



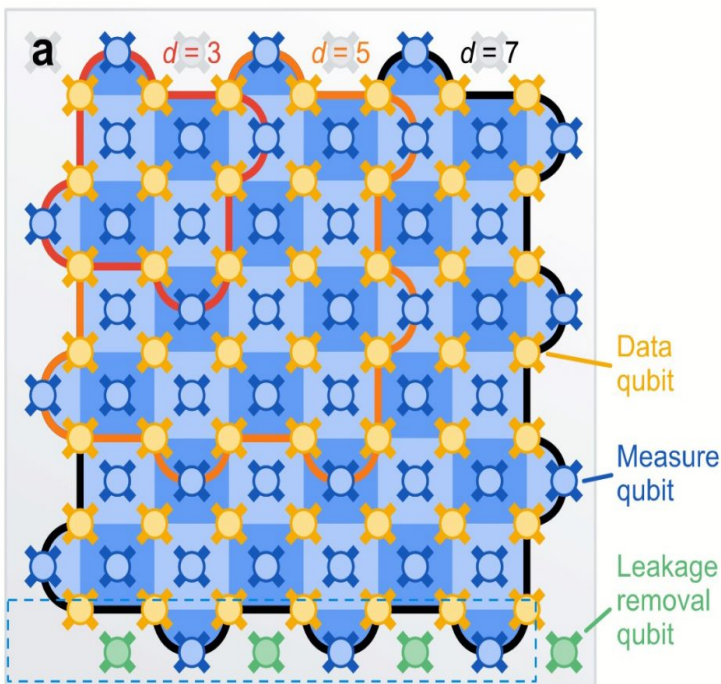
Opinions Libres

le blog d'Olivier Ezratty

Actualité quantique de décembre 2024 spécial Google Willow

Bienvenue dans ce 65^e épisode de Quantum, le podcast de l'actualité quantique. Nous avons initialement prévu de faire relâche en fin d'année et de reprendre le volant début février mais l'actualité nous a rattrapés. L'annonce récente de Google avec sa puce Willow a défrayé la chronique et surtout m'a bien titillé. Il se trouve que nous avons eu l'idée de lancer ce podcast en octobre 2019 à l'occasion de l'annonce de la suprématie quantique de Google (+ un post). L'histoire se répétant, il fallait bien assurer. Qui plus est, des chercheurs chinois ont surenchéri face à Google et c'est tout récent. Donc, cela mérite un approfondissement.

J'ai aussi publié un article détaillé sur le sujet « **Inside Google Willow** ».



Cet épisode est donc consacré à cette annonce, mais à la fin, nous évoquerons d'autres annonces concernant nos startups françaises, qui n'ont pas eu le même écho médiatique, et qui méritent le détour, chez **Quobly** et **Quandela**.

Voici un transcript approximatif de ce podcast.

Le papier de Google dans Nature sur la correction d'erreur

Fanny : Google a donc fait une annonce majeure le 9 décembre 2024 concernant sa puce Willow de 105 qubits.

De quoi s'agissait-il ?

Olivier : Cette puce est le dernier processeur à qubits supraconducteur développé par Google. Il vise principalement à explorer les techniques de correction d'erreurs. Il utilise un code de surface de distance 7, permettant la création d'un qubit logique de meilleure qualité que les qubits physiques du processeur.

Fanny : Cette annonce était liée à la publication d'un article dans *Nature*. Mais était-ce si nouveau ?

Olivier : L'article de *Nature* avait déjà été publié en **preprint arXiv** en août 2024 et nous en avons même parlé dans notre **podcast de rentrée de fin août 2024**. Le papier en question documente deux réalisations. La première est la mise en œuvre de portes logiques à un et deux qubits avec correction d'erreurs en temps réel sur une puce de 72 qubits, dénommée Sycamore, avec un code de surface de distance 5. La seconde est l'utilisation de cette puce Willow de 105 qubits pour démontrer un qubit logique fonctionnel basé sur un code de surface de distance 7. Cette correction d'erreurs n'est pas réalisée en temps réel sur Willow. Cet article s'inscrit dans la continuité des travaux de Google, notamment un **papier clé de 2022/2023** qui décrivait la stratégie pour atteindre ces résultats. À cette époque, Google avait présenté un plan pour utiliser un code de distance 7 avec environ 100 qubits physiques, ce qui a finalement conduit à la conception de Willow. Il leur a fallu deux ans pour y parvenir.

