49**(4), 327–335 (1995).

[53] L. Steffen, M. P. da Silva, A. Fedorov, M. Baur, et A. Wallraff. *Phys. Rev. Lett.* **108**, 260506 (2012).

[54] Yaoyun Shi. *arXiv preprint quant-ph/0205115* (2002).

[55] W Forrest Stinespring. Dans *Proc. Amer. Math. Soc*, volume 6, 19, (1955).

[56] Michał Horodecki, Paweł Horodecki, et Ryszard Horodecki. *Phys. Rev. A* **60**, 1888–1898 (1999).

[57] P.W. Shor. Dans *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, 124–134, (1994).

[58] Maarten Van Den Nest. *Quantum Info. Comput.* **11**(9-10), 784–812 (2011).

[59] John Von Neumann. *Mathematische grundlagen der quantenmechanik*, volume 38. Springer, (1932).

[60] M. B. Hastings. *Phys. Rev. Lett.* **93**, 140402 (2004).

[61] F. Verstraete et J. I. Cirac. *Phys. Rev. B* **73**, 094423 (2006).

[62] R. B. Laughlin. *Phys. Rev. Lett.* **50**, 1395–1398 (1983).

[63] J. Bardeen, L. N. Cooper, et J. R. Schrieffer. *Phys. Rev.* **108**, 1175–1204 (1957).

[64] J. Bardeen, L. N. Cooper, et J. R. Schrieffer. *Phys. Rev.* **106**, 162–164 (1957).

[65] Dorit Aharonov, Daniel Gottesman, Sandy Irani, et Julia Kempe. *Commun. Math. Phys.* **287**(1), 41–65 (2009).

[66] D. Gottesman et S. Irani. Dans *Foundations of Computer Science, 2009. FOCS '09. 50th Annual IEEE Symposium on*, 95–104, (2009).

[67] M B Hastings. *Journal of Statistical Mechanics : Theory and Experiment* **2007**(08), P08024 (2007).

[68] Itai Arad, Zeph Landau, et Umesh Vazirani. *Phys. Rev. B* **85**, 195145 (2012).

[69] Itai Arad, Alexei Kitaev, Zeph Landau, et Umesh Vazirani. *arXiv preprint arXiv :1301.1162* (2013).

[70] Steven R. White. *Phys. Rev. Lett.* **69**(19), 2863–2866 (1992).

[71] S Östlund et Stefan Rommer. *Phys. Rev. Lett.* **75**(19), 3537–3540 (1995).

[72] J Dukelsky, M.A. Martin-Delgado, T Nishino, et G Sierra. *EPL (Europhysics ...* **43**, 457 (1998).

[73] Ulrich Schollwöck. *Ann. Phys.* **326**(1), 96–192 (2011).

[74] R Blume-Kohout, Sarah Croke, et Michael Zwolak. *arXiv preprint arXiv :1201.6625* (1), 1–5 (2012).

[75] Carl W Helstrom. (1976).

[76] G. Vidal. *Phys. Rev. Lett.* **101**(11), 1–4 (2008).

[77] F Verstraete et JI Cirac. *arXiv preprint cond-mat/0407066* , 1–5 (2004).

[78] Norbert Schuch, Michael M. Wolf, Frank Verstraete, et J. Ignacio Cirac. *Phys. Rev. Lett.* **98**(14), 140506–4 (2007).

[79] J. Ignacio Cirac, Didier Poilblanc, Norbert Schuch, et Frank Verstraete. *Phys. Rev. B* **83**(24), 245134 (2011).

[80] Alexei Kitaev. *Ann. Phys.* **303**(1), 2–30 (2003).

[81] S. Sachdev. *Quantum phase transitions.* Cambridge University Press, (2001).

[82] Xie Chen, Zheng-Cheng Gu, et Xiao-Gang Wen. *Phys. Rev. B* **82**(15), 155138 (2010).

[83] XG Wen. *Phys. Rev. B* **40**(10), 7387 (1989).

[84] Daniel S. Rokhsar et Steven A. Kivelson. *Phys. Rev. Lett.* **61**(20), 2376–2379 (1988).

[85] N. Bonesteel. *Phys. Rev. B* **40**(13), 8954–8960 (1989).

[86] G. Misguich, D. Serban, et V. Pasquier. *Phys. Rev. Lett.* **89**(13), 137202 (2002).

[87] P.W. Anderson. *Mater. Res. Bull.* **8**(2), 153 – 160 (1973).

[88] D. J. Thouless. *Phys. Rev. B* **36**(13), 7187–7189 (1987).

[89] Emanuel Knill et Raymond Laflamme. *Phys. Rev. A* **55**, 900–911 (1997).

[90] S. Bravyi, M. Hastings, et F. Verstraete. *Phys. Rev. Lett.* **97**(5), 1–4 (2006).

[91] Carlos Fernández-González, Norbert Schuch, Michael M. Wolf, J. Ignacio Cirac, et David Pérez-García. *Phys. Rev. Lett.* **109**(26), 260401 (2012).

[92] Frank Wilczek. *Phys. Rev. Lett.* **49**(14), 957–959 (1982).

[93] J. Preskill. *Lecture Notes for Physics* **219** (2004).

[94] Alexei Kitaev et John Preskill. *Phys. Rev. Lett.* **96**(11), 110404 (2006).

[95] Michael Levin et Xiao-Gang Wen. *Phys. Rev. Lett.* **96**(11), 110405 (2006).

[96] Daniel Gottesman. *Stabilizer Codes and Quantum Error Correction.* Thèse de Doctorat, California Institute of Technology, (1997).

[97] Peter W. Shor. *Phys. Rev. A* **52**, R2493–R2496 (1995).

