Opinions Libres
le blog d'Olivier Ezratty

Decode Quantum avec Valentin Savin du CEA-Leti

Dans le 70e épisode des entretiens Decode Quantum, Fanny Bouton et moi-même sommes avec **Valentin Savin** du CEA-Leti, pour parler du sujet de la correction d'erreurs. Cet entretien est aussi diffusé par **Frenchweb**.

Valentin Savin a un master en mathématiques de l'ENS Lyon et de l'Université Joseph Fourier de Grenoble, puis a réalisé une thèse de doctorat dans cette même université en 2001. Entre 2002 et 2004, il était post-doc à l'Institut de Mathématiques de l'académie roumaine. Depuis 2005, il est au CEA-LETI à Grenoble, d'abord comme post-doc puis comme chercheur. Ses recherches portent sur les codes de correction d'erreurs classiques et quantiques, à la fois pour les communications, pour le stockage de données et pour le calcul. Il copilote de nombreux projets de recherche collaborative européens dans le domaine, notamment pour créer des systèmes quantiques à tolérance de pannes.

?

Ses débuts dans le quantique ont démarré assez tardivement. Il a commencé à travailler dans le domaine des codes de correction d'erreur classiques en 2005, notamment dans les télécoms et s'est intéressé aux codes quantiques en 2017 suite à une rencontre avec Maud Vinet, à l'époque également au CEA-LETI. Il fallait comprendre comment faire de la correction d'erreurs quantique. Il a tout de suite apprécié le formalisme mathématique associé. Il avait lancé une thèse sur le sujet sur des **codes polaires quantiques** puis s'est intéressé à d'autres aspects, en lien avec la tolérance aux fautes.

Sa thèse **Propriétés de bidualité des espaces de modules en terme de cohomologie de groupes**, 2001, a été soutenue à Université Joseph Fourier à Grenoble. Elle portait sur la géométrie complexe et algébrique. Il s'intéressait à des espaces classifiables par des invariants continus, liée à la notion d'espaces de modules. Cela n'avait pas de liens évidents avec les codes de correction d'erreurs classiques. Dans les années 1990, des constructions de correction d'erreur classiques à base de géométrie algébrique qui avaient un bon rendement

pouvaient être envisagées. Cela faisait principalement suite aux travaux de **Tsfasman, Vl?du? et Zink** qui ont permis de construire pour la première fois des familles de codes dépassant la borne de Gilbert-Varshamov. Il avait envie de regarder les choses de manière plus appliquée.

Les codes de correction d'erreur sont partout dans le numérique classique, notamment pour corriger les transmissions à faible rapport signal à bruit, qui nécessitent de la redondance dans la communication. Grâce aux **travaux de Claude Shannon**, on sait caractériser le niveau de redondance minimal permettant de communiquer de manière fiable sur un canal bruité. C'est ce qu'on appelle la capacité du canal, ou encore la limite de Shannon. Au début des années 2000, il restait beaucoup à faire, concernant entre autres la construction de familles de codes capables d'atteindre ou approcher cette limite théorique. Si les constructions dites « algébriques » (codes BCH, codes de Reed-Solomon, code de Goppa), cherchaient avant tout à optimiser la distance minimale du code, cherchant ensuite une algorithme de décodage efficace, les années 1990 voient apparaitre une nouvelle famille de codes, dont le mérite consistait particulièrement dans la façon de les décoder, plutôt que dans leurs propriétés en termes de distance minimale. Il s'agit des **turbo codes** inventés en France par **Claude Berrou** et **Alain Glavieux**, qui pouvaient faire bien mieux que l'existant, en utilisant un système de décodage ingénieux, avec un décodage itératif qui fonctionnait bien. Cela a aussi permis à la communauté de se rappeler les travaux plus anciens de **Robert G. Gallager**, qui avait proposé dès 1962 des codes LDPC classiques, également décodables par du décodage itératif.

