

Optimiser la consommation énergétique des calculateurs quantiques : un défi interdisciplinaire

Alexia Auffèves (alexia.auffeves@neel.cnrs.fr)

Institut Néel (UPR2940 CNRS et Université Grenoble Alpes)
25 avenue des Martyrs, BP 166, 38042 Grenoble Cedex 9

Jusqu'ici, la question de la facture énergétique d'un ordinateur quantique n'a guère été abordée par les programmes de recherche – bien qu'elle amène actuellement l'industrie des semi-conducteurs à ses limites.

Existe-t-il un avantage quantique de nature énergétique ?

Ou bien, au contraire, faudra-t-il une centrale nucléaire par ordinateur quantique ?

Cet article propose des pistes pour répondre à ces questions.

L'informatique quantique est un domaine en plein essor, mettant en synergie la recherche fondamentale, l'ingénierie et l'industrie. En exploitant des mécanismes quantiques tels que la superposition et l'intrication, elle promet de résoudre des problèmes complexes, classiquement intraitables. C'est la « suprématie quantique » [1], une première étape ayant été franchie en octobre 2019 par Google. Si le chemin jusqu'à l'ordinateur quantique résistant aux erreurs est long, il trace déjà de nouvelles voies prometteuses pour le traitement de l'information sur la base des dispositifs physiques actuels.

Cout énergétique du calcul classique

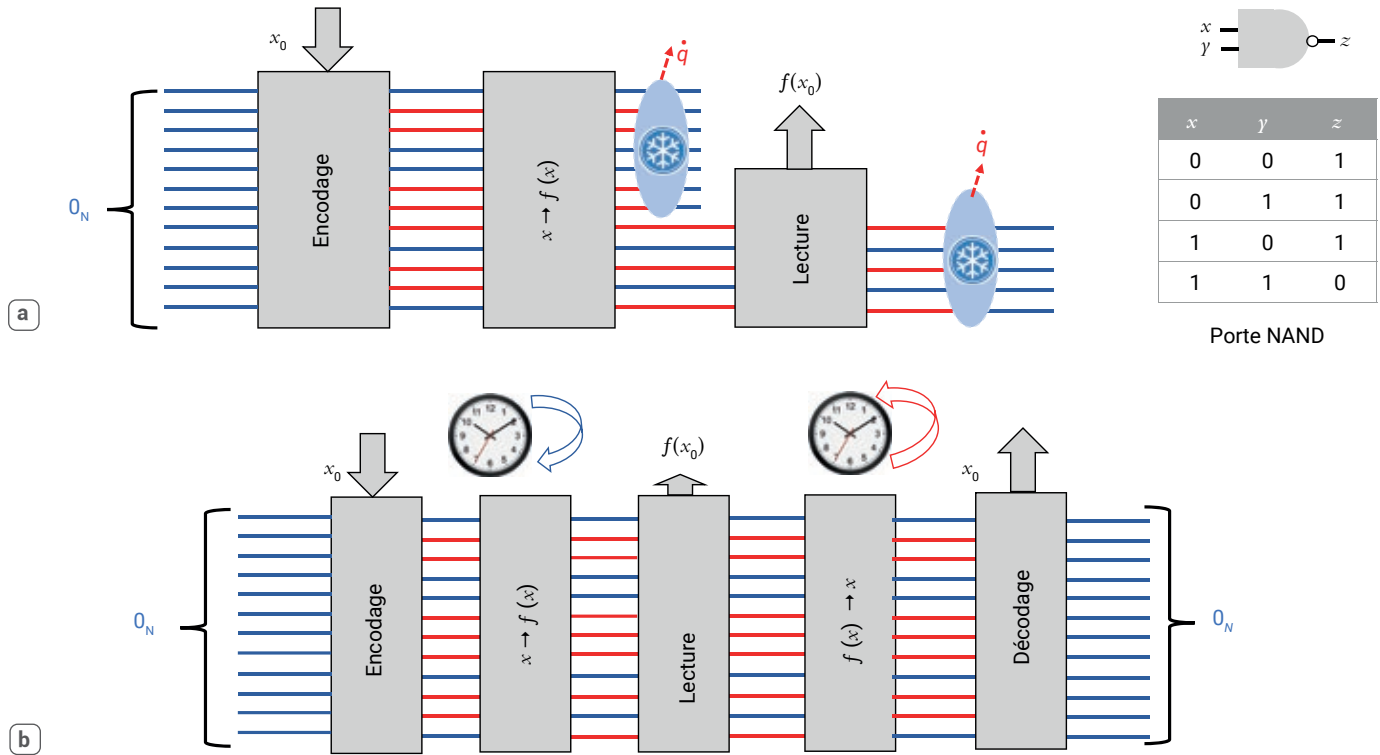
Avant d'aborder la question du calcul quantique, il est indispensable de préciser quelques notions de base sur le traitement de l'information classique. Celle-ci est encodée sur des « bits », à savoir des systèmes physiques ne pouvant prendre que deux états identifiés par convention comme 0 ou 1. Un exemple typique est celui des états « chargé » et « déchargé » d'un condensateur. Il peut également s'agir de l'aimantation d'un matériau. Un registre comportant N bits peut ainsi encoder une valeur d'une variable x parmi 2^N possibilités.

