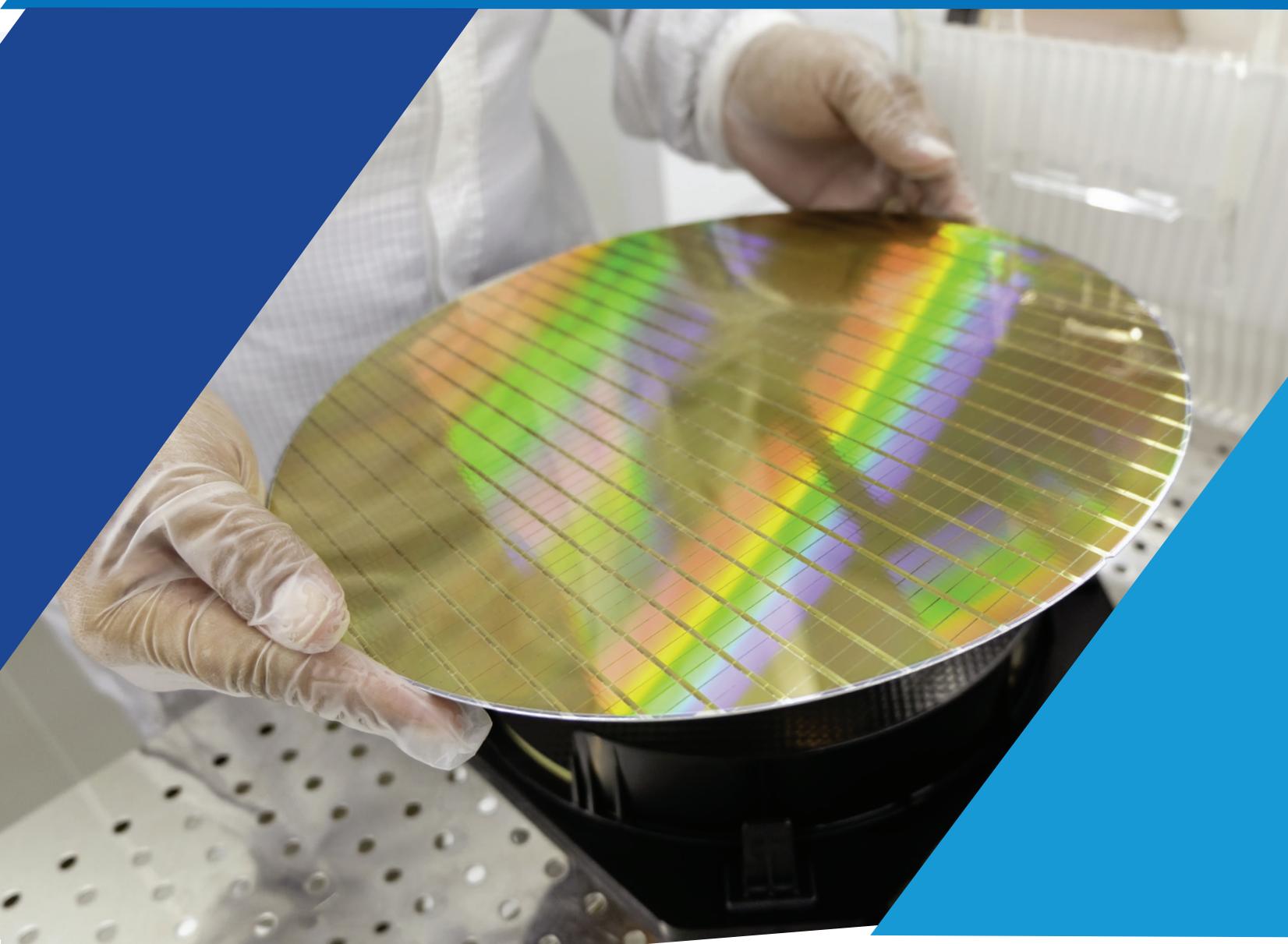


LES TECHNOLOGIES MATÉRIELLES SUPPORTS DU NUMÉRIQUE FUTUR

TROIS QUESTIONS

Rapport de l'Académie des technologies



Académie des technologies
Le Ponant – Bâtiment A
19, rue Leblanc
75015 PARIS
+33(0)1 53 85 44 44
secretariat@academie-technologies.fr
www.academie-technologies.fr

ISBN : 979-10-97579-38-8

photo de couverture :
©l'm Thongchai - stock.adobe.com

ACADÉMIE
POUR UN PROGRÈS
RAISONNÉ
CHOISI
PARTAGÉ

DES
TECHNOLOGIES

**LES TECHNOLOGIES MATÉRIELLES
SUPPORTS DU NUMÉRIQUE FUTUR**

TROIS QUESTIONS

RAPPORT DE L'ACADÉMIE DES TECHNOLOGIES

AVANT-PROPOS

Pendant une quarantaine d'années, la croissance fulgurante de la puissance des processeurs a été caractérisée par ce qu'il a été convenu d'appeler la loi de Moore. Cette loi empirique a été fondée sur l'observation des industriels des semi-conducteurs qui ont été capables de doubler tous les dix-huit mois, la densité de transistors gravés sur le silicium. Dans le même temps, la densité d'information pouvant être stockée sur des supports magnétiques puis électroniques n'a cessé de croître tout aussi rapidement tout en proposant des temps d'accès de plus en plus réduits. C'est la combinaison de ces deux phénomènes à évolution exponentielle, et exploitée par les logiciels les plus divers, qui est à l'origine du formidable développement de l'informatique et des télécommunications autrement dit du numérique.

Toutefois, depuis quelques années, le maintien du rythme de ces évolutions est confronté à des difficultés considérables. Dans le cas de la miniaturisation des transistors, les limites imposées par la puissance dissipée et la densité d'intégration, complexifient de plus en plus les progrès possibles. La problématique de stockage rapide et d'archivage de l'information numérique est, elle, confrontée au phénomène de « déluge de données » lié à la dissémination de capacités de calcul dans les objets les plus divers ainsi que par l'usage de l'Internet. La prise en compte de ce phénomène fait apparaître des besoins critiques en matériaux, énergie, eau et ressources financières.

Faut-il pour autant en conclure que la formidable puissance du numérique qui a permis et permet encore de transformer les individus et les sociétés atteint actuellement son point culminant ? Si une réponse négative peut être fournie à cette question alors quelles sont les technologies qui peuvent se substituer aux techniques actuelles et quand ?

La littérature regorge de rapports, articles scientifiques ou articles de presse pour proposer de nouvelles solutions. Dans ce contexte, l'Académie des technologies a souhaité, de manière totalement indépendante, faire le point sur le réalisme de prédictions largement diffusées, parfois utopiques. Au-delà des aspects purement scientifiques de ces prédictions, c'est

leur capacité à se concrétiser sur le plan industriel dans les dix à quinze ans à venir qui constitue le fondement de l'analyse.

Il ne s'agit pas d'évoquer ici toutes les pistes qui sont actuellement suivies pour proposer un futur aux phénomènes de croissance exponentielle passés en matière de puissance de calcul et de capacité d'archivage. En tirant parti de la compétence de ses experts, l'Académie a cherché à approfondir trois questions fréquemment débattues :

- la loi de Moore, jusqu'à quand et comment profiter des évolutions en cours dans la conception et la fabrication de composants pour réduire la dépendance de l'Europe ?
- l'usage de polymères comme l'ADN est-il une alternative crédible pour stocker de l'information en très grande quantité à des coûts réduits ?
- quelles pistes le calcul quantique peut proposer pour pallier les limites de la physique traditionnelle et offrir une solution réalisable à moyen terme pour soutenir la croissance de la puissance de calcul dans les domaines d'application où il peut apporter une solution ?

Il est intéressant de remarquer que toutes les réponses aux questions posées ont un point commun. Pour continuer à améliorer les performances des systèmes numériques, les solutions envisageables actuellement pour le moyen terme reposent sur l'emploi de dispositifs dédiés à chaque type d'application considérée et ceci quelle que soit la technologie employée (semi-conducteurs, polymères, quantique). Cet aspect est repris en conclusion de ce rapport.

Je tiens à remercier chaleureusement les académiciens Joel Hartmann, François Képès, Alain Pouyat et Boris Bourdoncle chargé d'études (lors de l'écriture de ce texte), pour avoir réuni les différents éléments qui permettent de répondre aux questions énoncées et en avoir rédigé une synthèse. Les compléments d'analyse formulés par Claude Weisbuch et Daniel Kaplan ont permis de préciser certains argumentaires. Les remarques et suggestions formulées par Albert Benveniste, Gérard Grumblatt, Marko Erman, François Lefauveux, Claude Weisbuch et le Comité de la Qualité de l'Académie des Technologies ont été aussi très utiles.

*Gérard Roucairol,
Président honoraire de l'Académie des technologies,
Président du pôle numérique*

TABLE DES MATIÈRES

AVANT-PROPOS	v
LA LOI DE MOORE DE L'INDUSTRIE DES SEMI-CONDUCTEURS	1
Introduction	3
La loi de Moore	4
Émergence, représentations et innovation	4
Adéquation produits-technologies et marché associé : vers un investissement optimum	7
Les limites à la loi de Moore	10
De « <i>More Moore</i> » à « <i>More than Moore</i> »	12
Des alternatives à la miniaturisation	13
D'autres alternatives à la miniaturisation : les circuits neuromorphiques et la photonique	18
La place de l'Europe dans l'industrie des semi-conducteurs	24
La stratégie américaine	25
Chips American Act	26
Le « silicium personnalisé »	27
Vers plus de souveraineté européenne ?	31
European Chips Act	32
La forte dépendance au numérique, de l'industrie européenne. Exemple : l'automobile	34
Conclusion pour l'Europe	35
UN NOUVEAU SOUFFLE POUR L'ARCHIVAGE DES BIG DATA	39
Motivation	41
La sphère globale des données	41
Centre de données	43
Pistes	43
Bref historique du stockage de données sur l'ADN	44
L'ADN	45

Avantages du stockage d'information sur l'ADN	46
Processus du stockage d'information sur ADN	48
Perspectives	53
Perspectives techno-scientifiques	54
Perspectives économiques	55
Perspectives nationales et européennes	56
LE CALCUL QUANTIQUE CONCEPTS ET ÉTAT DES LIEUX	59
Introduction	61
Concepts	63
Première et deuxième révolutions quantiques	63
Les technologies de l'information quantique	64
Calculabilité et complexité	66
Erreurs et bruits	68
Ordinateur quantique universel et calculateur spécialisé	69
Calcul analogique et calcul numérique	71
État des lieux	72
Les réalisations physiques de qubits	72
Les métriques de performance	74
Les entreprises de l'informatique quantique	75
Les plans nationaux et les investissements	79
Conclusion	80
CONCLUSION	83

LA LOI DE MOORE DE L'INDUSTRIE DES SEMI-CONDUCTEURS

VERS PLUS DE SOUVERAINETÉ EUROPÉENNE !

Joël Hartmann

Alain Pouyat

Résumé

À partir des années soixante, l'industrie de la microélectronique a connu un essor majeur porté par la course à la puissance de calcul. Basée sur les matériaux semi-conducteurs et mettant en œuvre une découverte scientifique nobélisée — l'effet transistor —, cette course fut possible grâce à une miniaturisation constante de la taille des composants élémentaires, reflétée par la loi dite de « Moore ».

Cependant, la hauteur des efforts financiers, techniques et humains pour soutenir cette course devint telle que quelques acteurs prirent la tête en investissant massivement et de façon continue. Aujourd'hui, pour des raisons économiques et physiques, le cycle vertueux entre investissements et bénéfices sous-tendu par la loi de Moore s'est ralenti et ne concerne plus que deux acteurs essentiellement pour les processeurs universels.

Mais, de façon concomitante, l'accès au grand public a fait surgir de nouveaux besoins et marchés appelant plus de fonctionnalités et ne requérant pas une miniaturisation extrême des composants élémentaires. Le mot d'ordre « *More than Moore* » remplace l'injonction « *More Moore* ».

Donc, par-delà la loi de Moore, en bénéficiant de la dynamique et du fort développement d'usages nouveaux, tels que l'IoT, la 5G, l'industrie 4.0, le véhicule autonome ou encore l'intelligence artificielle, les sociétés européennes de microélectronique, peuvent se positionner de manière plus dynamique et permettre ainsi aux industriels de ces secteurs d'être plus innovants et compétitifs sur leurs marchés. Pour renforcer et développer ses industries, l'Europe ne peut passer à côté de l'opportunité historique que constitue la révolution actuelle des composants et notamment celle du « silicium personnalisé ».

INTRODUCTION

La démultiplication et le foisonnement actuel des applications électroniques prennent naissance avec la première utilisation industrielle de l'électricité en 1799 et l'invention de la pile électrique par le professeur Volta. Cette découverte, qui a révolutionné le XIXe siècle, a grandement contribué à l'amélioration de la condition humaine du fait de son déploiement généralisé à travers des usages de plus en plus complexes et diversifiés. Cette révolution électrique sera suivie, un siècle et demi plus tard, par l'avènement de l'électronique.