Fanny : Qu'est-ce qui a permis à Google d'obtenir ces résultats ?

Olivier : Depuis son **expérience de "suprématie quantique" en 2019** avec Sycamore (53 qubits), Google a concentré ses efforts sur deux axes: améliorer la fidélité de ses qubits physiques et développer des techniques de correction d'erreurs pour créer des qubits logiques performants. Ce travail s'appuie également sur des innovations comme le « gap engineering » des jonctions Josephson pour réduire les erreurs corrélées et l'usage de qubits dédiés à la suppression des fuites (*Dynamic Qubit Leakage Removal*), des défis qui n'avaient pas encore été traités en 2019 et en 2022.

Fanny : Peux-tu expliquer ce que sont ces erreurs corrélées ?

Olivier : Il s'agit d'erreurs qui se produisent simultanément à plusieurs endroits dans une puce et qui ont une cause commune. La plus connue est l'effet des rayons cosmiques qui génèrent une erreur sur environ un million d'opérations. Avec la technique du gap engineering, Google semble avoir réussi à faire baisser ce taux d'erreurs de 5 ordres de grandeur. On passe maintenant à une heure de calcul sans erreurs. Ce n'est pas suffisant pour les grands circuits.

Fanny : Revenons à la correction d'erreurs. Pourquoi est-elle si essentielle pour les ordinateurs quantiques tolérants aux fautes ?

Olivier : Les qubits physiques sont intrinsèquement bruyants, avec des taux d'erreurs qui empêchent l'exécution de circuits quantiques complexes. La correction d'erreurs permet de regrouper plusieurs qubits physiques pour créer des qubits logiques, capables de maintenir une cohérence sur de longues périodes et des fidélités apparentes plus élevées. Cependant, cela a un coût élevé: un seul qubit logique nécessite entre 100 et 10'000 qubits physiques, selon la fidélité des qubits, leur connectivité et le type de correction d'erreur utilisé. Sachant d'ailleurs que plusieurs codes de correction d'erreurs sont utilisés. Il y en a au moins trois : pour les portes à un qubit, pour celles à deux qubits, et pour les portes particulières qui sont hors du groupe de Clifford comme la porte T et la porte Toffoli.

Fanny : De quoi s'agit-il ?

Olivier : Ce sont des portes quantiques à un et trois qubits qui sont indispensables car elles seules permettent

d'obtenir une accélération exponentielle dans le calcul quantique. Or elles coûtent plus cher à corriger et nécessitent des codes de correction d'erreur particuliers. D'ailleurs, ces portes n'étaient pas encore corrigées dans Willow au moment de la sortie du papier dans Nature. Google y travaille pour la suite. Ils ont publié un arXiv récent sur la Magic State Cultivation qui permettrait de drastiquement réduire les ressources nécessaires pour corriger ces portes T.

Fanny : Google a affirmé que la réduction des erreurs est “exponentielle”. Cela veut dire quoi ?

Olivier : La réduction des erreurs est “exponentielle” car à chaque augmentation de la distance du code de surface, les taux d'erreur logique diminuent exponentiellement. Cependant, cela exige une augmentation polynomiale du nombre de qubits physiques nécessaires. Par exemple, passer d'un code de distance 5 à 7 divise d'un ordre de grandeur les erreurs logiques, mais double le nombre de qubits physiques requis. En pratique, pour faire des simulations chimiques, il faudra des taux d'erreur tellement faibles qu'il sera nécessaire d'avoir des milliers de qubits physiques pour créer un qubit corrigé.

Fanny : ils parlent aussi de deux notions : “sous le seuil” (below threshold) et “à l'équilibre” (breakeven). A quoi est-ce que cela correspond ?

Olivier : “Sous le seuil” signifie que les taux d'erreur des qubits physiques sont suffisamment bas pour que la correction d'erreurs soit possible. Cela permet de réduire les erreurs logiques à mesure que la taille des codes de correction et le nombre de qubits physiques augmente. “À l'équilibre” signifie que les qubits logiques surpassent les qubits physiques en termes de performance, par exemple en maintenant une cohérence plus longue. Willow a atteint ces deux étapes, ce qui est une avancée importante.