Un ordinateur peut être décrit de façon très idéalisée comme une machine programmable ayant pour fonction de délivrer, à partir de la variable x , la fonction $f(x)$ (fig. 1). Pour ce faire, les N bits d'un « registre de calcul » sont initialement préparés dans l'état 0. Cela correspond à l'état « 0_N » du registre, sur lequel la variable x est copiée. Une séquence de portes logiques est ensuite appliquée sur les différents bits du registre, à une fréquence typique appelée fréquence d'horloge. La porte NAND (insérée en haut à droite de la figure 1a) est un exemple de porte logique à deux bits. Elle est universelle : tous les calculs classiques peuvent en principe être réalisés sous forme de séquences de portes NAND. Au terme des opérations logiques, le résultat $f(x)$ peut être lu sur le registre, qui doit finalement être réinitialisé dans l'état 0_N pour qu'un nouveau calcul puisse être effectué.

Actuellement, le traitement de l'information classique est caractérisé par son irréversibilité, qui conduit à une dissipation de chaleur. L'irréversibilité est d'abord de nature thermodynamique : par exemple, la résistivité des composants électroniques induit une dissipation de chaleur lors des commutations logiques, qui augmente avec la fréquence d'horloge et la densité des transistors [2]. Il existe également une irréversibilité de nature fondamentale :

Remerciements

L'autrice remercie chaleureusement Olivier Ezratty, Marco Fellous-Asiani, Benjamin Huard, Yvain Thonnart et Robert Whitney pour leurs relectures attentives.



1. Calcul irréversible versus calcul réversible. Un calcul consiste en une séquence d'opérations physiques effectuées sur un registre comportant N bits. La couleur bleue (resp. rouge) symbolise la valeur 0 (resp. 1) du bit correspondant. Le registre est préalablement initialisé dans l'état 0_N , puis la variable x_0 est encodée sur le registre. Une séquence de portes logiques permet le calcul de $f(x_0)$, résultat lu par l'opérateur. Le registre doit être réinitialisé avant le lancement du calcul suivant.

a) Paradigme actuel (dit de Landauer). Les portes logiques utilisées pour le calcul sont irréversibles et le registre est réinitialisé en fin de calcul. Information et chaleur \dot{q} sont dissipées dans un réservoir froid (symbolisé par un flocon de neige). Insert : symbole et table de vérité d'une porte NAND (qui, à deux entrées x et y, pouvant avoir chacune la valeur 0 ou 1, associe une sortie z qui a la valeur 1 seulement si l'une des deux entrées a la valeur 0).

b) Paradigme réversible. Les portes logiques sont réversibles, ce qui permet d'inverser physiquement toutes les étapes du calcul après extraction du résultat. L'énergie dépensée pour réaliser le calcul est récupérée et le registre d'entrée est réinitialisé sans dissipation de chaleur. Le déploiement d'une technologie réversible est à l'étude dans de nombreux laboratoires de R&D.

l'irréversibilité logique, liée à la perte d'information au cours du calcul. Elle est due à la conception physique des portes NAND, avec deux bits en entrée et un seul en sortie (insert de la figure 1a). Elle est également liée à la réinitialisation du registre de calcul (fig. 1a). Comme l'irréversibilité thermodynamique, l'irréversibilité logique a des conséquences énergétiques. Rolf Landauer [3] a ainsi prédit que la perte d'un bit d'information dans un environnement à température T correspond à une dissipation minimale de chaleur de $k_B T \log 2$, typiquement 10^{-20} J à température ambiante^(a).