Figure 1: Image du premier transistor réalisé en 1947 dans les laboratoires Bell aux États-Unis, inventé par les chercheurs américains William Shockley, John Bardeen et Walter Brattain dont les travaux ont été reconnus par un prix Nobel de Physique en 1956.

source : clintonwhitehouse4.archives.gov

L'aventure de l'électronique moderne prend très certainement naissance en 1947 avec l'invention du premier transistor aux États-Unis, dans les Bell Labs, par William Shockley, John Bardeen et Walter Brattain. Cette invention repose notamment sur la mise en œuvre de matériaux de base dont les caractéristiques électriques peuvent être altérées à dessein pour obtenir soit les propriétés d'un conducteur, soit celles d'un isolant : on parle alors de matériau semi-conducteur. Parmi les éléments présentant ces propriétés électriques se trouve le silicium, élément chimique très répandu dans la nature et bon marché.

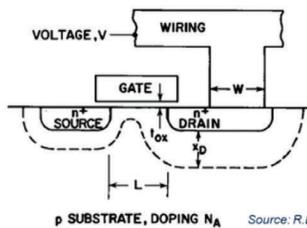
LA LOI DE MOORE

L'industrie des semi-conducteurs est née de la capacité à combiner à grande échelle ces dispositifs électroniques sur un même substrat en silicium et à produire ainsi une large gamme de circuits électroniques dits « intégrés » (par opposition aux circuits dits imprimés, combinaison de circuits discrets soudés sur une carte organique).

ÉMERGENCE, REPRÉSENTATIONS ET INNOVATION

Cette intégration allait connaître à partir des années soixante une accélération grâce à la réduction systématique des dimensions physiques de ces composants. La régularité de cette réduction se trouva formulée en 1965 par Gordon Moore, un des cofondateurs de la société Intel Corporation, en une loi empirique portant son nom, qui prédit le doublement du nombre de transistors par mm^2 de silicium tous les deux ans. La loi de Moore rythme ainsi depuis des décennies la mise sur le marché de circuits de plus en plus denses et performants qui sous-tendent des systèmes de plus en plus complexes. Cette croissance, au cours des cinquante dernières années, est illustrée par la Figure 2.

Paramètres critique pour les circuits	Facteur de reduction d'échelle
Dimension transistor (T_{ox} , L , W)	$1/k$
Concentration des dopant (N_A)	k
Tension (V)	$1/k$
Courant (I)	$1/k$
Capacité (C)	$1/k$
Temps de propagation ($C \cdot V / I$)	$1/k$
Puissance dissipée ($V \cdot I$)	$1/k^2$
Densité de puissance ($V \cdot I / A$)	1



Source: R.DENNARD (IEEE) - 1974

Source: A.ASENOV (université de Glasgow) - 2015

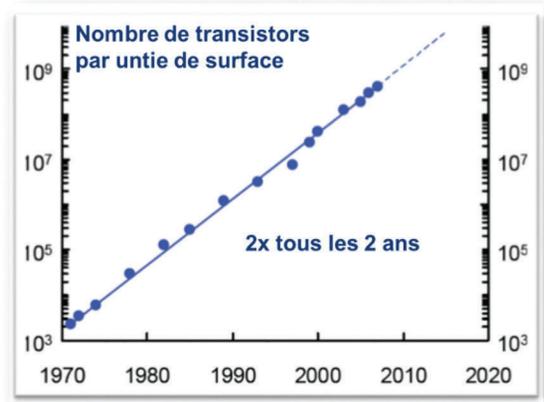


Figure 2 : Illustration des paramètres critiques du transistor sur le facteur de réduction d'échelle (à gauche) et croissance exponentielle du nombre de transistors par unité de surface (à droite)

source : A. Asenov (université de Glasgow)-2015

En 1974, Robert H. Dennard, chercheur d'International Business Machines Corporation (IBM), a modélisé le facteur de réduction d'échelle (appelé k dans la partie gauche de la Figure 2). Ce coefficient permet d'exprimer de multiples bénéfices : économique grâce à la miniaturisation, et technique avec l'amélioration simultanée de la vitesse de calcul et de la consommation d'énergie. Ainsi, pour une contraction des dimensions du transistor d'un facteur k , la surface du circuit et l'énergie consommée sont divisées par k^2 et k^3 respectivement. Ce sont ces rapports extrêmement favorables de la loi de Moore qui sont à l'origine de son déploiement à grande échelle dans l'industrie des semi-conducteurs pour les circuits intégrés.

D'un point de vue économique, la mise à disposition d'une nouvelle technologie tous les deux ans induit une contraction significative du coût de fabrication d'un transistor unitaire (voir la Figure 3). Ainsi, depuis son invention, ce coût s'est réduit d'un facteur de l'ordre du million pour les processeurs et même proche du milliard pour les mémoires. C'est probablement l'une des performances économiques les plus notables obtenues par une industrie, tous domaines d'activité confondus, au cours des dernières décennies.

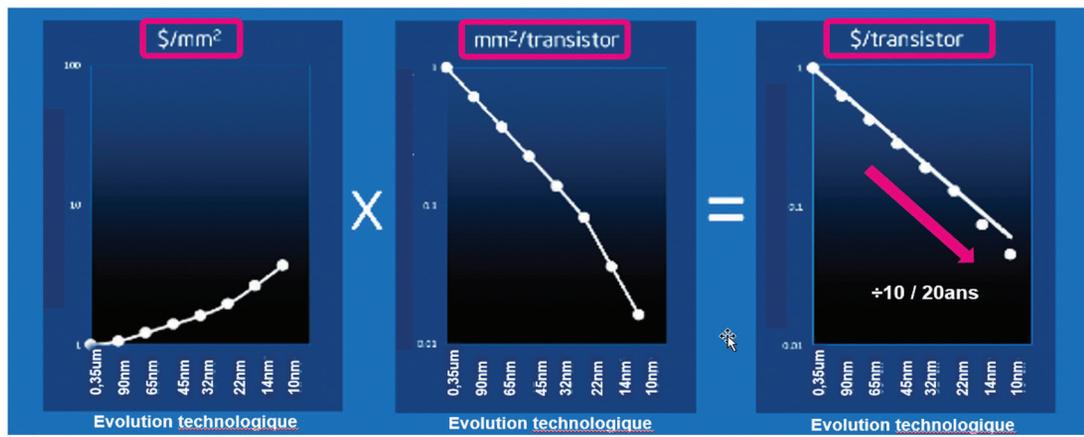


Figure 3 : Intérêt économique de la loi de Moore soutenant une contraction du prix de revient d'un transistor de -50% par génération.

source : Intel Corporation - janvier 2014

Cette performance économique induite par cet extraordinaire effort de miniaturisation s'accompagne d'une augmentation spectaculaire des performances et puissances de calcul. Comme mentionné précédemment, à chaque nouvelle génération, le facteur de réduction d'échelle k prédit une diminution simultanée du temps de calcul de $1/k$ et de l'énergie consommée de $1/k^3$. La Figure 4 illustre la transposition du gain en performance des circuits intégrés à la consommation kilométrique d'une voiture.

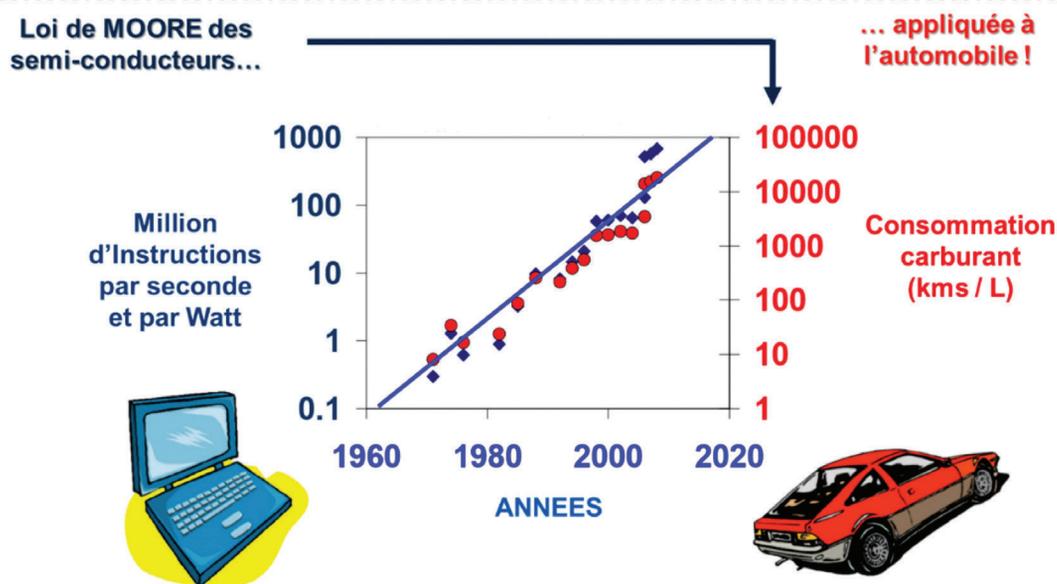


Figure 4 : Illustration comparative de l'amélioration exponentielle de l'efficacité énergétique des circuits électroniques (environ 1 décade tous les 10 ans) avec sa transposition théorique à une automobile.

source : Jupiterimage Corp—2009

La marche visible sur le graphique de la figure 4 (points bleus au-dessus de la droite de régression vers l'année 2010) correspond à l'adoption de la technologie (ou « nœud technologique ») 7 nm, une finesse de lithographie qui met en œuvre une rupture technique majeure. De quoi s'agit-il ? Un circuit intégré est réalisé physiquement de façon additive par l'enchaînement consécutif de couches et de motifs. L'ensemble des motifs issus de la conception informatique du circuit est transféré au fur à mesure de la réalisation. Ce passage du monde de la conception numérique au monde physique sur silicium s'opère à l'aide de masques qui sont éclairés par une lumière particulière pour projeter les motifs sur des résines photosensibles. Cette étape clef, appelée lithographie, repousse les limites de la physique en mettant dorénavant en œuvre une lumière ultraviolette extrême t . Ceci permet de maîtriser le nombre de passages des plaques de silicium sur les équipements de lithographie (très coûteux), de garantir les rendements et de contenir le temps de cycle de développement et de fabrication des circuits intégrés (déjà de plusieurs mois). Cependant, ces équipements de lithographie EUV sont extrêmement onéreux à l'achat et en entretien, et le coût des masques à mettre en œuvre devient également prohibitif : si le prix d'un jeu de masques pouvait être de l'ordre d'un million d'euros dans les gammes de technologies entre 32 et 22 nm, le prix en revanche s'envole très largement au-delà de plusieurs millions d'euros pour les nœuds technologiques inférieurs à 10 nm.

À cet investissement en équipements et en infrastructure de salle blanche se couple aussi l'effort associé à l'apprentissage. Une fois ces équipements disponibles, il faut prévoir de nombreux lots successifs de plaquettes de silicium pour développer, prototyper puis stabiliser et qualifier les briques innovantes des nouvelles technologies. Cet apprentissage permettant d'ouvrir ces solutions au prototypage de produits d'abord puis à la production de volume ensuite, nécessite l'utilisation d'un nombre croissant de plaquettes ; en effet, les outils de modélisation et de simulation actuels ne sont pas capables de rendre compte avec précision des comportements et des interactions électroniques, quantiques, thermiques et électromagnétiques à ces échelles nanométriques.

Toutefois, cette course à la réduction des dimensions conduit à relever des défis technologiques toujours plus ardues lorsque l'on se rapproche de l'échelle atomique. Dans les années soixante-dix, la jonction d'un transistor faisait de l'ordre d'une dizaine de micromètres. Aujourd'hui, il est question d'une dizaine de nanomètres, soit mille fois moins qu'à l'époque. Cela n'est possible qu'au prix d'une innovation technique grandissante couplée à une innovation dans l'architecture même des transistors : les transistors au cœur des produits numériques aujourd'hui (transistor à effet de champ dit FET – Field Effect Transistor) ont ainsi connu une évolution très significative comme illustrée dans la Figure 5.

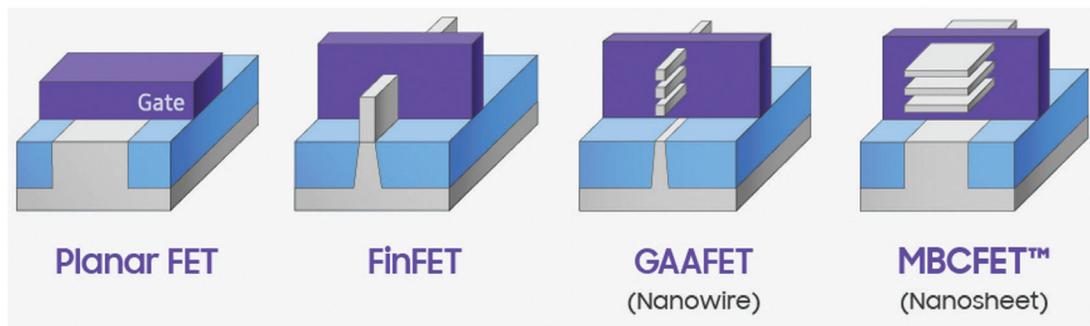


Figure 5 : représentation des différentes architectures de transistors avancés: transistor planaire (incluant les substrats FDSOI), transistor à ailettes (FinFET), transistor à grille entourant des nanofils (GAAFET) ou des nanofeuilles (MBCFET).

Source : Samsung Foundry

ADÉQUATION PRODUITS-TECHNOLOGIES ET MARCHÉ ASSOCIÉ : VERS UN INVESTISSEMENT OPTIMUM

La course à la miniaturisation visait au départ à accompagner des besoins croissants en puissance de calcul dans les secteurs scientifiques, académiques, industriels et non civils, mais l'accessibilité croissante de l'industrie électronique au grand public a fortement contribué à son plein essor.