Au début des année 2000, des **travaux de Richardson, Shokrollahi et Urbanke** ont permis de montrer que la construction de ces codes pouvait être optimisée, de façon à approcher de très près la limite théorique de Shannon pour une large gamme de canaux. Ils ont intégré depuis un bon nombre de standards de télécommunication et ont remplacé petit à petit les turbo codes qui dominaient.

Pour la fiabilisation des mémoires ou systèmes de stockage, le principe de la correction d'erreur est similaire à celui de la transmission des données, avec toutefois des modèles de canaux et des contraintes spécifiques. On pourra retrouver des mécanismes de correction d'erreurs à base de codes LDPC, notamment dans les mémoires non-volatiles. Pour la partie de calcul classique, c'est différent. Si la fiabilisation du calcul à travers des mécanismes de correction d'erreur était effectivement envisagée au début du calcul classique, la très grande fiabilité de la technologie CMOS, qui s'est imposée à partir des années 1980, a rendu inutile leur utilisation. La question s'est néanmoins reposée au début des années 2010 avec l'augmentation très forte de la densité d'intégration, et la fin pressentie de la loi de Moore (**projet i-RISC**, 2013-2016). Le niveau de redondance acceptable est relativement faible, à la différence du calcul quantique, car dans le cas classique on dispose déjà d'une technologie fiable. Ainsi, réduire la taille des transistors à une échelle où ils perdraient de leur fiabilité, pour ensuite ajouter la redondance nécessaire à la correction d'erreur, doit être mis en balance contre un nœud technologique moins réduit, mais fiable. Une question intéressante est celle de la **tolérance aux fautes des décodeurs** eux-mêmes.

Valentin explique la différence entre la correction d'erreurs en informatique classique et en informatique quantique. Le problème du décodage quantique est souvent ramené à du décodage classique. La formalisation par les codes stabilisateurs qui permet d'établir des liens avec les codes classiques non binaires. Les codes CSS avec deux codes classiques binaires utilisés pour le décodage des erreurs X (bit flip) et Z (phase flip). Ce qui est intrinsèquement quantique est l'extraction du syndrome d'erreur, que l'on décode ensuite pour identifier l'erreur qui s'est produite. Ce décodage est toujours classique, puis une correction quantique est appliquée aux qubits via une boucle de rétroaction.

On évoque l'historique des codes de correction d'erreur avec le papier fondateur de **Peter Shor** en 1996, le Steane code avec 5 qubits, l'introduction des codes stabilisateur par Gottesman, puis le code torique de de Kitaev et sa version code de surface. La plupart des développements ont été réalisés par des physiciens, alors que les codes classiques sont venus des mathématiques.

Les chercheurs se sont alors intéressés aux codes LDPC quantiques, ou qLDPC, mais pour des raisons différentes du classique. Dans le classique, l'intérêt principal des codes LDPC vient de la manière dont ils sont décodés. La "sparsité" de la matrice de parité (creuse) permet de mettre en place des algorithmes de décodage itératifs par passage de message, qui sont non seulement capables de fournir une excellente capacité de correction, mais sont aussi très bien adaptés aux implémentations matérielles, d'où leur succès. Dans le quantique, l'intérêt principal des codes LDPC réside dans le fait que l'extraction du syndrome d'erreur nécessite de faire parler entre eux qu'un faible nombre de qubits. Cela génère une faible propagation d'erreurs et implique que l'extraction du syndrome peut être faite de manière tolérante aux fautes. Par contre le décodage des codes qLDPC, c'est-à-dire, l'algorithme de décodage classique qui consiste à déterminer l'erreur qui s'est produite à partir du syndrome extrait, est plus difficile qu'en classique. Cela est dû au problème de dégénérescence des codes qLDPC, qui rend les algorithmes par passage de message inefficaces, nécessitant l'ajout de techniques de post-traitement (e.g., **stabilizer inactivation**). Ce problème n'est pas anodin, car au-delà d'être précis (identifier correctement l'erreur qui s'est produite), le décodeur doit répondre à des contraintes spécifiques du système quantique, comme la latence ou la consommation de puissance, et peu de techniques utilisées actuellement pour le post-traitement sont compatibles avec ces contraintes (e.g., **check-agnosia**). La classe des qLDPC est très large, elle comprend en particulier les constructions topologiques (e.g., codes de surface, codes couleurs) et dans une certaine mesure d'autres familles de codes, comme les Bacon-Shor.