Calcul réversible

Dans nos ordinateurs actuels, la dissipation de chaleur est due à l'irréversibilité thermodynamique. Elle est supérieure de trois à

quatre ordres de grandeur à la limite fondamentale de Landauer, mais cette dernière pourrait rapidement devenir un obstacle sérieux du fait de l'explosion de la demande en puissance de calcul (big data, intelligence artificielle, simulations...). Un changement de paradigme pourrait être apporté par le calcul réversible : le calcul est opéré dans le sens direct, puis après lecture du résultat, est « rembobiné » comme dans un film. Cette opération inverse réinitialise le registre d'entrée sans dissiper de chaleur, et permet de récupérer l'énergie dépensée pour réaliser le calcul (fig. 1b). Ces idées révolutionnaires, conçues par Charles Bennett dans les années 1980, sont considérées comme une solution de plus en plus sérieuse [4]. Plusieurs pistes sont explorées, en particulier le CMOS adiabatique^(b), et les circuits supraconducteurs réversibles.

Un avantage énergétique quantique ?

A priori, on pourrait s'attendre à ce qu'un calcul quantique soit moins coûteux en énergie qu'un calcul classique équivalent, et ce pour deux raisons. D'une part, la notion de superposition d'états de la logique quantique permet d'envisager de diminuer le nombre d'étapes nécessaires à l'obtention du résultat. Ainsi, l'expérience de suprématie quantique [5] réalisée par Google [6] a mis en jeu des puissances typiques de 25 kW, trois ordres de grandeur en dessous des puissances consommées par des supercalculateurs de type IBM Summit^(c).

D'autre part et même si ce n'est pas une caractéristique usuellement mise en avant, le calcul quantique constitue un exemple typique de traitement réversible de l'information. Celle-ci est codée sur des registres



.....

A priori, on peut s'attendre à ce qu'un calcul quantique soit moins couteux en énergie qu'un calcul classique équivalent, pour deux raisons. D'une part, la notion de superposition d'états de la logique quantique permet d'envisager de diminuer le nombre d'étapes nécessaires à l'obtention du résultat. D'autre part,... le calcul quantique constitue un exemple typique de traitement réversible de l'information.

.....

>>>

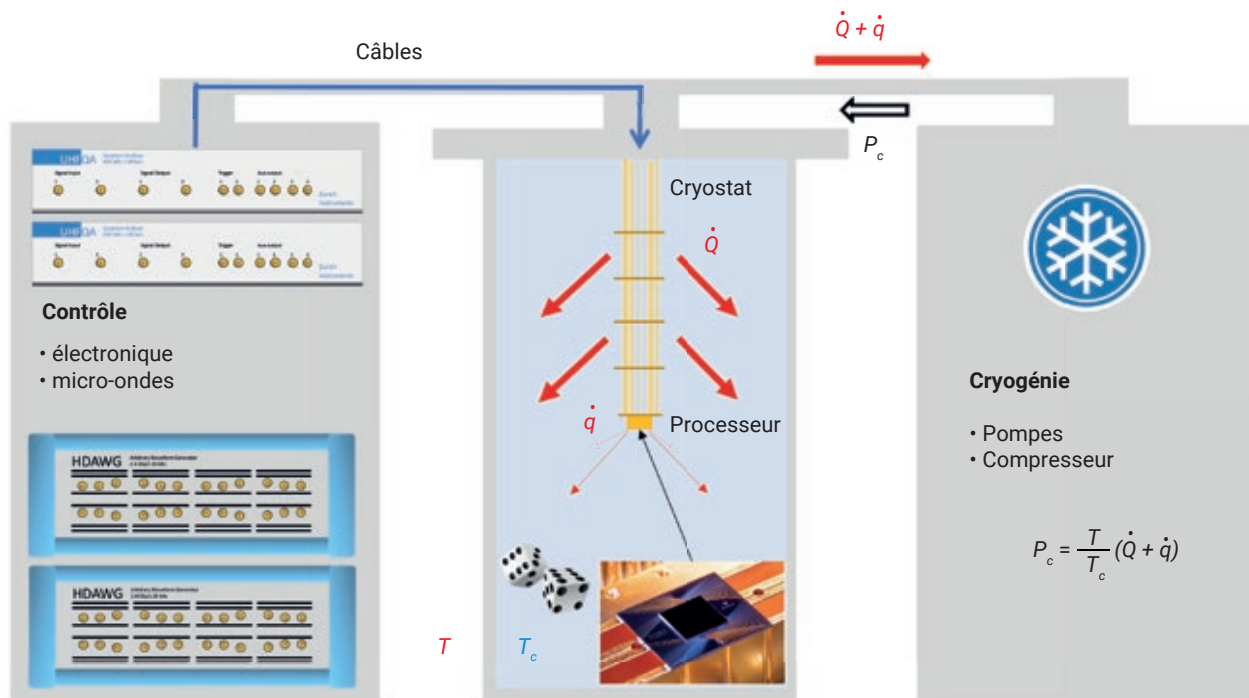
composés de bits quantiques ou qubits. Contrairement à leurs homologues classiques, les qubits peuvent non seulement se trouver dans l'état « 0 » ou dans l'état « 1 », mais aussi dans des superpositions cohérentes d'états « 0 et 1 ». Un calcul quantique repose ainsi sur la préparation de superpositions cohérentes d'un grand nombre de qubits. Ces états peuvent être vus comme des « chats de Schrödinger », à la fois morts et vivants tant qu'aucune mesure de leur état n'est réalisée. C'est sur cette cohérence à grande échelle que repose une part de l'avantage quantique computationnel, et celle-ci doit être préservée tout au long du calcul.