Cela s'est traduit par une pénétration massive de ces circuits électroniques intégrés dans les objets du quotidien (TV/multimédia, téléphonie, PC, Internet... mais aussi de l'industrie (contrôle numérique, systèmes MRP, robotisation). Depuis le début des années 2010, leur prolifération se matérialise selon trois axes : le décuplement du nombre d'objets connectés, le lancement de la voiture (semi) autonome et l'émergence de l'intelligence artificielle.

Cette course à la réduction des dimensions conduit à relever des défis technologiques toujours plus ardues lorsque l'on se rapproche de l'échelle atomique, et requiert donc des investissements financiers et humains de plus en plus conséquents. Afin de se positionner au mieux dans cette course et de dominer ainsi le marché associé, quelques acteurs prirent le pari de soutenir cet effort au cours des dernières années par le biais d'investissements annuels dépassant la dizaine de milliards de dollars.

Alors que l'utilisation des circuits intégrés est planétaire, la géographie de la recherche et de la production est, elle, très inhomogène. En effet, pour soutenir le rythme effréné de la miniaturisation, les énormes investissements financiers et humains nécessaires favorisent la concentration des moyens, facteur essentiel des économies d'échelle, et réduisent *in fine* le nombre d'acteurs capables de financer cette évolution. À ce jour, seules deux firmes, toutes deux asiatiques, sont capables d'offrir des technologies de génération inférieures à 7 nm : Samsung et TSMC. Concomitante à une tension géopolitique accrue affectant le commerce mondial des circuits intégrés, cette récente ultraconcentration pose nécessairement la question de l'autonomie stratégique de la France et de l'Europe face à ces technologies avancées. Ces questions sont étroitement liées à la capacité d'autofinancement des efforts de R & D par l'industrie européenne et à la disponibilité d'une ressource humaine hautement qualifiée pour maîtriser de telles technologies.

Pour cela, l'industrie doit soutenir un effort important en recherche et développement, requérant un niveau d'investissement massif, couvrant à la fois les aspects matériels (équipements de nouvelle génération, substrats de silicium, matériaux et produits chimiques) et humains (formation scientifique et technique, compétences...) afin de rendre possibles les gains promis par la loi de Moore.

Ces gains ne sont cependant obtenus qu'au prix d'investissements manufacturiers et en R & D sans cesse croissants. La pression financière et en moyens humains nécessaires à ces développements a connu une accélération soudaine au-delà du nœud 10 nm. Cette marche est décrite dans la Figure 6 et montre une augmentation de 50 % de l'investissement en nouveaux équipements de salle blanche pour la fabrication des puces dès l'année 2017.

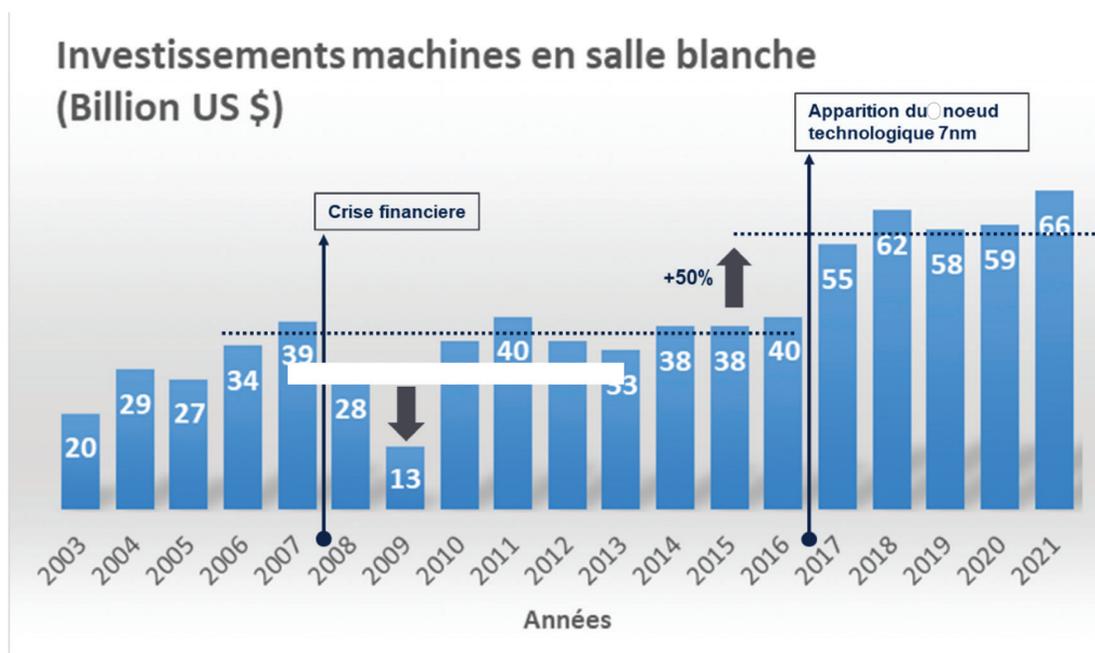


Figure 6 : Évolution de l'investissement mondial réalisé dans les usines de fabrication de circuits intégrés sur silicium : fort accroissement des montants investis au-delà de la technologie 10 nm

source : blog.semi.org - 2020

Ainsi, pour réaliser le développement et l'industrialisation d'une technologie 7 nm, il est communément admis qu'il faut réaliser un apprentissage sur plusieurs dizaines de milliers de plaquettes de silicium. Enfin, en marge de ces aspects matériels déjà lourds, ces développements complexes requièrent l'accès à une source abondante d'ingénieurs très qualifiés.

Pour compléter l'inventaire des différents aspects de l'investissement pluriel à consacrer au soutien de la loi de Moore, il nous faut considérer également ceux liés aux activités de conception numérique des produits. De façon inhérente à la densification des circuits, le niveau de complexité des schémas électroniques explose en fonction de la réduction des dimensions. La gestion de cette complexité de conception des circuits intégrés entraîne elle aussi un accroissement exponentiel des coûts associés. Ainsi, comme illustré par la Figure 7, on estime que le surcoût de la conception d'un produit est multiplié par dix entre une technologie dont les règles de dessin sont en 28 nm et celle en 5 nm.

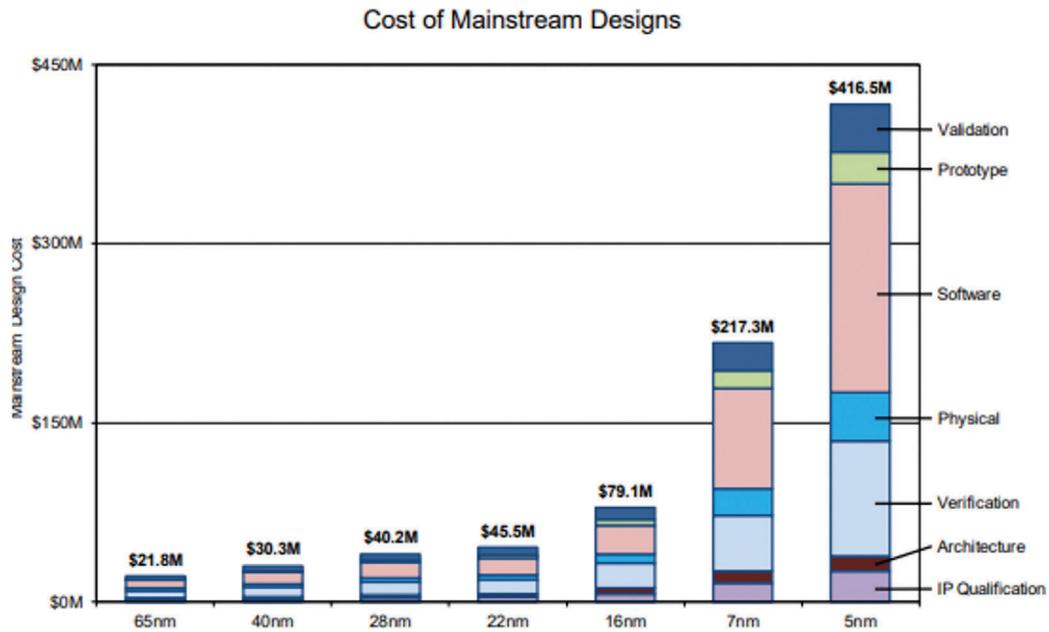


Figure 7 : Diagramme décrivant l'évolution du coût associé aux activités de conception des circuits intégrés en fonction du nœud technologique associé.

Source : IBS

Au regard de ces différents aspects, nous pouvons observer que l'ensemble des contraintes financières associées n'a pas, malgré tout, suffi à mettre en défaut la loi de Moore de l'industrie des semi-conducteurs : on parle aujourd'hui de technologie 2 nm en phase de développement amont chez les acteurs les plus avancés. Ceux-ci firent le pari de soutenir leur effort dans cette course durant les dernières années par le biais d'investissements annuels au-delà d'une dizaine de milliards de dollars, l'objectif étant de se positionner au mieux et ainsi de dominer le marché mondial.

LES LIMITES À LA LOI DE MOORE

Aujourd'hui, il reste deux grands acteurs asiatiques dans cette course, le Taïwanais TSMC et le Coréen Samsung. Ainsi, même la compagnie américaine Intel — à l'origine de la loi de Moore — va se tourner dorénavant vers TSMC dont la construction d'une usine de fabrication sur le territoire américain (en Arizona) à l'état de l'art de l'industrie (5 nm) est en cours. Pour illustrer cette compétition par élimination, nous voyons sur la Figure 8 l'évolution du nombre d'acteurs de l'industrie des semi-conducteurs dans le temps, et donc au fil des générations successives des nœuds technologiques .

On voit très clairement sur la Figure 8 qu'au cours des dix dernières années, une première sélection entre les acteurs s'est opérée lors du passage de 45 nm à 28 nm (migration vers des grilles à forte constante diélectrique), puis une seconde lors du basculement à 22 nm, avec l'introduction de l'architecture des transistors à ailettes (FinFET). La prochaine rupture sera celle des transistors à grille entourant des nanofils/nanofeuilles (GAA, MBCFET) tels qu'illustrés dans la Figure 5.

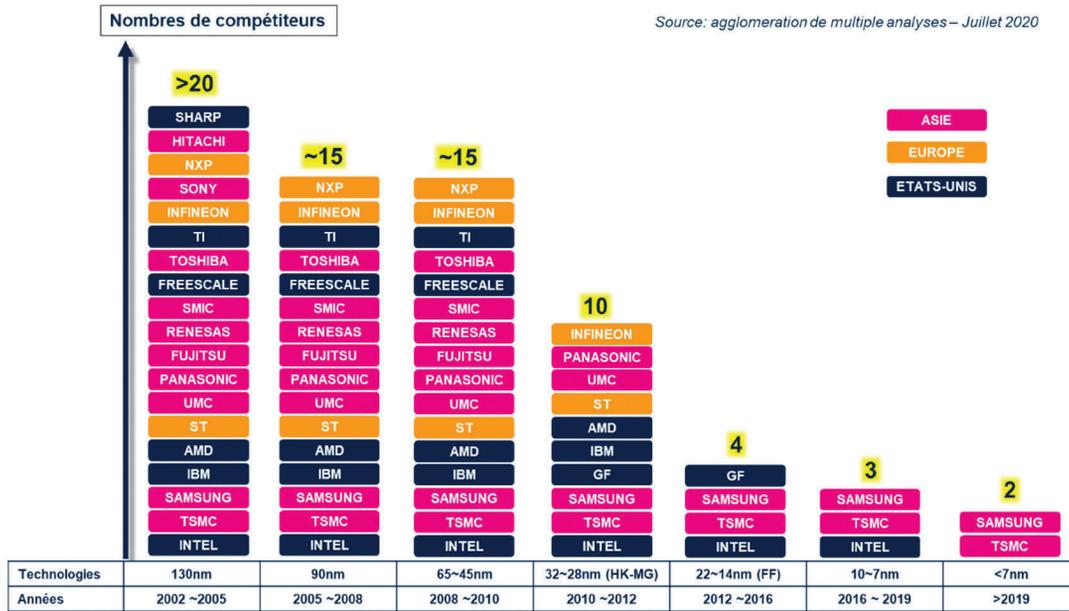


Figure 8 : Diagramme mettant en évidence la contraction du nombre de participants à la course à la réduction d'échelle.

La miniaturisation a rendu possible l'amélioration régulière de plusieurs paramètres. L'un d'entre eux est la fréquence d'horloge, liée au temps nécessaire pour effectuer une opération de base. La croissance de la fréquence d'horloge assurée pendant plusieurs décennies a ainsi permis d'augmenter la vitesse de calcul. Toutefois, cette fréquence a atteint une valeur limite au début des années 2000, au-delà de laquelle la chaleur dissipée devient trop importante. Pour continuer à tirer parti de la miniaturisation pronostiquée par Moore sans augmenter la fréquence d'horloge, les processeurs multicœurs se sont développés. En permettant de traiter plusieurs ensembles d'instructions en parallèle, ceux-ci ont permis de continuer à faire croître la vitesse de calcul sans augmenter la fréquence d'horloge. Toutefois, pour un programme d'instructions donné, profiter de l'architecture multicœurs nécessite de le réécrire en blocs d'instructions pour permettre leur exécution en parallèle, ce qui n'est pas évident et pas toujours possible.