Le benchmarking de comparaison avec 10^{25} années de calcul classique

Fanny : L'article publié dans *Nature* a-t-il suscité des critiques?? Comment a-t-il été accueilli dans la communauté scientifique ?

Olivier : Sur la correction d'erreurs, les avis sont unanimes qu'il s'agit d'un progrès certain. Par contre, Google a communiqué sur la réalisation d'un benchmark de manière très contestable. Ils affirment que leur processeur Willow pourrait résoudre en quelques minutes un problème qui prendrait un “septillion d'années” (10^{25} puissance 25) à un superordinateur. Ces chiffres faits pour impressionner reposent sur un benchmark spécifique (le *Cross-Entropy Benchmarking*), qui, bien que difficile à simuler classiquement, n'a aucune application pratique. Ce n'est même pas véritablement du calcul !

Fanny : pourquoi ?

Olivier : cela ne mesure pas la capacité d'un processeur à résoudre un problème concret. Il n'utilise aucune donnée en entrée et ne génère aucune donnée utile en sortie. Il s'agit d'un circuit dont les portes sont générées classiquement de manière aléatoire. C'est une sorte de générateur de nombres aléatoires coûteux et complexe. En plus, il ne fonctionne pas bien. Il génère un résultat correct dans 0.1% des cas. L'évolution par rapport à 2019 est que Google est passé de 20 à 40 cycles de portes quantiques dans ce benchmark. Le défi donné aux supercalculateurs classiques est donc de simuler un système physique imparfait et bruité. C'est absurde. Qui plus est, par le passé, ces benchmarks n'ont pas tenu longtemps. Google revendiquait en 2019 que Sycamore pouvait réaliser avec 53 qubits un calcul en 200 secondes qui prendrait 10^7 000 ans sur un superordinateur classique. En 2022, les tests de Google ont montré que ce temps descendait à 6 secondes, grâce à la technique des réseaux de tenseurs. Cependant, lorsque l'on augmente le nombre de qubits, la simulation classique devient difficile, même avec un réseau de tenseurs. Mais cela sert toujours à rien et ne fournit pas d'indication sur la capacité de calcul pratique. D'ailleurs, si cette capacité était si grande, pourquoi s'embêterait-on à créer un seul qubit logique sur un tel processeur ?

Fanny : Quels seraient des benchmarks plus pertinents pour évaluer Willow??

Olivier : Un benchmark plus pertinent serait de tester la capacité de Willow à créer et maintenir des états GHZ (*Greenberger–Horne–Zeilinger*) de grande taille. Ces états sont des superpositions maximales, comme $|000\dots000\rangle$ et $|111\dots111\rangle$, qui nécessitent une intrication entre tous les qubits. La taille maximale d'un état GHZ réalisable avec une haute-fidélité est un indicateur clé de la capacité d'un processeur quantique à manipuler des états complexes. Mais ce n'est pas le seul. Ce benchmark pourrait être réalisé sans artifices comme la post-sélection utilisée par Quantinuum (ions piégés) et QuEra (atomes froids), ce qui le rendrait beaucoup plus représentatif des performances pratiques. Autre benchmark possible : essayer de réaliser un algorithme NISQ avec (sans correction d'erreur). On en verra sûrement apparaître en 2025 et on se rendra compte qu'ils utiliseront moins de qubits et le feront avec des circuits peu profonds.

Fanny : OK, ce benchmark cache-t-il autre chose ?

Olivier : peut-être le manque d'algorithmes quantiques capables d'apporter une accélération exponentielle, en tout cas en régime NISQ bruité. Car le benchmark ne relève pas de la tolérance aux fautes et n'utilise pas de qubits logiques. C'est un benchmark « NISQ ». En régime FTQC, on est à peu près sûr d'obtenir une accélération exponentielle pour des simulations chimiques mais c'est bien moins évident pour les algorithmes quantiques de résolution d'optimisations combinatoires ou de machine learning. On a besoin de beaucoup de recherche dans le domaine.