Une autre famille de codes très intéressants sont les codes polaires. Ils atteignent la capacité « one-shot » de tout canal quantique (**purely quantum polar codes**) et, à la différence des codes qLDPC, sont équipés d'un algorithme de décodage efficace, hérité de leurs homologues classiques. Par contre la préparation tolérante aux fautes de ces codes est plus difficile, avec néanmoins des avancées récentes pour les codes polaires encodant un qubit logique (e.g., **Q1-codes**, **factory-based preparation**).

Le code de surface s'est imposé par la connectivité requise, dite du « plus proche voisin ». Les codes qLDPC (ou polaires) ont besoin d'interactions longue distance entre les qubits, mais ils deviennent intéressants grâce au fait qu'ils sont capable d'encoder plus de qubits logiques ou qu'ils possèdent une capacité de correction supérieure à celle du code de surface. IBM prévoit d'utiliser une version de **codes LDPC définis via des graphes bi-planaires** dans les qubits supraconducteur, afin de contourner le problème de connectivité à distance. Dans les atomes neutres, on peut envisager le déplacement des qubits, comme **l'a montré récemment Mikhail Lukin** à Harvard.

Du côté des performances, on recherche des codes avec une profondeur constante du circuit quantique d'extraction du syndrome d'erreur. C'est le cas des codes de surface et qLDPC. Le décodage se fait sur une fenêtre temporelle de syndromes, afin de garantir la tolérance aux fautes pendant l'extraction du syndrome. La distance minimale du code est de l'ordre de grandeur de la racine carrée du nombre de qubits pour le code de surface, mais elle peut atteindre le même ordre de grandeur que le nombre de qubits pour les qLDPC (e.g., **quantum Tanner codes**). Nous évoquons la notion de longueur du code, liée au nombre de qubits physiques pour encoder un qubit logique. N qubits physiques peuvent encoder plusieurs qubits logiques K. N/K est un indicateur de la redondance.

Le décodage des syndromes d'erreurs demande un algorithme classique qui est complexe. Le temps est limité pour le faire, qui correspond au temps nécessaire pour faire l'extraction du syndrome, soit quelques centaines de ns en général. On a besoin de parallélisme et d'algorithmes en complexité linéaire et pas polynomiale.

Comment corrige-t-on les erreurs corrélées ? Il y a relativement peu de choses qui sont faites au niveau de la correction d'erreur (**des modèles de corrélation sont sortis récemment**), moins que ce qui est fait pour les modèles de bruit biaisé. Les erreurs corrélées avec des qubits voisins sont liées au contrôle et au crosstalk et pas seulement aux rayons cosmiques.

L'objet de ses recherches au CEA-Leti et leur lien avec l'écosystème du calcul quantique français. Au Leti, les qubits de spin de silicium y sont développés depuis plusieurs années. Il a un lien avec la startup Quobly via des projets collaboratifs et avec le CNRS et UGA. Et aussi au niveau national via le PEPR quantique, avec les projets **NISQ2LSQ** (coordonné par Anthony Leverrier d'Inria) et PRESQUILE (en lien avec la technologie des qubits silicium CMOS). Au niveau européen, le projet **Quantera EQUIP** coordonné par Valentin regarde ces problématiques en s'intéressant au décodage temps-réel ainsi qu'à la mitigation d'erreurs. L'écosystème français de la correction d'erreurs comprend aussi des équipes à Bordeaux ainsi que Christophe Vuillot d'Inria Nancy, ou encore Omar Fawzi d'Inria Lyon.

Voilà pour cet épisode. Dans le suivant, nous accueillons **Jan Goetz**, le CEO d'IQM !

Cet article a été publié le 2 mai 2024 et édité en PDF le 2 mai 2024.