Les limites imposées par la puissance dissipée et la densité d'intégration, complexifient de plus en plus les progrès possibles en matière de miniaturisation des transistors. Sur le plan de l'intégration, des problèmes nouveaux apparaissent. L'extrême densité des transistors (plusieurs dizaines de millions par mm² en technologie 5 nanomètres) impose d'empiler

de plus en plus de niveaux de métallisation pour les interconnecter, ce qui a tendance à faire perdre en partie l'avantage de l'augmentation de la performance des transistors par l'augmentation des délais de propagation des signaux dans les circuits. L'impossibilité de continuer à réduire régulièrement en 2D les dimensions critiques de la lithographie nécessite de passer à des solutions innovantes d'intégration en trois dimensions des transistors.

Par ailleurs, l'industrie des semi-conducteurs se trouve aujourd'hui confrontée à une autre limite physique : les transistors nanoscopiques qui figurent à la droite de la Figure 5 approchent de la taille de l'atome (0,1 nm). À cette échelle, le comportement des particules est décrit par la physique quantique, dont les lois non déterministes mettent en défaut le fonctionnement attendu des transistors développés jusqu'alors. Bien que, pendant plusieurs décennies, la miniaturisation seule ait suffi à stimuler l'innovation, en permettant d'inventer de nouveaux usages et en assurant ainsi la croissance, l'avenir de l'industrie des semi-conducteurs doit désormais se construire par-delà la loi de Moore.

DE « *MORE MOORE* » À « *MORE THAN MOORE* »

Cette course à la miniaturisation visait au départ à accompagner des besoins croissants en puissance de calcul dans les secteurs scientifiques, académiques, industriels et non civils. Aujourd'hui, elle répond aux besoins en puissance de calcul des microprocesseurs ou processeurs de cartes graphiques, que l'on trouve dans les ordinateurs, les tablettes, les téléphones intelligents (smartphones), les consoles de jeux, les serveurs.

Ainsi, l'accès des produits électroniques au plus grand nombre a accéléré l'innovation « produit » et a suscité l'émergence de nouveaux besoins utilisateurs qui n'étaient pas concevables jusqu'alors. Cette prolifération d'applications s'accompagne d'une diversification en termes d'exigences d'intégration. L'enjeu n'est plus seulement de miniaturiser, mais aussi de faire cohabiter sur une même puce des fonctions différentes et complémentaires afin de créer un système complet répondant à une application déterminée.

Cette nouvelle tendance, qui dépasse donc la seule miniaturisation, porte le nom de « *More than Moore* » (plus que Moore). Elle permet maintenant de développer des technologies qui apportent des solutions plus globales en intégrant sur un même composant une grande variété de fonctions électroniques et de proposer des produits optimisés tant en coût de production, qu'en performance globale du système, tout en accélérant leur introduction sur le marché.

DES ALTERNATIVES À LA MINIATURISATION

Ainsi, alors que durant les cinquante dernières années, la loi de Moore a motivé la recherche d'une plus grande performance des fonctions gravées sur la puce (processeur de plus en plus puissant ou mémoire de plus en plus grande), ces nouvelles possibilités permettent :

- soit une forte intégration des différentes fonctions d'un système sur une même puce de silicium. Ce sont les : *System On Chip (SOC)* ;
- soit à l'inverse, un système complexe et « décomposé » (dés-intégré), en fonctions plus simples pour être ensuite « réintégré » en un seul composant et ainsi ne payer que pour le service effectif des nœuds technologiques les plus adaptés. Ce sont les *System in Package (SiP)*.

La Figure 9 illustre cette tendance récente vers une intégration dite hétérogène (*a contrario* de monolithique).

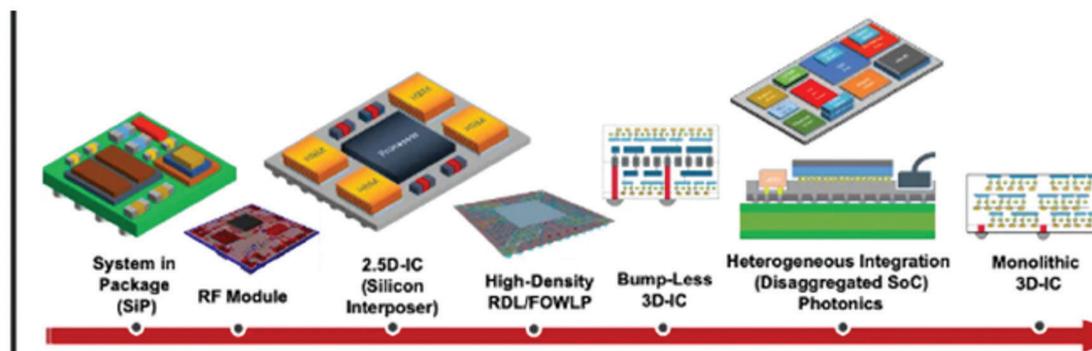


Figure 9 : Exemple de *roadmap* d'assemblage hétérogène

Source : CADENCE

Les Systèmes intégrés sur une puce (SOC) & le Parallélisme

Grâce à la miniaturisation poussée, il devient possible désormais de graver sur une même puce un ensemble de fonctions (ou de circuits) « hétérogènes » (ou différentes) permettant de composer un système complet, appelé SOC (*System On a Chip*) dédié à une application spécifique. Par exemple, on peut mettre pratiquement tout le contenu d'un ordinateur classique (processeur, carte graphique, mémoire), sur une seule et même puce, alors que, auparavant, tous ces composants distincts étaient réunis sur : un circuit imprimé, la carte « mère » d'un ordinateur.

L'exemple suivant (Figure 10) est celui d'un smartphone où toutes les fonctions principales sont réunies en un seul composant.



Figure 10 : SOC équipant les mobiles 5G gravé en 5 nm

La puce « Snapdragon » de Qualcomm est l'exemple d'un SOC gravé en 5 nm et composé de :

- un processeur ARM 4 cœurs ;
- un processeur graphique GPU ;
- un processeur neuronal NPU (*Neuronal Processing Unit*) dédié à l'intelligence artificielle pour la reconnaissance de son et d'image ou encore la reconnaissance de contexte, utile pour la caméra et l'assistant connecté ;
- un processeur de traitement d'image (ISP) qui gère les différents capteurs photo et vidéo et qui rend accessibles certaines fonctions comme le ralenti, le HDR ou le mode portrait ;
- de la mémoire centrale ;
- des modems 5/4/3/2G, GPRS, Edge, HSDPA, USB 3.1 ;
- des communications radio : WiFi, Bluetooth, RFID ;
- la localisation : GPS, GLONASS, Beidou, Galileo, QZSS, SBAS.

Dans la suite du document, on verra d'autres exemples de SOC hétérogènes comme la puce M1 de Apple ou la puce FSD de Tesla.

De la même manière, il devient également possible de réaliser des SOC multicœurs permettant de développer des processeurs massivement parallèles. Ces SOC sont indispensables au développement des supercalculateurs et à la course à l'Exaflop, qui nécessite d'assembler plus d'un million de cœurs. C'est l'exemple de Kalray et le parallélisme massif (*MPPA : Massively Parallel Processor Array Architecture*). Cette société française propose des processeurs de 80 à 256 cœurs utilisables en parallèle pour des applications de réseau 5G, de robotique, d'IA et d'aide à la conduite (Figure 11).

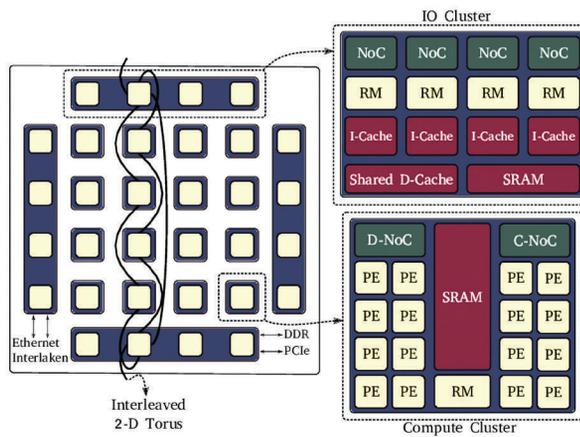


Figure 11 : architecture d'une puce Kalray de 256 processeurs !

Les coprocesseurs graphiques de Nvidia sont un autre exemple de processeurs massivement parallèles selon une architecture totalement différente. Cette dernière est spécialisée pour exécuter des opérations simultanées sur les composantes de données régulières comme un vecteur ou un tableau. Au-delà des traitements graphiques, les possibilités offertes par cette architecture trouvent un large domaine d'applications qui relèvent du calcul à haute performance ou de l'entraînement de réseaux de neurones.

Bien sûr les algorithmes qui utilisent ces différents types de processeurs diffèrent selon l'architecture de ces derniers. Ceci entraîne une tendance lourde vers la spécialisation du triplet *domaine d'application- classe d'algorithmes- architecture de processeurs*.

Les Systèmes encapsulés (SIP)

Dans le cas des SIP (*systems in package*), contrairement aux SOC, seules les fonctions de traitement du signal et processeurs sont réalisées en technologie avancée pour optimiser la consommation et les performances tout en réduisant le temps de conception et les risques de redesign. Les fonctions complémentaires du produit sont implémentées dans des structures silicium à la fois plus établies et matures, donc moins onéreuses, mais néanmoins suffisamment performantes pour la spécification envisagée. Un substrat de silicium assure la connexion entre les différents circuits et garantit un encombrement spatial minimal.

Être ainsi capable d'empiler des puces en 3D, donne la possibilité de réutiliser, pour des fonctions plus accessoires, des puces moins performantes (voire même des puces existantes), mais suffisantes pour l'application et donc offrir un meilleur retour sur investissement. C'est l'opportunité de créer des systèmes performants, de nouvelle génération, sans être contraint de maîtriser les filières sub-10 nm et d'en supporter les investissements considérables.

La Figure 12 illustre l'exemple d'un produit réalisé avec une architecture hétérogène.

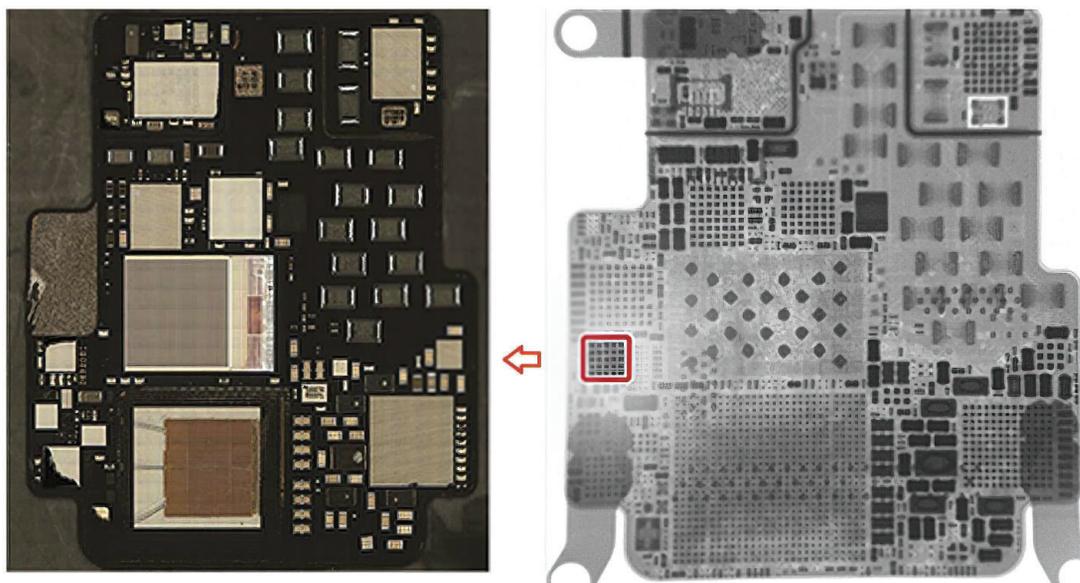


Figure 12 : l'Apple Watch (S2) contient un SiP comprenant plus de 42 circuits assemblés par la société ASE

source : Apple

Les techniques d'intégration hybrides ont fait beaucoup de progrès récemment (Figure 13), et ont rendu possible une intégration 3D, avec une densité pratiquement identique à du monolithique, mais avec la possibilité de faire des optimisations plus poussées. Progressivement cela a permis d'aboutir à la technologie « *Wafer-Level Packaging* » (WLP), c'est-à-dire réaliser le packaging du (ou des) composant directement sur le silicium.

L'activité de ces « *Advanced Packaging* » ne cesse de croître, passant de 20 milliards de dollars de chiffre d'affaires en 2014 (représentant 38 % de l'activité de packaging), à \$29B en 2019 (soit 43 %), et on prévoit plus de \$40B en 2025 soit près de 50 % du chiffre d'affaires mondial de l'activité packaging. La valeur stratégique est encore plus importante, car la fabrication de SiP est couplée aux activités de conception et surtout de test. Elle devient de fait une activité à forte valeur ajoutée.

Les acteurs industriels des SiP viennent de deux horizons :

- assez naturellement, du monde de l'assemblage ou du test, appelés OSAT (*Out-sourced Semiconductor Assembly and Test*). Les sociétés les plus importantes sont ASE 11,87B\$ (Taiwan) et AMKOR 4,31B\$ (US), mais aussi : JCET (Chine) 3,97B\$, SPIL 2,79B\$, Power tech (Taiwan) 2,17B\$, qui migrent du packaging classique vers le Wafer Level Packaging (WLF). Plus de 90% des usines sont en Asie.