Fanny : est-ce que Willow permettrait de casser des clés RSA 2048 bits ?

Olivier : on en est très loin. Pour donner un ordre d'idée, Willow comprend un qubit logique avec 10^{-3} d'erreurs. Pour casser une clé RSA, il faut au minima 4000 qubits logiques avec 10^{-12} erreurs. Chaque qubit logique de cette qualité va demander au moins 4000 qubits physiques. A la fin, on voit qu'il faudrait l'équivalent de 16M qubits physiques, répartis sur 160 000 Willow ou 1600 puces à 10K qubits pour casser RSA. Il y a encore beaucoup de pain sur la planche pour y arriver. Mais le progrès technologique aidant, ce nombre de qubits et de puces pourrait se réduire avec le temps et réduire d'autant l'addition.

Les mondes multiples

Fanny : Google prétendait dans une déclaration que leurs résultats soutiendraient l'idée des mondes multiples.

Olivier : Cette déclaration provient d'une interprétation spécifique de la mécanique quantique, connue sous le nom d'Interprétation des Mondes Multiples (*Many-Worlds Interpretation*, MWI). Selon cette interprétation, chaque fois qu'un événement quantique survient, comme une mesure ou une interaction, l'univers se "divise" en plusieurs branches, chacune correspondant à un résultat possible. Google suggère que leur calcul, qui implique des qubits dans des superpositions complexes, serait une manifestation de ces branches multiples dans lesquelles des calculs parallèles se dérouleraient. Cette affirmation est controversée. Bien que l'interprétation des mondes multiples soit fascinante, elle reste une hypothèse philosophique, et aucune expérience à ce jour, y compris celles avec Willow, ne peut confirmer ou infirmer cette interprétation. De nombreux physiciens préfèrent d'autres interprétations, comme celle de Copenhague, qui est moins spéculative, la théorie de l'onde pilote notamment promue par Aurélien Drezet de l'Institut Néel à Grenoble, et aussi celle de Philippe Grangier, avec CSM.

Fanny : Est-il vraiment nécessaire d'invoquer les mondes multiples pour expliquer le fonctionnement de Willow??

Olivier : Ce n'est pas nécessaire. La mécanique quantique fournit une description précise du fonctionnement des qubits sans qu'il soit indispensable d'invoquer les mondes multiples. Les qubits dans Willow suivent les

règles bien établies de la théorie quantique?: ils évoluent dans un espace d'états appelé espace de Hilbert, où des superpositions et des intrications se produisent. Ces concepts sont mathématiques et ne nécessitent pas une explication en termes de multivers. L'invocation des mondes multiples peut être vue comme une simplification pour expliquer intuitivement les calculs parallèles effectués par un ordinateur quantique. Cependant, ces "calculs parallèles" sont mieux compris comme des interférences dans un espace quantique, et non comme des processus se déroulant dans des univers distincts.

Fanny : Pourquoi cette mention des mondes multiples a-t-elle été perçue comme problématique??

Olivier : elle mélange des considérations philosophiques avec des résultats scientifiques, ce qui peut prêter à confusion. En associant leurs résultats aux mondes multiples, Google a semblé valider une interprétation particulière de la mécanique quantique qui reste spéculative. Cela détourne l'attention des véritables avancées techniques réalisées par Willow, qui n'ont absolument pas besoin de cette hypothèse pour être expliquées. En outre, certaines voix dans la communauté scientifique ont critiqué cette déclaration comme étant davantage une tentative de rendre leurs travaux plus "mystiques" ou attrayants pour un large public, plutôt qu'un reflet précis des implications scientifiques.

Fanny : est que cette déclaration sur les mondes multiples pourrait avoir un impact sur la perception publique du calcul quantique??

Olivier : cela peut captiver l'imagination. Mais cela sème la confusion, en plus des affirmations sur les septillions d'années, en particulier auprès du grand public. Cela risque de donner l'impression que les ordinateurs quantiques réalisent littéralement des calculs dans des univers parallèles, ce qui n'est pas exact. Cela peut également détourner l'attention des véritables défis techniques et scientifiques du domaine. D'un autre côté, si cette mention attire davantage d'intérêt pour le calcul quantique, elle peut avoir un impact positif en suscitant la curiosité. Mais il est crucial que cette curiosité soit ensuite dirigée vers une compréhension précise des principes sous-jacents, sans amplification inutile des implications philosophiques.