Cela impose en particulier que les portes logiques quantiques ne dissipent aucune information dans l'environnement : cela reviendrait en effet à mesurer le registre, et à transformer les superpositions cohérentes « 0 et 1 » en états « 0 » ou « 1 », appelés mélanges statistiques, sans intérêt pour le calcul quantique. C'est le phénomène de décohérence, qui rend toute expérience de physique quantique si délicate à réaliser. Les portes quantiques doivent donc être *logiquement réversibles*. Dans un monde idéal, un calcul quantique devrait pouvoir être effectué dans le sens direct au prix

d'une certaine énergie, puis l'énergie récupérée en réalisant ensuite le calcul en sens inverse, conduisant à un coût énergétique nul ! Une telle gratuité présuppose cependant que le calcul est effectué sans erreurs, ce qui est irréaliste. Comme expliqué ci-dessous, la lutte contre les erreurs est l'origine fondamentale du coût énergétique du calcul quantique.

Le coût énergétique de la lutte contre le bruit

Un processeur est toujours couplé à un environnement : par exemple, les qubits supraconducteurs sont en interaction avec des phonons, des charges électriques... Les interactions incontrôlées qui en résultent sont à l'origine d'un « bruit quantique », c'est-à-dire des perturbations aléatoires de l'état des qubits qui conduisent à des erreurs dans les algorithmes, d'autant plus sévères que le taux de bruit est plus important. Ultimement, ce bruit tend à dégrader l'état des qubits dans des états de type « 0 » ou « 1 » sans intérêt pour le calcul. Les expérimentateurs doivent donc relever le défi de taille d'isoler le processeur du bruit quantique (fig. 2).



2. Un calculateur quantique est un système hors-équilibre. Le processeur est refroidi à la température $T_c < T$, où T est la température ambiante. Les dés symbolisent le bruit quantique rémanent à l'intérieur du cryostat. Le processeur est contrôlé depuis l'extérieur par des câbles qui dissipent un taux de chaleur \dot{Q} (flèches épaisses rouges), le processeur dissipant \dot{q} (flèches fines rouges). La puissance dépensée ne peut pas être inférieure à la puissance de Carnot $P_c = (T/T_c) (\dot{Q} + \dot{q})$

D'un point de vue thermodynamique, il s'agit typiquement de réaliser un processus « hors-équilibre », à savoir préserver la cohérence et l'intrication quantiques alors même que le bruit tend à les détruire. Le deuxième principe de la thermodynamique nous enseigne que maintenir des états hors-équilibre a un coût fondamental, qui est à l'origine de la facture énergétique du calcul quantique. Ramener le bruit à des taux compatibles avec une certaine probabilité de succès du calcul coûte une puissance minimale – d'autant plus importante que le taux de bruit visé est plus faible. On liste ci-dessous les principaux postes de dépense dans le cas typique de qubits solides, tels que les supraconducteurs ou les qubits de *spin* d'électron piégé dans une boîte quantique en matrice de silicium. Les pistes pour réduire ces coûts sont également indiquées. La stratégie de réduction du bruit comporte généralement trois étapes à réaliser successivement.