- mais aussi du monde des « fondeurs » pour les technologies avancées de type WLF, souvent comme un plus à leurs filières. C'est en particulier le cas de TSMC. (Voir l'exemple de « Dojo » page 28)

ST Microélectronique a développé une capacité interne en particulier pour l'assemblage de ses circuits pour caméras avec des circuits laser et le numérique. C'est un bon niveau de maîtrise, mais sur un segment étroit (qui est leur « cœur business »). L'autre activité importante de ST, les microcontrôleurs, ne nécessite pas ce type de packaging ; ce sont des circuits monolithiques classiques.



Figure 13 : ces schémas présentent les principales technologies actuelles permettant d'empiler des puces en 3D

(source : Industrie & Technologies février 2022)

D'AUTRES ALTERNATIVES À LA MINIATURISATION : LES CIRCUITS NEUROMORPHIQUES ET LA PHOTONIQUE

Comme on l'a vu, les ordinateurs traditionnels sont en train d'atteindre progressivement une limite dans la miniaturisation et si on veut continuer à progresser on peut aussi revoir la façon de concevoir les ordinateurs. Il y a, actuellement, au moins deux technologies émergentes qui offrent cette possibilité :

- les ordinateurs quantiques, bien sûr, qui font l'objet du dernier chapitre de ce document ;
- mais aussi d'autres approches dont on parle également beaucoup en ce moment, car elles sont très prometteuses pour le développement de l'intelligence artificielle : les ordinateurs neuromorphiques et/ou photoniques.

Les puces neuromorphiques

L'objet des ordinateurs neuromorphiques est d'émuler le fonctionnement du cerveau humain, non pas à partir de puces « classiques » comme avec les processeurs neuronaux actuels, mais avec des puces dont l'architecture s'inspire de l'humain, composées de neurones et de synapses artificiels. Ces neurones artificiels parlent les uns avec les autres, fonctionnent sans ordre prescrit, comme le font les neurones biologiques. Ils ont ainsi des capacités d'apprentissage beaucoup plus développées, et économisent beaucoup d'énergie. Par ces caractéristiques en rupture, ces circuits représentent probablement le futur de l'intelligence artificielle.

Comparé au cerveau humain, un processeur traditionnel fonctionne d'une manière totalement différente. 99 % des puces modernes utilisent l'architecture de Von Neumann. Cela signifie qu'elles sont conçues pour fonctionner de manière synchrone, que toutes les opérations sont rythmées par une même horloge et qu'elles doivent échanger en permanence avec une mémoire externe. Cette architecture consomme du temps, limite les capacités de traitement et demande beaucoup d'énergie.

À l'opposé, le cerveau humain fonctionne de manière totalement asynchrone, il n'utilise pas d'horloge spécifique et il consomme infiniment moins d'énergie.

Ainsi, à son image, un ordinateur neuromorphique traduit et transporte l'influx nerveux qui circule de neurone en neurone, tout en leur permettant de travailler tous ensemble simultanément, et seulement quand cela est nécessaire. Comme l'influx nerveux qui est constitué de rafales de sollicitations, les puces neuromorphiques, composées de neurones indépendants sont stimulées par la circulation de flux composé de bits.

Un neurone de silicium peut être réalisé par un circuit analogique ou un circuit numérique, c'est-à-dire soit par un intégrateur analogique, soit par un compteur numérique complexe. Chacun de ces neurones peut recevoir une entrée sous la forme d'impulsions provenant de n'importe quel autre neurone de la même puce ou d'une autre puce, puis il intègre les différentes sollicitations reçues et en fonction de sa programmation, lorsqu'un seuil est atteint, il utilisera le résultat pour l'envoyer à d'autres neurones.

Loihi de Intel : La puce Loihi de la société INTEL est un des premiers exemples concrets de cette conception. Le processeur neuromorphique « Loihi » développé par Intel Lab est conçu comme un circuit ouvert utilisant des zones de mémoire qui peuvent agir comme des neurones artificiels capables à la fois de calculer et de stocker des données. Ce composant est commercialisé depuis 3 ans pour l'expérimentation applicative et la recherche en laboratoire.

Constitué de 128 cœurs entièrement asynchrones, tous connectés par un réseau NOC (*network on chip*), ce circuit permet l'évolutivité et l'efficacité énergétique pour des transferts de données entre les différents cœurs. Et grâce aux interfaces NOC sur les quatre côtés de la puce, Loihi peut être connecté à d'autres puces neuromorphiques dans toutes les directions (Figure 14). Cette première version de Loihi comprend 128 000 neurones et 128 millions de synapses.

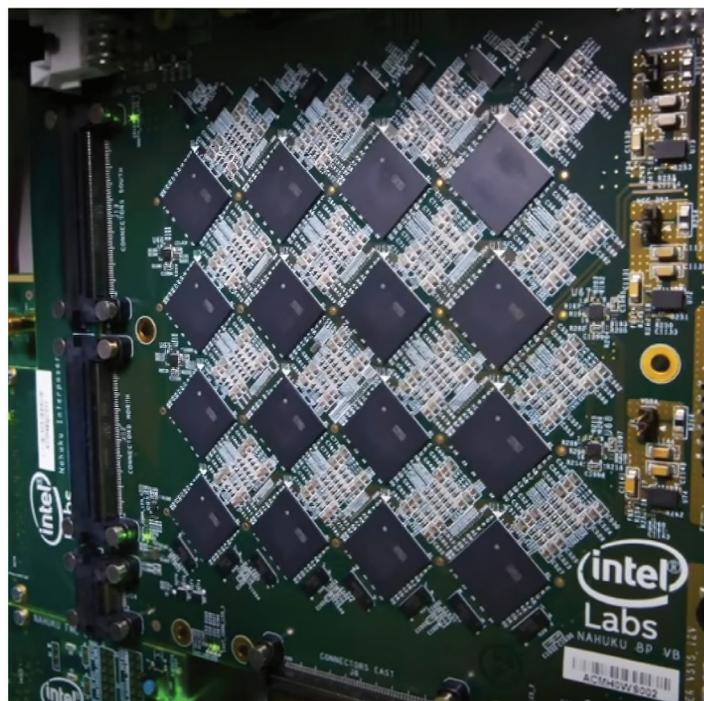


Figure 14 : Circuit imprimé qui assure l'interconnexion de 16 puces Loihi.

Même si le système proposé par Intel est très évolutif, on est encore très loin du cerveau humain qui comprend 86 milliards de neurones et 10 000 milliards de synapses. En combinant des centaines ou des milliers de puces, on peut construire un très grand système neuromorphique et se rapprocher un peu plus de l'humain.

Au cours des trois dernières années, Loihi a donné de très bons résultats dans une large gamme d'applications : reconnaissance gestuelle, cartographie, traitement d'horaires de train, repérage de mot-clé et contrôle adaptatif. Dans tous ces exemples, Loihi a consommé moins d'un watt alors que les CPUs et GPUs traditionnels en consomment des centaines. C'est donc une véritable percée en matière d'efficacité énergétique.

Fin septembre 2021, les laboratoires d'Intel ont présenté une deuxième génération : « Loihi 2 » comprenant 1 million de neurones et 120 millions de synapses.

Mais Intel n'est pas le seul sur le marché des puces neuromorphiques.

Rain Neuromorphics. Cette start-up basée en Californie développe des puces neuromorphiques ouvertes à base de réseau neuronal de nanofils constitués de mems¹ résistif. Cette puce est également hautement évolutive. Rain Neuromorphique peut composer un réseau neuronal de plus d'un milliard de neurones.

SpiNNaker. Ce supercalculateur massivement parallèle est le fruit de travaux de recherche de l'université de Manchester. SpiNNaker s'intéresse au mécanisme de l'excitation d'un réseau neuronal. À l'origine, il a été conçu davantage comme un simulateur. Mais déjà, la machine spiNNaker avec 500 000 cœurs est capable de rivaliser avec les superordinateurs traditionnels et les résultats obtenus sont impressionnants tant la consommation d'énergie est faible. La deuxième génération « SpiNNaker 2 » apporte d'importantes améliorations. Elle est basée sur un cortex ARM M4 18 cœurs, 128 kilo-octets sram local et accélérateur MAC. C'est une sorte de grande NPU (*Neural Process Unit*) qui a probablement inspiré « Dojo », le superordinateur de Tesla en cours de construction.

Akida Nsoc. Ce processeur Neural SoC a été développé par la société « Brainchip » Il comprend 1,2 million de neurones et 10 milliards de synapses. Le circuit est en technologie 28 nm ; il est fabriqué par TSMC.

Si les ordinateurs classiques sont beaucoup plus puissants que le cerveau humain pour un grand nombre de tâches comme le calcul, ils lui restent très inférieurs pour tout ce qui est cognitif ou interprétations subtiles (reconnaissance de formes, reconnaissance de par-

1 *Memory processors.*

fums), pourtant très utiles en intelligence artificielle. Pour apprendre à un réseau neuronal à reconnaître un objet, il faut lui présenter des milliers de photos du même objet. Pour le cerveau humain, quelques photos suffisent et souvent même une seule. C'est ce que l'on attend des réseaux neuromorphiques, dont les capacités d'apprentissages seront ainsi beaucoup plus aisées.

Un second avantage important est la consommation d'énergie. Les puces neuromorphiques seront les meilleurs supports pour réaliser des réseaux de neurones capables de calculer, mais avec une consommation d'énergie des milliers ou des millions de fois inférieure à celle des meilleures puces d'IA numériques actuelles. On peut donc parler d'une véritable « informatique verte ».

Cependant, la conception de ces puces est beaucoup plus complexe. Plus sujettes aux erreurs, elles nécessitent plus d'itérations et font appel à des outils de développement non standards.

Les puces neuromorphiques seront utilisées pour l'IA dans des applications telles que le calcul scientifique, l'analyse graphique et différents systèmes de recommandation. Elles ont également un grand potentiel dans les domaines de la robotique et de la conduite autonome. Et sans doute, à terme, on trouvera les puces neuromorphiques dans toute l'électronique grand public.

Des puces neuromorphiques dopées à la photonique ?

La photonique est le fruit de la combinaison de deux inventions, parmi les plus importantes du ^{xx}e siècle : le circuit intégré en silicium et le laser semi-conducteur. Elle permet des transferts de données très rapides sur longue distance comparée à l'électronique traditionnelle, tout en bénéficiant des évolutions du silicium. En effet, la photonique est déjà très largement utilisée pour communiquer sur des câbles en fibres optiques. Les fibres constituent désormais l'épine dorsale de tous les centres de données connectés à Internet et la photonique est le moyen de communications interurbaines ou internationales dominant.

La photonique appliquée à l'électronique, connue également sous le nom « d'informatique optique », accélère le transfert de données et réduit la consommation d'énergie. En effet, la vitesse à laquelle un courant électrique circule entre composants microélectroniques est une contrainte potentielle pour l'avenir, car les ordinateurs actuels dépensent beaucoup de temps en transfert de données. L'architecture Von Neumann constitue, comme on l'a déjà vu, une contrainte importante en matière de performance.

En janvier 2021 des chercheurs d'IBM ont présenté une puce photonique neuromorphique constituée d'un tenseur expérimental avec stockage de données photoniques et traitement des données colocalisés.

Plusieurs articles de recherche récents, notamment en février 2021 dans « *Nature* », ont également détaillé les avantages potentiels de l'utilisation de la photonique pour l'informatique neuromorphique. Ils ont montré la possibilité de créer des circuits massivement parallèles et distribués. En permettant un niveau de bande passante et de traitement élevé, les futurs systèmes photoniques offriront une latence bien inférieure, nécessiteront moins d'énergie et généreront moins de chaleur. L'efficacité énergétique des processeurs photoniques peut surpasser considérablement l'électronique classique qui reste tributaire des interconnexions.

En résumé l'objectif principal de la photonique neuromorphique est d'émuler un réseau de neurones efficacement et rapidement grâce à une bande passante élevée, une faible latence, une mémoire répartie qui réduit le nombre de transferts de données ; le tout à la vitesse de l'optique (Figure 15).

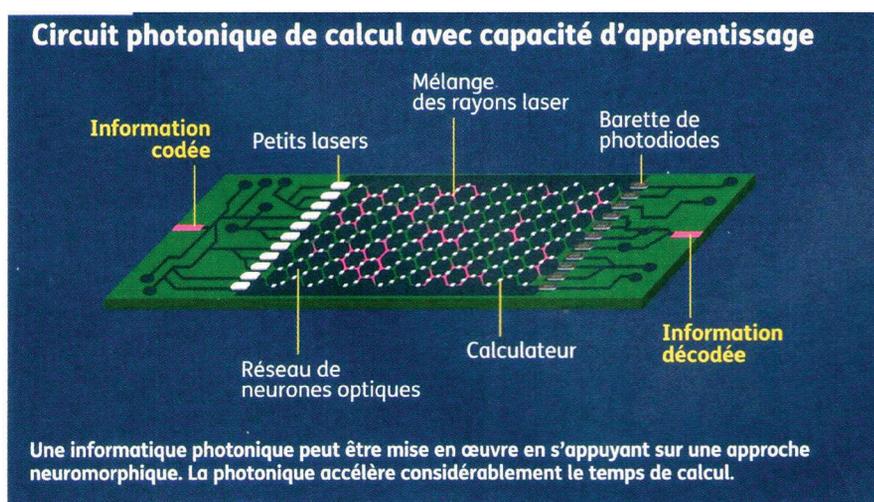


Figure 15 : principe d'un circuit neuromorphique photonique

(Source : *Industrie & Technologies* - février 2022)

Autre avantage de la photonique, ces puces seraient capables d'exécuter plusieurs opérations en même temps sur un même cœur. En effet, l'équipe d'IBM a montré qu'un processeur photonique intégré peut exécuter des opérations parallèles sur un même noyau physique en utilisant le multiplexage en longueur d'onde ou WDM (*Wavelength Division Multiplexing*). Une technologie, déjà très utilisée, qui multiplexe un certain nombre de signaux porteurs optiques sur une seule fibre optique en utilisant différentes longueurs d'onde du laser. De cette façon, on pourrait exécuter des calculs en parallèle sur un même cœur en utilisant différentes longueurs d'onde (ou couleurs de lumière) simultanément.