La communication de Google

Fanny : D'où vient ce nom Willow ?

Olivier : C'est le nom d'un film de fantasy. Mais tu en sais peut-être plus ?

Fanny : Willow est un personnage nain mais je ne sais pas pourquoi ils ont choisi ce nom.

Fanny : Google est connu pour ses annonces spectaculaires. Que penser de cette stratégie de communication ?

Olivier : Expliquer la correction d'erreur étant difficile pour toucher le grand public, ils ont déplacé le curseur vers les comparaisons de temps de calcul, mais elles sont abusives. Cela pose un problème d'intégrité scientifique.

Fanny : Les affirmations comme le calcul "septillionnaire" ou les références aux mondes multiples sont-elles un atout ou un risque pour leur crédibilité scientifique ?

Olivier : Ces affirmations sont un atout sur le plan marketing, car elles attirent énormément l'attention des médias et du public. Elles positionnent Google comme un leader dans le domaine du calcul quantique. Mais cela peut aussi nuire à leur crédibilité scientifique. Les comparaisons avec des superordinateurs, par exemple, occultent les limitations actuelles des ordinateurs quantiques, qui sont encore loin de surpasser les solutions classiques pour la résolution de problèmes pratiques.

Fanny : Pourquoi Google choisit-il d'insister sur des benchmarks qui n'ont pas d'application pratique??

Olivier : Peut-être parce que peu de gens se rendent compte de l'entourloupe. Seuls les spécialistes pointus du calcul quantique connaissent l'astuce. Pas vu, pas pris ! C'est comme en politique. C'est d'ailleurs encore débattu entre spécialistes du domaine.

Fanny : Google a-t-il une responsabilité dans la manière dont il communique auprès du grand public??

Olivier : Absolument. En tant que leader technologique, Google influence fortement la perception publique du calcul quantique. Leur communication devrait donc équilibrer l'enthousiasme et la rigueur scientifique. Exagérer les capacités des ordinateurs quantiques ou simplifier les concepts pourrait créer des attentes irréalistes, ce qui risque de décevoir à long terme. Une communication plus précise, qui explique à la fois les avancées et les limites, permettrait de construire une compréhension plus solide et durable du domaine.

Fanny : Comment se comportent les concurrents de Google, tels qu'IBM ?

Olivier : Ils sont bien plus sages et respectueux de l'intégrité scientifique. En 2023, par exemple, IBM a communiqué plus prudemment sur la notion de « quantum utility » autour d'algorithmes NISQ pour qubits bruités à partir de 127 qubits, et exploitant la technique de la mitigation d'erreurs quantiques. Ils ont publié avec leurs partenaires des benchmarks documentés. Mais ils n'ont pas prétendu avoir atteint un avantage quantique ni brandi de chiffres mirobolants. Il en va de même de ce qu'ont pu faire Quantinuum et QuEra par exemple.

Fanny : Quels seraient tes conseils pour améliorer la communication de Google et de ses concurrents dans ce domaine??

Olivier : Google pourrait adopter une approche plus équilibrée, en mettant davantage l'accent sur les défis qu'ils rencontrent et sur les implications à long terme de leurs travaux.

Par exemple?:

- Insister sur les étapes intermédiaires vers des applications pratiques, plutôt que sur des comparaisons extrêmes.
- Être transparents sur les limites actuelles des technologies quantiques, tout en expliquant comment ces défis sont abordés.
- Proposer des récits qui intègrent des perspectives académiques et industrielles, pour montrer comment leurs travaux s'inscrivent dans une dynamique collective.
- Éviter de mélanger des spéculations philosophiques, comme les mondes multiples, avec des résultats techniques, afin de maintenir une clarté scientifique.

Une telle stratégie renforcerait leur crédibilité et favoriserait une compréhension plus claire des progrès réels dans le domaine du calcul quantique. Mais libre à eux évidemment de faire ce qu'ils veulent.