1) *Calculer à froid*. Pour la plupart des qubits à l'état solide, la première action est de réduire le bruit thermique en plaçant le processeur en environnement cryogénique. Le calcul est effectué au cœur de cryostats à dilution où la température est de quelques millikelvins (pour les qubits supraconducteurs), et jusqu'à 1 K (pour les qubits de *spin*). Dans les expériences actuelles, le processeur est relié à l'extérieur *via* des câbles transportant des signaux analogiques : ce sont des impulsions micro-ondes activant les portes à un et deux qubits et servant à la mesure de ceux-ci.

La puissance minimale à consacrer au maintien de la température T_c des qubits est la puissance de Carnot. Dans une approche simplifiée^(d), celle-ci est proportionnelle au rapport des températures T/T_c , où T est la température ambiante. Cela motive à augmenter la température opérationnelle T_c . L'amélioration de l'efficacité des cryostats fait l'objet d'importantes recherches en cryogénie. D'autre part la puissance de Carnot évolue comme \dot{Q} , le taux de chaleur dissipée à l'intérieur du cryostat, qui a deux sources principales de nature pratique (fig. 2) : (i) les impulsions micro-ondes doivent être amplifiées, puis atténuées fortement pour filtrer le bruit thermique venant de l'extérieur, ce qui occasionne d'importantes dissipations de chaleur ; (ii) les câbles conduisent la chaleur de l'extérieur dans le cryostat. Il est délicat de donner une valeur précise pour \dot{Q} , tant

celle-ci dépend d'une multiplicité de paramètres tels que la technologie des câbles, les stratégies de multiplexage, les températures aux différents étages du cryostat, le nombre de portes logiques activées en parallèle... On donne des ordres de grandeur dans le cas d'une porte à 1 qubit (voir l'encadré). L'optimisation de \dot{Q} sera un enjeu essentiel, qui repose sur un travail largement interdisciplinaire.

2) *Calculer plus vite que le bruit*. La cryogénie permet de ramener le bruit affectant les qubits à des niveaux tolérables. La deuxième action est de réaliser les calculs sur des échelles de temps courtes pour réduire la probabilité qu'une erreur se produise au cours de son déroulement. Cette stratégie a ses limites, car elle augmente la puissance consommée et donc la chaleur dissipée à l'intérieur du cryostat par unité de temps, notamment pour activer les portes quantiques. Qui plus est, la puissance de contrôle est bornée supérieurement sous peine d'induire des erreurs^(e), voire de détruire le qubit.

3) *Corriger les erreurs*. Réaliser pleinement les promesses de l'informatique quantique implique la manipulation de registres de

grande taille et l'enchaînement d'un grand nombre de séquences de portes quantiques. C'est par exemple le cas de l'algorithme de factorisation de nombres entiers de Shor [7], qui pourrait être utilisé pour décrypter une clé RSA. Cela constitue un exemple de « suprématie quantique » [5], à même de générer des failles de sécurité redoutables. Une telle opération requiert l'activation de millions de portes logiques – augmentant d'autant la probabilité d'échec de l'algorithme complet [8]. Les stratégies 1) et 2) présentées ci-dessus ne sont en général pas suffisantes pour réduire le bruit aux niveaux souhaités^(f). Il faut donc développer des dispositifs tolérants aux erreurs, fondés sur la redondance de l'information [9]. Un bit d'information quantique (le qubit dit « logique ») est alors codé sur plusieurs qubits physiques (de 5 pour les codes correcteurs d'erreur les plus simples, à plusieurs dizaines de milliers). Les erreurs sont détectées et corrigées au cours même du calcul, mais le prix à payer est élevé : le nombre de qubits contenus par le processeur tolérant aux erreurs peut alors atteindre plusieurs millions, ce qui impose une importante quantité d'électronique de contrôle et de connexions avec l'extérieur du cryostat.