Même si les procédés de fabrication actuels sont potentiellement capables de fabriquer de telles puces, il est sans doute plus raisonnable de prévoir que ce sera plutôt dans les années 2030 que nous aurons des systèmes d'apprentissage automatique photoniques-neuromorphiques fondamentalement plus puissants, plus réactifs et plus efficaces énergiquement comparés aux applications actuelles de l'IA.

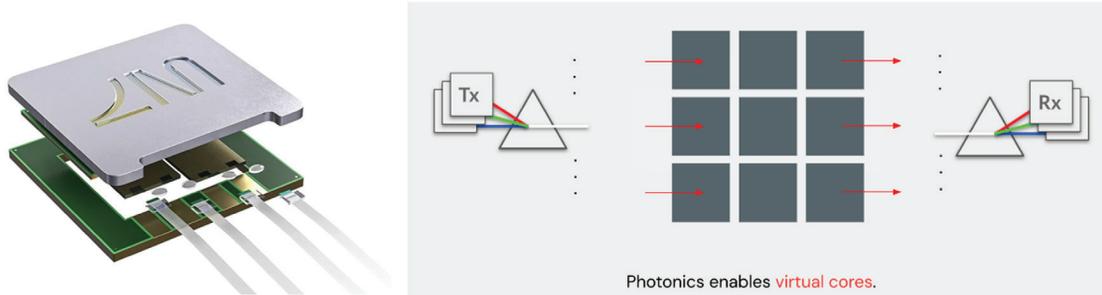


Figure 16 : Lightmatter a développé "Envisage". Ce processeur photonique, dédié à l'apprentissage profond, démontre le concept de "cœur virtuel". Ainsi chaque cœur physique du processeur peut lui-même exécuter plusieurs tâches en parallèle grâce au multiplexage WDM.

(source Lightmatter)

Il faut aussi noter que tandis que les interactions des photons sont directement analogues aux interactions électrochimiques des neurones dans le cerveau humain, la grande différence est que les neurones artificiels photoniques fonctionnent des millions de fois plus rapidement que leurs homologues biologiques. Si nous devenons capables de construire des processeurs photoniques neuromorphiques avec des dizaines de milliards de neurones, alors nous serons sans doute en mesure de dépasser les performances humaines dans le domaine cognitif.

Pendant des décennies, les humains ont construit, des ordinateurs à partir de composants microélectroniques avec une architecture Von Neumann, maintenant nous sommes susceptibles de nous affranchir de ces contraintes et de construire une IA de toute nouvelle génération.

L'ingénierie neuromorphique, tout comme l'informatique quantique, donne la possibilité de faire un pas gigantesque en avant dans la façon de déduire des informations à partir de données.

Cela ne veut pas dire que la photonique et les ordinateurs neuromorphiques remplaceront l'électronique des ordinateurs plus traditionnels, mais nous nous dirigeons vers un monde où nous utiliserons différents types de matériel pour différents types d'application.

LA PLACE DE L'EUROPE DANS L'INDUSTRIE DES SEMI-CONDUCTEURS

Si donc l'Europe n'a jamais été un grand leader dans le domaine des composants, elle a quand même réussi à créer des acteurs de classe mondiale. En voici quelques exemples :

- ST Microélectronique, Infineon, NXP et Bosch pour les puces ;
- ASML : leader mondial pour les matériels de lithographie, essentiels à la fabrication des composants ;
- SOITEC : produit les wafers silicium les plus performants ;
- ARM développe des designs de processeurs repris par tous les fabricants du monde entier.

Mais comme on l'a vu les énormes besoins en investissements pour accéder aux technologies de dernière génération ont concentré très fortement cette industrie, autour de quelques leaders mondiaux dont principalement le Taiwanais TSMC et le Coréen Samsung.

Ainsi aujourd'hui, vouloir accéder à une totale indépendance technologique de l'Europe paraît un objectif ambitieux et difficile à atteindre à court terme pour plusieurs raisons :

- l'accès aux technologies les plus miniaturisées pour certaines applications ne peut se faire qu'en Asie, auprès des deux acteurs majeurs que sont TSMC et Samsung, qui sont particulièrement en avance ;
- la conception des circuits intégrés repose aujourd'hui sur des fournisseurs américains d'outils logiciels clefs ; de façon complémentaire, la production reste également tributaire d'un grand nombre d'équipementiers américains ;
- concernant le « *front-end* », les circuits et les fonderies, l'Europe ne représente que 10 % de l'activité mondiale. C'est ce chiffre qui a alerté, parmi sans doute bien d'autres considérations, la Commission européenne pour lancer un plan et viser pour l'Europe une part de 20 % ;
- par ailleurs, la part européenne dans les activités de packaging (dont plus la moitié sera bientôt du SiP) n'est que de 3 % ! Il faut s'en alarmer autant que pour le reste et se fixer un objectif : par exemple, 10 % du marché. De plus, l'Europe ayant une industrie en pointe dans le B to B, les SiP apportent la flexibilité nécessaire à une grande variété d'applications et donc de spécifications variées, mêlant analogique et numérique ;
- la sécurisation d'une filière d'expertise technique doit également passer par un accroissement de l'appétence de nos jeunes pour les sciences, les techniques, l'industrie, et ceci, dès le collège. Ainsi, une promotion accrue pour les formations scientifiques afin d'en augmenter l'attractivité est vitale pour assurer la pérennité de l'innovation, et plus particulièrement pour l'industrie des semi-conducteurs.

Paradoxalement, malgré toutes ces difficultés, il renaît en Europe quelques espoirs, car cette industrie arrive maintenant à un nouveau stade de son évolution. Les investissements nécessaires pour générer des gains de performance sont devenus si colossaux qu'ils ont réduit à un duopole le nombre d'acteurs industriels encore en mesure de les financer. Mais, par ricochet, le ticket d'entrée demandé à leurs clients pour accéder à ces nouvelles générations de procédés est devenu tellement élevé qu'une rupture est apparue. La course à l'intégration monolithique des circuits intégrés se ralentit et l'assemblage de technologies hétérogènes devient une alternative compétitive.

Effectivement, les SOC et les SIP, développés au paragraphe « *More than Moore* » (page 12) constituent cette véritable rupture technologique, en offrant :

- de bien meilleures performances pour une application donnée ;
- une plus grande efficacité énergétique ;
- ainsi qu'un coût optimisé.

C'est à l'occasion de telles ruptures que des opportunités apparaissent et permettent à de nouveaux acteurs de venir sur un marché que l'on croyait définitivement verrouillé. Ces alternatives ouvrent une fenêtre d'opportunité stratégique pour l'industrie des semi-conducteurs française et européenne, car les vecteurs de croissance tels que l'industrie 4.0 et l'automobile, dans lesquels l'Europe excelle traditionnellement, ne bénéficient encore que marginalement de la loi de Moore.

Mais avant de développer les opportunités qui s'offrent ainsi à l'Europe, regardons comment réagissent les États-Unis face aux « menaces » du « monopole » asiatique.

LA STRATÉGIE AMÉRICAINE

Les Américains qui, depuis l'origine, ont été les leaders du secteur, ont dominé ce marché pendant des décennies, ont perdu pied progressivement au cours des dernières années face à la montée des entreprises asiatiques, particulièrement Samsung et TSMC. Aujourd'hui, seulement 12 % des semi-conducteurs dans le monde sont fabriqués aux États-Unis contre 37 % en 1990 et même 100% à l'époque des Bell Labs et des débuts de la Silicon Valley.

Consciente du danger, la première action de l'administration Trump a été d'inciter (voire de contraindre) le duo Samsung/TSMC à transférer une partie de ses fabrications (celles qui concernent les clients américains) sur le sol national grâce à l'implantation d'usines en contrepartie d'aides gouvernementales. Cette action a porté ses fruits. Samsung, déjà

présent à Austin vient tout récemment d'annoncer la création d'une seconde usine toujours dans le Texas pour un montant de 17 Md\$ et TSMC construit actuellement une usine de fabrication de puces (5 nm) en plein milieu du désert de l'Arizona. Cette usine d'une superficie totale de 215.000 m² représente un investissement de 12 milliards de dollars.

CHIPS AMERICAN ACT

Mais la pénurie de puces, qui a affecté très négativement la production automobile, aurait entraîné une perte de 210 milliards de dollars pour l'ensemble de l'industrie américaine en 2021. Sans véritable visibilité sur la fin de cette crise, l'industrie automobile et les entreprises technologiques ont aussi appelé à une production accrue de puces basées aux États-Unis.

En réponse, en avril 2021, l'administration Biden a annoncé un « *Chips American Act* », destinée à aider l'industrie des semi-conducteurs. Il est intégré dans le cadre du « *Biden's infrastructure package* ».

Le Sénat américain vient de donner son approbation pour doter ce « *Chips for America Act* » de 52 milliards de dollars. Il vise à relancer l'industrie américaine des semi-conducteurs au cours de la prochaine décennie en encourageant de nouveaux investissements dans la fabrication et particulièrement :

- la fabrication de puces de 3 nm et en dessous ;
- le packaging microélectronique avancé (SIP) ;
- les matériels et machines de fabrication des semi-conducteurs ;
- la chaîne d'approvisionnement.

Ce « *Chips American Act* » est un soutien à tout un écosystème technologique complexe. L'administration américaine estime pour sa part que cet investissement de l'état devrait permettre la construction de 19 nouvelles usines aux États-Unis au cours des dix prochaines années.

C'est dans ce cadre qu'Intel a annoncé la construction de deux usines dans l'Ohio pour un montant de 20 milliards \$. L'une est dédiée à la technologie 3 nm et devrait être opérationnelle en 2025. La seconde supporte la nouvelle stratégie « Intel IDM 2.0 » qui est de concurrencer TSMC et Samsung sur l'activité de fonderie.

Parallèlement, les grandes sociétés américaines ont compris tout le parti qu'elles pouvaient maintenant tirer du « *More than Moore* », c'est-à-dire des SOC et des SIP et donc de la conception de puces dédiés à des applications spécifiques. Jusqu'ici les entreprises achetaient des composants « sur étagères » pour développer leurs produits. Par exemple,

chez Intel ou AMD, qui conçoivent et fabriquent des processeurs qui sont ensuite intégrés par les fabricants de serveurs, de PC, de tablettes, de mobiles et aussi, plus récemment, d'électroménager ou de voitures.

Mais ces dernières années, les grandes sociétés de la Tech (Amazon, Apple, Google, Tesla...) plutôt que d'utiliser des composants *standards*, préfèrent concevoir elles-mêmes leurs puces, beaucoup plus adaptées et optimisées pour leurs besoins spécifiques, leur assurant ainsi une plus grande efficacité et compétitivité. Le design des composants une fois réalisé, ces entreprises en sous-traitent la fabrication au « fondeur » de leur choix.

LE « SILICIUM PERSONNALISÉ »

Le « *Custom silicium* » (le « silicium personnalisé »), prend ainsi rapidement de l'importance dans l'ensemble des secteurs industriels aux États-Unis, voici quelques exemples de sociétés ayant conçu des SOC ou des SIP personnalisés :

- Apple qui « motorisait » ses Mac à partir de processeurs Intel a conçu son propre processeur « M1 » qui maintenant remplace dans tous ses appareils les composants d'Intel. Le processeur M1 Apple comporte un très haut niveau d'intégration (un système sur une puce : SOC) réunissant : un processeur dix cœurs, un processeur graphique trente-deux cœurs, un processeur neuronal, une architecture de mémoire unifiée (tous les processeurs accèdent aux mêmes données), ainsi que : processeurs d'image, de sécurité, contrôleur de stockage, moteurs d'encodage/décodage vidéo, contrôleur thunderbolt, etc. Ces puces M1 sont fabriquées par TSMC en technologie 5 nm ;
- Tesla qui utilisait initialement pour son AutoPilot, son application d'assistance à la conduite, un processeur produit par nVIDIA, a préféré le remplacer par un SOC de sa conception : « FSD » (*Full Self Driving*) qui est désormais intégré dans tous ses véhicules depuis 2019. Il intègre 8 cœurs ARM, un processeur graphique, deux processeurs neuronaux et de la mémoire partagée. La puce FSD est 20 fois plus puissante, 4 fois moins énergivore et 20 % moins chère que la puce précédente de nVIDIA. Une nouvelle version (FSD2), 4 fois plus puissante, est annoncée pour courant 2022. Les deux versions sont fabriquées par Samsung, FSD en technologie 14 nm, FSD2 en 7 nm.