Défis restants

Fanny : Quels sont les principaux défis qui restent à relever pour Google et pour l'industrie en général afin de construire des ordinateurs FTQC??

Olivier : Les défis pour atteindre des ordinateurs quantiques tolérants aux fautes sont multiples et interdépendants?:

- **Correction d'erreurs pour les portes non-Clifford.** Les portes T, essentielles pour les accélérations exponentielles, nécessitent des ressources considérables, notamment la distillation d'états magiques. Google et d'autres travaillent sur des méthodes plus efficaces comme la "cultivation" d'états magiques.
- **Élargissement des puces.** Construire des puces avec 10?000 qubits physiques ou plus, tout en maintenant les fidélités actuelles, est un défi technologique majeur. Cela nécessite également une maîtrise de la fabrication en salle blanche.
- **Interconnexions entre puces.** Les puces supraconductrices, comme Willow, devront être reliées par des guides d'ondes micro-ondes ou des transducteurs optiques, des technologies encore à un faible niveau de maturité.
- **Électronique et cryogénie.** Avec des milliers de qubits, le câblage, les systèmes de contrôle et les solutions cryogéniques devront être repensés pour rester scalables et économiquement viables.
- **Applications pratiques.** Développer des algorithmes adaptés à ces architectures est tout aussi crucial. Les progrès en chimie quantique et en optimisation combinatoire, par exemple, nécessiteront des ressources massives, avec des millions de qubits physiques pour atteindre des qubits logiques de haute fidélité.

Fanny : est-ce que l'annonce de Willow modifie le calendrier vers le FTQC ?

Olivier : cela reste un objectif long terme. Entre 5 et 10 ans pour obtenir une centaine de qubits logiques et au moins une quinzaine d'années pour obtenir 4000 qubits logiques. Au nez.

Réponse chinoise

Fanny : quelle a été la réponse des chinois ?

Olivier : Elle est arrivée en deux temps. D'abord avec l'annonce réchauffée d'un processeur de 504 qubits. Puis d'une suprématie quantique obtenue avec 105 qubits, le même nombre de qubits que Willow. Cette synchronicité est étonnante mais peut s'expliquer. Le nombre 105 est lié à la disposition en maillage à 45° des qubits.

China Introduces 504-Qubit Superconducting Chip by Matt Swayne, The Quantum Insider, December 2024.

Establishing a New Benchmark in Quantum Computational Advantage with 105-qubit Zuchongzhi 3.0 Processor by Dongxin Gao, Jian-Wei Pan et al, arXiv, December 2024 (16 pages).

99.9%-fidelity in measuring a superconducting qubit by Can Wang, Jian-Wei Pan et al, arXiv, December 2024 (12 pages).

Actualité des startups françaises

Parlons des startups française et de leurs annonces récentes.

Quobly et STMicroelectronics

Lors de la Q2B Santa Clara, Quobly et STMicroelectronics annonçaient un partenariat pour la production de puces pour les qubits et pour l'électronique basse température. Cette annonce est importante car elle permet à Quobly de travailler avec un partenaire qui peut customiser en mode « codéveloppement » la fabrication de ses

puces.

https://www.ecranmobile.fr/Quobly-et-STMicroelectronics-un-partenariat-strategique-pour-revolutionner-le-calcul-quantique-a-grande-echelle_a77392.html

Quandela

Bien plus discrètement, l'équipe de Quandela a publié il y a quelques jours un intéressant preprint sur la génération de cluster states de photons intriqués qui a l'air d'être plusieurs ordres de grandeur plus efficace que ce que leur concurrent PsiQuantum essaye de faire. C'est un progrès certain dans la direction du calcul à tolérance aux fautes à base de photons.

Minimizing resource overhead in fusion-based quantum computation using hybrid spin-photon devices by Stephen C. Wein, Timothée Goubault de Brugière, Luka Music, Pascale Senellart, Boris Bourdoncle, and Shane Mansfield, arXiv, December 2024 (22 pages).

On se retrouve, là, c'est sûr, fin janvier ou début février pour le prochain épisode.

Cet article a été publié le 19 décembre 2024 et édité en PDF le 21 décembre 2024.
(cc) Olivier Ezratty – “Opinions Libres” – <https://www.oezratty.net>