>>>

Consommation énergétique d'une porte à un qubit supraconducteur

On donne ici des ordres de grandeur de consommation énergétique dans le cas d'un processeur élémentaire, composé d'un unique qubit supraconducteur. Un algorithme se réduit à une séquence de portes quantiques à un qubit, c'est-à-dire un train d'impulsions micro-ondes. La puissance typique d'une telle impulsion est $P_g = 10^{-12}$ W, ce qui correspond à une dizaine de photons de fréquence 10 GHz échangés pendant 1 ns.

On distingue les deux scénarios ci-dessous.

- **Scénario non autonome** : les signaux sont mis en forme à l'extérieur du cryostat et amenés au qubit *via* un câble, dont la conduction induit un taux de chaleur typique au niveau du processeur $\dot{Q}_{cond} (T \rightarrow T_c) = 1$ mW (valeur pour un mètre de coaxial entre $T = 300$ K et $T_c = 1$ mK). D'autre part, le câble conduit également le rayonnement thermique à température ambiante. Afin d'améliorer le rapport signal sur bruit, les impulsions sont amplifiées lors de leur génération, puis atténuées au niveau du qubit de typiquement $A = T/T_c$. L'atténuateur dissipe ainsi un taux de chaleur $\dot{Q}_{diss} = AP_g = 10^{-7}$ W.
- **Scénario autonome** : les signaux sont mis en forme à l'intérieur du cryostat. La dissipation de chaleur se réduit à celle de l'électronique de contrôle, $\dot{Q}_{CMOS} = 1$ mW par porte active. Cette valeur pourrait être encore réduite en exploitant des technologies alternatives.

Les pistes autonomes

Les connexions du processeur froid avec le monde extérieur sont à l'origine de la plus grosse partie de la facture énergétique, du fait de la dissipation de chaleur interne au cryostat par les câbles. L'amélioration des architectures de calcul, de la connectivité des codes et de l'adressage des qubits est indispensable, mais permettra difficilement une réelle mise à l'échelle. Pour pallier cette difficulté, il est envisagé à moyen terme que la génération des signaux analogiques qui contrôlent le processeur, tout comme le traitement de l'information, soient effectués à l'intérieur du cryostat, le calcul se déroulant en autonomie complète [9]. La communication avec l'extérieur se réduirait alors à l'envoi d'instructions et à l'extraction du résultat, respectivement au début et en fin de calcul, le tout sous la forme de signaux numériques.

Dans cette situation, la dissipation de chaleur \dot{q} (fig. 2) est localisée au niveau du processeur, et essentiellement due à l'électronique contrôlant *in situ* la génération et la lecture des micro-ondes, ainsi que la correction des erreurs. C'est dans cette perspective qu'est actuellement développée la technologie de cryo-CMOS, qui vise à

améliorer les performances énergétiques des transistors en environnement cryogénique. Cependant, les meilleures projections estiment une réduction de la dissipation à quelques $\dot{q}_{\text{CMOS}} = 100 \mu\text{W}$ par porte active. Au vu de leur performance actuelle, la consommation d'énergie des cryostats devrait atteindre 1 MW pour supporter un million de qubits physiques fonctionnant à 10 mK^(g).

Ces estimations mettent en lumière tout l'intérêt pour le calcul quantique d'explorer des pistes alternatives telles que le calcul réversible, qui pourraient réduire drastiquement la facture énergétique de la partie classique de l'ordinateur quantique. Elles montrent aussi que les performances énergétiques du traitement classique et quantique de l'information sont irrévocablement liées. Eu égard au nombre de qubits potentiellement en jeu, réduire le coût énergétique de l'électronique de contrôle apparaît essentiel. Dans une perspective encore plus futuriste, les mécanismes de correction d'erreur (mesure et rétroaction) pourraient être intégralement réalisés à l'échelle quantique sans mettre en jeu aucune information classique – réduisant d'autant la facture énergétique.