Les architectures SOC & SIP permettent également de développer des processeurs massivement parallèles, pouvant ainsi permettre la construction de superordinateurs :

- Tesla a annoncé en août dernier la conception et la construction en cours, par ses équipes, d'un superordinateur baptisé « Dojo », dont la puissance sera de 1,1 exaflops et qui sera dédié à l'apprentissage de la conduite des voitures et à la simulation. Il est basé sur la puce D1, conçue par Tesla, d'une surface de 645 mm². D1 est un SOC, constitué de 354 nœuds identiques fonctionnant à 2GHz. Chaque nœud est composé

de processeurs de « *machine learning* », de processeurs de calcul matriciel, de mémoire et d'I/O permettant l'interconnexion de 1 000 000 nœuds. Ce SOC est fabriqué par TSMC en technologie 7 nm.



Figure 17 : Tuile « Dojo ». 120 tuiles de ce type constitueront le superordinateur « Dojo » construit par Tesla pour apprendre à conduire à ses voitures.

La **Figure 17** montre la « tuile dojo » qui intègre sur un wafer, 25 puces D1 ainsi que les équipements permettant l'interconnexion entre toutes ces puces. Pour l'intégration, Tesla s'est appuyée sur la technologie 3D de TSMC appelée « InFO (Integrated Fan-Out) Wafer-Level Packaging » qui consiste à « coller » sur un même Wafer, comportant déjà un certain nombre de composants et de circuits, les 25 puces D1 et tous les équipements permettant de communiquer et fonctionner. Par exemple les puces sont entourées de connecteurs d'entrées/sorties à large bande assurant un débit impressionnant de 36 Tb/s avec toutes les autres tuiles. Comme on l'a vu au chapitre des puces neuromorphique, Dojo semble inspiré de Spinnaker et des travaux de l'université de Manchester.

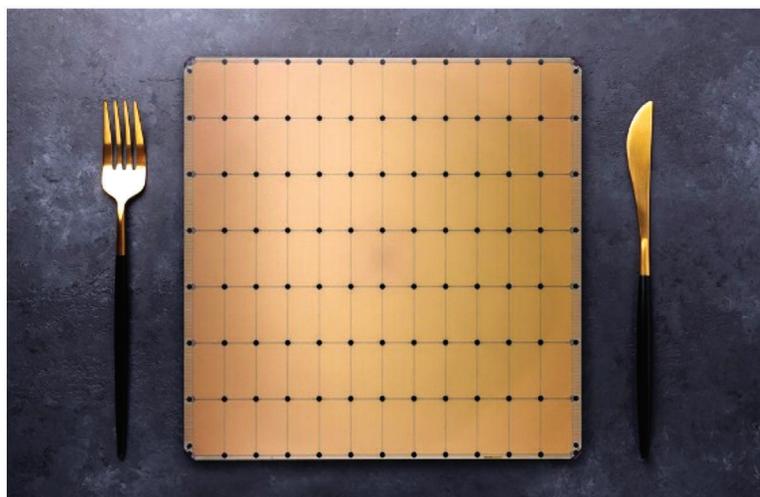


Figure 18 : La puce de dernière génération pour l'IA de Cerebras est la plus grande du monde.

- Cerebras System est une société californienne qui conçoit des puces pour accélérer les calculs liés à l'intelligence artificielle. Cette puce de très grande taille (Figure 18), réalisée selon une architecture « *Wafer scale* » utilise une tranche entière de silicium. Elle contient 2 600 milliards de transistors, 850 000 cœurs. D'une surface de 42.225 mm², c'est la plus grande puce au monde. Elle est fabriquée par TSMC en 7 nm. Optimisée pour les calculs intensifs d'IA, elle est capable de rivaliser avec certains supercalculateurs.

Quels que soient les besoins, il y a donc actuellement une forte tendance qui incite les entreprises utilisatrices de puces, comme ici Apple ou Tesla, à s'emparer de leur conception pour mieux répondre à leurs besoins spécifiques afin d'optimiser les performances, d'améliorer l'efficacité énergétique et d'abaisser les coûts.

On peut penser que cette tendance pour le « silicium personnalisé », qui concerne principalement des sociétés technologiques, va s'amplifier et s'étendre à bien d'autres secteurs d'activités.

On voit donc que le monde des composants se scinde en deux grands domaines, du fait des besoins de spécialisation et de personnalisation :

- d'un côté les concepteurs/utilisateurs, qui sont en train d'acquérir et de rapatrier chez eux le savoir-faire en matière de conception et qui ne souhaitent plus dépendre de tiers pour cette phase à très forte valeur ajoutée ;
- les « fondeurs » dont le métier est beaucoup plus axé sur les process de fabrication à la demande de séries petites ou moyennes, à partir des technologies les plus avancées.

Dans ce nouveau contexte, les rapports de force et de dépendance ne sont plus tout à fait les mêmes !

Mais malgré l'engouement des entreprises les plus avancées pour le silicium personnalisé, il faut reconnaître que le succès et l'attrait des offres américaines « sur étagères » demeurent pour les sociétés qui maîtrisent, encore mal, le numérique.

C'est notamment le cas de la solution « *Snapdragon digital chassis* » (Figure 19) proposée et présentée par Qualcomm au CES'22 et qui a recueillie l'adhésion de nombreux constructeurs automobiles mondiaux (et pas des moindres) comme BMW, GM, Hyundai, Renault, Volvo et Honda.

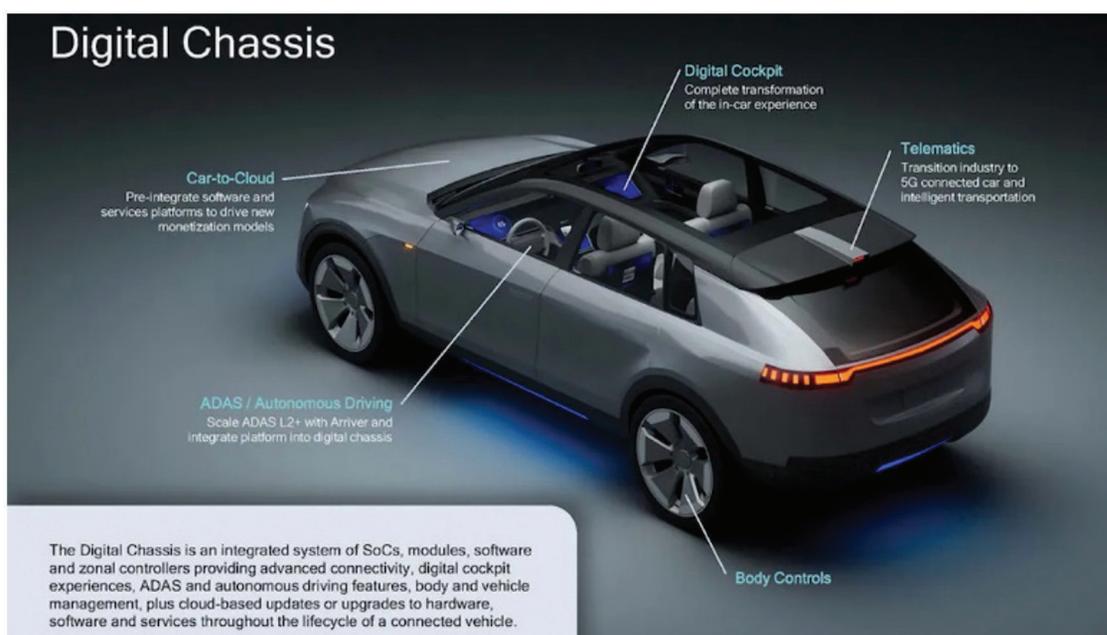


Figure 19 : “Snapdragon digital chassis” et ses quatre modules.

Leader mondial des composants pour la téléphonie mobile, Qualcomm (Figure 19) propose maintenant aux constructeurs automobiles, confrontés au développement rapide des véhicules électriques et de l'autonomie, une solution « clefs en main » en une plateforme numérique complète composée de quatre modules :

- *Snapdragon Ride* : technologies d'aides à la conduite et à la conduite automatisée ;
- *Snapdragon Cockpit* : ensemble de services multimédias à bord du véhicule ;
- *Snapdragon Auto Connectivity* : liaisons LTE 5G, Wi-Fi et GPS et connexions des voitures à internet, au cloud et à d'autres véhicules ;
- *Snapdragon Car-to-Cloud Services* : fonctions de cybersécurité, plateformes de mises à jour en direct et de services payants ainsi que du suivi des véhicules et de leur conduite, afin de permettre aux constructeurs automobiles de monétiser les services à bord des véhicules.

Comme pour les mobiles, Qualcomm conçoit les composants qui sont ensuite fabriqués par TSMC.

Renault espère implanter cette nouvelle architecture sur les nouveaux véhicules dès 2026. Puis, à terme, le constructeur fera migrer l'ensemble de ses gammes et proposera également cette solution à ses partenaires.

VERS PLUS DE SOUVERAINETÉ EUROPÉENNE ?

Comme on l'a vu précédemment, les conditions de marché et les besoins qui ont mené à la super-intégration ont continué d'évoluer. Elles ne sont plus seulement orientées vers une recherche de performance, mais génèrent également une forte demande vers plus d'agilité, de flexibilité, de spécialisation applicative, comme illustré en Figure 20.

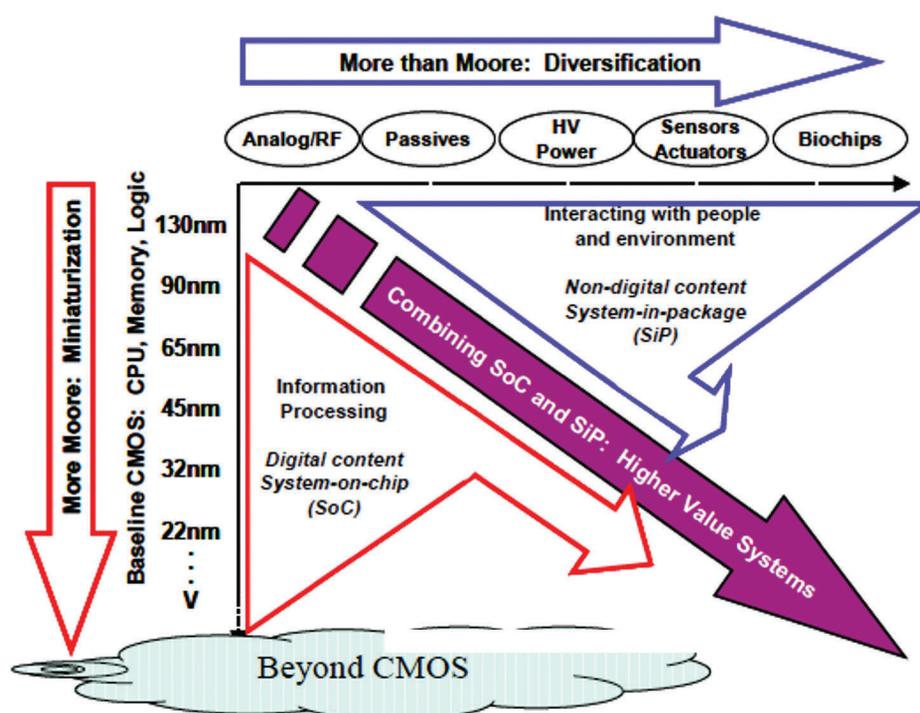


Figure 20 : Illustration d'une offre différenciée bâtie sur la diversification et complémentarité des technologies de semi-conducteurs et d'assemblage.

Outre la compréhension fine des enjeux de l'industrie des semi-conducteurs, l'Europe est également à la fois un marché et un acteur mondial majeur pour les applications qui bâtissent le futur : infrastructure 5G, véhicules hybrides/électriques et autonomes, industrie 4.0, Internet des objets...

Tout ce qui précède a montré qu'il existe maintenant des alternatives à la loi de Moore et que l'Europe et la France sont bien placées pour les exploiter et que la maîtrise des éléments différenciants nécessaires à nos industries et à nos marchés en est la pierre angulaire.

Dans ce contexte, il est également crucial de pouvoir garantir la protection de nos données, de nos informations. En synergie avec la maîtrise des compétences nationales en matière de sécurité informatique et de cryptographie, et en utilisant les solutions développées par

les acteurs européens de l'industrie des semi-conducteurs (tels que STMicroelectronics avec ses microcontrôleurs), une approche encore plus ambitieuse pourrait être proposée.

Des architectures innovantes pourraient en effet bénéficier des avancées en technologies 3D et permettre de sécuriser efficacement des circuits complexes embarquant notamment des technologies extraeuropéennes.

La position proposée pour l'Europe et pour la France est donc de continuer d'innover et d'investir pour développer une plus grande indépendance et attractivité, par la différenciation technologique européenne, et mettre en place une stratégie de complémentarité permettant de garder le contrôle de nos applications lorsque des technologies externes doivent être utilisées.

L'Europe peut, *a minima*, inciter les industriels des composants à produire plus localement en s'appuyant sur les industries qui sont les plus grandes consommatrices de composants comme, par exemple, l'industrie automobile européenne. Mais son potentiel est-il suffisant à lui seul pour inciter les fondeurs à créer des usines de composants en Europe et quels seraient les autres secteurs industriels européens, suffisamment significatifs en volume, qui pourraient venir conforter cette démarche ?