[98] Jeongwan Haah. *Phys. Rev. A* **83**, 042330 (2011).

[99] Guillaume Duclos-Cianci. Mémoire de Maîtrise, Département de physique, Université de Sherbrooke., (2010).

[100] IB Damgard, S Fehr, L Salvail, et C Schaffner. Dans *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, 449–458. IEEE, (2005).

[101] Mario Berta, Matthias Christandl, Roger Colbeck, Joseph M Renes, et Renato Renner. *Nature Physics* (2010).

[102] H. Bombin et M. A. Martin-Delgado. *Phys. Rev. Lett.* **97**, 180501 (2006).

[103] Michael Levin et Xiao-Gang Wen. *Phys. Rev. B* **71**(4), 1–21 (2005).

[104] Robert Koenig, Greg Kuperberg, et Ben W. Reichardt. *Annals of Physics* **325**(12), 2707 – 2749 (2010).

[105] Sergey Bravyi, Matthew B Hastings, et Spyridon Michalakis. *Journal of Mathematical Physics* **51**, 093512 (2010).

[106] Sergey Bravyi et Barbara Terhal. *New J. Phys.* **11**(4), 043029 (2009).

[107] Sergey Bravyi, David Poulin, et Barbara Terhal. *Phys. Rev. Lett.* **104**, 050503 (2010).

[108] Jeongwan Haah et John Preskill. *Phys. Rev. A* **86**, 032308 (2012).

[109] R Alicki, M Fannes, et M Horodecki. *Journal of Physics A : Mathematical and Theoretical* **42**(6), 065303 (2009).

[110] Alastair Kay et Roger Colbeck. *arXiv preprint arXiv :0810.3557* (2008).

[111] L. Lamata, J. León, D. Pérez-García, D. Salgado, et E. Solano. *Phys. Rev. Lett.* **101**, 180506 (2008).

[112] Sergey Bravyi et Jeongwan Haah. *arXiv preprint arXiv :1112.3252* (2011).

[113] Sergey Bravyi et Jeongwan Haah. *Phys. Rev. Lett.* **107**, 150504 (2011).

[114] Jeongwan Haah. *arXiv preprint arXiv :1204.1063* (September), 1–39 (2012).

[115] Kamil Michnicki. *arXiv preprint arXiv :1208.3496* , 1–19 (2012).

[116] Eric Dennis, Alexei Kitaev, Andrew Landahl, et John Preskill. *Journal of Mathematical Physics* **43**, 4452 (2002).

[117] Robert Alicki, Michal Horodecki, Pawel Horodecki, et Ryszard Horodecki. *Open Systems & Information Dynamics* **17**(01), 1–20 (2010).

[118] Tian-Heng Han, Joel S Helton, Shaoyan Chu, Daniel G Nocera, Jose A Rodriguez-Rivera, Collin Broholm, et Young S Lee. *Nature* **492**(7429), 406–410 (2012).

[119] Alioscia Hamma, Claudio Castelnovo, et Claudio Chamon. *Phys. Rev. B* **79**(24), 1–6 (2009).

[120] Stefano Chesi, Beat Röthlisberger, et Daniel Loss. *Phys. Rev. A* **82**(2), 022305 (2010).

[121] Fabio L Pedrocchi, Adrian Hutter, James R Wootton, et Daniel Loss. *arXiv preprint arXiv :1209.5289* (2012).

[122] Cyril Stark, Lode Pollet, Ata ç Imamoğlu, et Renato Renner. *Phys. Rev. Lett.* **107**, 030504 (2011).

[123] James R. Wootton et Jiannis K. Pachos. *Phys. Rev. Lett.* **107**, 030503 (2011).

[124] Fernando Pastawski, Lucas Clemente, et Juan Ignacio Cirac. *Phys. Rev. A* **83**, 012304 (2011).

[125] Matthew B Hastings. *arXiv preprint arXiv :1207.1671* (2012).

[126] L. Cincio et G. Vidal. *Phys. Rev. Lett.* **110**, 067208 (2013).

[127] Michael P Zaletel, Roger SK Mong, et Frank Pollmann. *arXiv preprint arXiv :1211.3733* (2012).

[128] Guifré Vidal. *Phys. Rev. Lett.* **91**(14), 12–15 (2003).

[129] D. Perez-Garcia, F. Verstraete, M.M. Wolf, et J.I. Cirac. *Quantum Inf. Comput.* **7**(5), 401 (2007).

[130] Leo P. Kadanoff, Wolfgang Götze, David Hamblen, Robert Hecht, E. A. S. Lewis, V. V. Palciauskas, Martin Rayl, J. Swift, David Aspnes, et Joseph Kane. *Rev. Mod. Phys.* **39**, 395–431 (1967).

[131] Kenneth G. Wilson. *Rev. Mod. Phys.* **47**, 773–840 (1975).

[132] Michael E. Fisher. *Rev. Mod. Phys.* **70**, 653–681 (1998).

[133] Christoph Holzhey, Finn Larsen, et Frank Wilczek. *Nucl. Phys. B* **424**(3), 443–467 (1994).

[134] G. Vidal, J. I. Latorre, E. Rico, et A. Kitaev. *Phys. Rev. Lett.* **90**, 227902 (2003).

[135] G. Evenbly et G. Vidal. *Phys. Rev. B* **79**(14), 144108 (2009).

[136] Kenneth R Davidson. *C*-algebras by example*, volume 6. American Mathematical Soc., (1996).

[137] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer, (1997).

[138] S. Bravyi et M. Vyalyi. *Quantum Inf. Comput.* **5**, 187 (2005).