Conclusion et perspectives

Le dimensionnement énergétique des calculateurs quantiques apparaît comme une contrainte fondamentale pour leur mise à l'échelle, et définit les conditions de possibilité d'un nouvel avantage quantique de nature purement énergétique. La compréhension et l'optimisation de leur consommation trace de nouvelles pistes d'exploration à l'interface entre recherche fondamentale et ingénierie. Les pistes évoquées ici peuvent servir de base à la réalisation d'une étude globale et systématique de la consommation énergétique pour les différentes technologies de qubits actuellement développées [10].

Ultimement, c'est bien la présence du bruit qui empêche d'inverser le sens d'un calcul et de récupérer l'énergie dépensée pour le réaliser. L'un des défis de la recherche sera de dériver de nouvelles bornes fondamentales reliant quantitativement le taux de bruit à la consommation énergétique des calculateurs, en particulier dans le régime quantique. L'enjeu, de taille dans un monde dont les ressources paraissent toujours plus limitées, sera d'optimiser leur consommation et de dégager tout l'avantage énergétique que l'on peut attendre de ces calculateurs d'un nouveau type. ■

Références

- 1• M. Le Bellac, « Peut-on parler de suprématie quantique ? », *Reflets de la physique* **67** (2020) 4-8.
- 2• J.L. Hennessy et D. A. Patterson, "A New Golden Age for Computer Architecture", *Communications of the ACM* **62** (2019) 48-60.
- 3• R. Landauer, "Irreversibility and heat generation in the computing process", *IBM Journal of Research and Development*, **5** (3) (1961) 183-191.
- 4• C.H. Bennett, "Logical Reversibility of Computation", *IBM Journal of Research and Development*, **17** (6) (1973) 525-532.
- 5• J. Preskill, "Why I Called It 'Quantum Supremacy'", *Quanta Magazine* (2 octobre 2019).
- 6• F. Arute, "Quantum supremacy using a programmable superconducting processor", *Nature* **574** (2019) 505-510.
- 7• P.W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring", *Proceedings 35th Annual Symposium on Foundations of Computer Science*, IEEE Comput. Soc. Press (1994) 124-134.
- 8• C. Gidney et M. Ekerå, "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits", *Quantum* **5** (2021) 433.
- 9• M.A. Nielsen et I.L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press (2010).
- 10• J. Bardin et al., "A 28nm Bulk-CMOS 4-to-8GHz <2mW Cryogenic Pulse Moderator for Scalable Quantum Computing", *Proceedings of the 2019 IEEE International Solid-State Circuits Conference* (2019) 456-458.
- 11• O. Ezratty, *Comprendre l'informatique quantique* (2020), e-book téléchargeable sur www.oezratty.net/wordpress/2020/comprendre-informatique-quantique-edition-2020/

(a) Il s'agit d'une valeur microscopique, typiquement l'énergie cinétique acquise par un électron accéléré par une différence de potentiel d'un volt.

(b) CMOS (Complementary Metal Oxide Semiconductor) est une technique de fabrication de composants électroniques.

(c) Il est probable qu'un tel avantage énergétique pourra être conservé dans les évolutions à court terme du calcul quantique de type "Noisy Intermediate Scale Quantum" ou NISQ, qui utilise des calculateurs exploitant quelques dizaines de qubits physiques en l'absence de toute correction d'erreur. Certains algorithmes quantiques présentent déjà un intérêt pratique.

(d) En toute rigueur, il faut tenir compte des températures aux différents étages du cryostat.

(e) Dans les qubits supraconducteurs, il s'agit notamment de l'excitation de niveaux d'énergie plus élevés.

(f) Ainsi, la factorisation d'une clef de 2048 qubits requiert l'activation de $L = 2048^2$ portes, chacune ayant une probabilité d'erreur p_{err} . Pour que l'algorithme total ait une probabilité d'erreur inférieure à 1/3, il convient que $p_{\text{err}} < 1/3L \approx 10^{-6}$, largement en deçà des capacités technologiques actuelles (autour de 10^{-4} [5]).

(g) La puissance de Carnot s'écrit : $P_C = (T/T_C) N_q P_{\text{CMOS}} \sim 1 \text{ MW}$, où on a introduit le nombre de qubits physiques $N_q = 10^6$ (typique pour un processeur résistant aux pannes), la puissance dissipée $P_{\text{CMOS}} = 100 \mu\text{W}$, et le facteur de magnification $(T/T_C) = 10^4$.