Néanmoins, cela semble le cas puisque INTEL & TSMC ont manifesté, ces derniers temps et à plusieurs reprises, leur souhait d'implanter des usines en Europe.

EUROPEAN CHIPS ACT

Mi-septembre 2021, faisant le constat de la situation de l'Europe dans le domaine des composants, Thierry Breton, commissaire européen, a annoncé un « *European chips Act* » visant à remettre l'Europe dans la course à la technologie.

« *Alors que la demande mondiale a explosé, la part de l'Europe dans l'ensemble de la chaîne de valeur, de la conception à la capacité de fabrication, a diminué. Nous dépendons de puces de pointe fabriquées en Asie. Il ne s'agit donc pas seulement d'une question de compétitivité. C'est aussi une question de souveraineté technologique. Alors, concentrons-nous là-dessus.* »

L'*European Chips Act* devrait permettre d'intégrer les efforts dans ce domaine de chaque pays membre dans une vision et une stratégie européenne cohérente.

Début février 2022, le commissaire européen Thierry Breton a confirmé ce plan doté de 50 milliards € destiné à couvrir trois domaines :

- une stratégie européenne de recherche sur les semi-conducteurs s'appuyant sur les acteurs en place : IMEC, CEA-LETI ainsi que le Fraunhofer Institute ;
- un plan pour renforcer la production européenne (multiplier par quatre la production en 2030) avec tous ses aspects, y compris le packaging, ainsi que la production à grand volume de semi-conducteurs les plus avancés (2 nm et moins) ;
- un cadre de coopération et de partenariats internationaux, car tout ne peut pas être produit en Europe par les seuls Européens et donc aider Samsung, Intel et TSMC à s'implanter en Europe.

Par ailleurs la Commission européenne, pour dynamiser le secteur, avait déjà lancé le programme *European Processor Initiative* (EPI) qui vise à élaborer une famille de processeurs à faible consommation pour les besoins allant des supercalculateurs les plus puissants aux applications émergentes de l'Informatique embarquée. Ce programme réunit 27 acteurs de la recherche, des semi-conducteurs, et de l'informatique, mais aussi des industriels comme BMW.

Au-delà des entreprises du Numérique, l'industrie européenne doit à son tour, pour diminuer sa dépendance et accroître sa part de valeur ajoutée, acquérir et maîtriser les compétences dans la conception de puces et d'architectures de machines, afin d'utiliser voire mettre en œuvre par elle-même des systèmes les plus adaptés à leurs besoins spécifiques, sous peine de perdre rapidement en compétitivité face aux concurrents qui ont fait cette démarche.

Il faut donc sensibiliser les industriels européens, les motiver et les inciter à constituer leurs équipes pour développer leurs propres circuits, accroître leur compétitivité pour entrer dans la problématique de la conception de puce et d'architecture de machine en organisant les compétences, la formation et la création d'écosystèmes plus favorables.

Quelques industriels européens ont déjà entrepris cette démarche :

- parmi les participants au projet EPI, citons STmicroelectronics avec des processeurs neuromorphiques pour l'Industrie automobile, Infineon pour l'Industrie automobile allemande, Atos pour les supercalculateurs, Kalray qui depuis plusieurs années conçoit une puce dédiée aux applications nécessitant un calcul massivement parallèle (encodage vidéo, conduite autonome), mais aussi BMW ;
- de son côté Volkswagen (Audi, Porsche), à l'image de Tesla, a annoncé en mai dernier vouloir se lancer dans la conception de puces pour déployer plus facilement la conduite autonome pour prendre ainsi de l'avance sur ses concurrents et se différencier. Volk-

swagen semble ainsi vouloir devenir une société plus technologique et plus agile.

LA FORTE DÉPENDANCE AU NUMÉRIQUE, DE L'INDUSTRIE EUROPÉENNE. EXEMPLE : L'AUTOMOBILE

Au-delà du projet EPI, il est nécessaire de créer un écosystème européen des composants (compétences, formation) pour aider les industriels à progresser dans ce domaine et des filières d'expertise technique en s'appuyant sur les grands acteurs européens du secteur.

Il est particulièrement intéressant de faire un « zoom » sur l'industrie automobile européenne qui est confrontée, aujourd'hui, à une violente mutation, le passage au véhicule électrique, qui peut jouer un rôle déterminant et peser fortement sur l'industrie européenne des semi-conducteurs.

De « plieurs et soudeurs de tôle » les constructeurs européens sont contraints de devenir rapidement des grands du numérique, s'ils ne veulent pas perdre de fortes parts du marché mondial et s'ils ne veulent pas voir déferler en Europe d'autres constructeurs non seulement américains, mais aussi et surtout chinois.

Une voiture électrique contiendra beaucoup plus de lignes de code qu'un avion et le nombre important de composants déjà présents dans les véhicules sera décuplé dans les années qui viennent.

Au moins trois constats en témoignent déjà :

- en 2021, on n'avait jamais vu autant d'arrêts de lignes de production de voitures pour cause de manque de composants au détriment de la production annuelle, alors que d'autres constructeurs moins dépendants ont doublé leur production pendant la même période !
- jamais les constructeurs automobiles européens n'ont été aussi présents au CES de Las Vegas qu'en janvier 2022, à la recherche de partenariats avec les acteurs du numérique ;
- une voiture contient beaucoup plus de composants mixtes analogique-digital, de capteurs et de microcontrôleurs que de processeurs ultra-avancés.

On le voit donc, ces deux grandes industries européennes, composants et automobile, ont tout intérêt à collaborer et à saisir ensemble les opportunités technologiques du « *more than Moore* » qui s'offrent à elles. Pour les semi-conducteurs, c'est ainsi renforcer leur position, et pour les constructeurs automobiles européens c'est retrouver ou maintenir leur leadership. Pour tous, les deux années à venir seront décisives.

CONCLUSION POUR L'EUROPE

Durant les cinquante dernières années, les progrès spectaculaires réalisés par l'industrie des semi-conducteurs ont contribué à la croissance économique mondiale. La loi de Moore a structurellement permis plus de capacité de calcul pour un coût de production maîtrisé de la technologie. Sans la dynamique qu'elle illustre, il n'y aurait pas eu d'ordinateurs personnels, de smartphones, ni d'Internet. Concomitamment, les technologies à base de semi-conducteurs comme le silicium se sont améliorées pour approcher l'échelle de l'atome et faire progresser les méthodes de fabrication.

Cependant, comme nous avons pu le voir, cette industrie arrive maintenant à un nouveau stade de son évolution. Les investissements nécessaires pour générer des gains de performance supplémentaires sont devenus colossaux et ont réduit à un duopole le nombre d'acteurs industriels encore en mesure de les financer. Par ricochet, le ticket d'entrée demandé à leurs clients pour accéder à ces nouvelles générations de procédés est également devenu tellement élevé qu'une rupture est apparue :

la course à l'intégration monolithique des circuits intégrés et à la réduction des dimensions des transistors se ralentit. L'assemblage de technologies hétérogènes redevient une alternative compétitive, en permettant l'association avec d'autres composants amenant des fonctionnalités complémentaires comme mémoire embarquée, analogique, RF, capteurs. Aucun de ces composants ne nécessite de technologie à résolution ultime, ils s'appuient sur des générations antérieures allant du 40 au 28 nm notamment.

Ce changement ouvre une fenêtre d'opportunité stratégique pour l'industrie des semi-conducteurs française et européenne, et peut-être encore plus pour les industriels qui les utilisent, car les vecteurs de croissance tels que l'industrie 4.0 et l'automobile, dans lesquels l'Europe excelle traditionnellement, ne bénéficient encore que marginalement de la loi de Moore.

- Donc même si le « packaging » n'était pas jusqu'ici son cœur de métier, l'industrie européenne des semi-conducteurs, peut trouver là une véritable opportunité pour rester ou revenir dans la course en proposant des solutions innovantes et pragmatiques à ses clients.
- Par ailleurs, la pénurie de composants (*chips shortage*) devrait maintenant inciter les industriels européens, fortement dépendants des composants (et destinés à le devenir de plus en plus !) à réfléchir à la nécessité absolue de réduire fortement cette dépendance vis-à-vis de leurs fournisseurs. Comment ne pas s'inspirer de Tesla qui, pendant cette période, a été capable de trouver de nouvelles sources de composants

et, pour les accueillir, a pu modifier les circuits imprimés et le firmware de ses voitures en seulement quelques jours. Depuis le début de la crise, Tesla a également développé près d'une vingtaine de microcontrôleurs. En 2021, alors que ses concurrents avaient tous, périodiquement, des chaînes à l'arrêt, Tesla a doublé sa production annuelle.

- Plus généralement, toutes ces industries, « composants dépendantes », si elles veulent bénéficier pleinement, non seulement de la loi de Moore, mais aussi des possibilités d'optimisation qu'offre maintenant « le silicium personnalisé », doivent acquérir progressivement les compétences, pour mieux analyser et définir leurs besoins et devenir elles-mêmes capables de spécifier (voire même de concevoir) les composants qui leur sont nécessaires, plutôt que de simplement puiser dans des catalogues de fournisseurs.
- Pour cela, les organisations professionnelles et les États européens doivent se mobiliser et s'organiser pour créer les écosystèmes indispensables tant du point de vue des compétences, des équipements, que de la formation et de l'assistance, pour permettre aux entreprises qui le souhaitent (et inciter les autres !) à se mettre à niveau.
- Enfin n'oublions pas la recherche dans le domaine des puces neuromorphiques qui pourraient nous faire faire un saut « quantique », à très court terme, dans le domaine de l'intelligence artificielle et de l'apprentissage profond.

Maintenant que les processeurs sont partout et que la domination de la planète numérique par quelques processeurs d'origine américaine est en voie de réduction, pour satisfaire les nouveaux besoins, profitons des opportunités offertes par les ruptures technologiques (qui sont à la portée de l'Europe) pour se repositionner, renforcer notre industrie des composants, mieux maîtriser leur conception et leurs usages, reprendre des parts de marchés et diminuer notre dépendance.

Pour renforcer et développer ses industries, l'Europe ne peut passer à côté de l'opportunité historique de la révolution en cours des composants et particulièrement celle du « silicium personnalisé ».

LA VISION D'UNE ENTREPRISE EUROPÉENNE : STMICROELECTRONICS

Dans le cas de STMicroelectronics, le développement technologique cible préférentiellement un portefeuille de technologies complémentaires, hétérogènes et performantes, en bénéficiant des compétences en technologie d'assemblage et d'encapsulation. Cela permet de couvrir les marchés de l'automobile électrifiée et numérique, de l'Internet des Objets (IoT), des réseaux téléphoniques cellulaires de 5e génération (5G), l'industrie intelligente et connectée (I 4.0), les capteurs améliorant l'expérience utilisateur, l'intelligence artificielle embarquée (Edge AI)...

Pour illustrer ce type de situation, STMicroelectronics est devenue l'un des acteurs majeurs du marché des microcontrôleurs. Les microcontrôleurs sont des systèmes sur silicium qui intègrent des fonctions performantes de traitement numérique (plusieurs centaines de MHz pour un cœur 32-bit), des mémoires non volatiles de grande capacité (plusieurs mégaoctets) et des fonctions analogiques haute tension pour servir les applications multiples qui interagissent avec notre monde analogique (actuateurs). On les retrouve dans une multitude d'objets du quotidien tels que les robots ménagers, les télécommandes diverses, les terminaux bancaires... Il peut y en avoir des dizaines dans chaque voiture pour le contrôle des freins, de l'airbag, des lève-vitres, tableau de bord ou encore des essuie-glaces. Grâce à son modèle économique de fabricant de produits intégrés (Integrated Device Manufacturer — IDM) et à sa compréhension de l'application, STMicroelectronics fournit des solutions innovantes optimisant la consommation pour les produits opérant sur batterie, la performance radiofréquence pour les objets connectés, le stockage sécurisé pour les applications sensibles ou encore la fiabilité pour les applications automobiles, et bien sûr, tout cela pour un coût compatible avec un marché de fort volume. En ce sens, STMicroelectronics maîtrise la définition, la conception et la production interne des produits, et se différencie des fonderies, dont l'activité est centrée sur la production en volume de produits qu'elles n'ont pas conçus.

Une autre illustration de cette approche est le choix fait par STMicroelectronics dans l'adoption et le développement d'une technologie optimisée en performance, coût et consommation, appelée FDSOI. Cette innovation provient de l'écosystème grenoblois regroupant STMicroelectronics, le CEA-Leti et SOITEC. Elle repose sur l'introduction d'un substrat silicium contenant une fine couche d'isolant sous une fine couche de silicium. En améliorant le contrôle électrostatique des transistors et en réduisant les fuites de courant dans le substrat, cela permet de repousser les limites des technologies planaires et de surseoir à la nécessité de basculer vers un changement coûteux d'architecture transistor (passage au FinFET), comme illustré dans la Figure 21.

Cette figure met en évidence les différentes évolutions des architectures des transistors matérialisées dans l'espace puissance/vitesse, et dont l'amélioration du compromis est visualisée par des ruptures technologiques. Chacune des grandes bifurcations correspond à des innovations contournant les points de blocage techniques rencontrés, avec toutefois des options stratégiques différentes. Ainsi, comme évoquée précédemment, l'architecture FDSOI privilégie la modération de la consommation énergétique correspondant bien aux besoins des produits plébiscités par les clients.

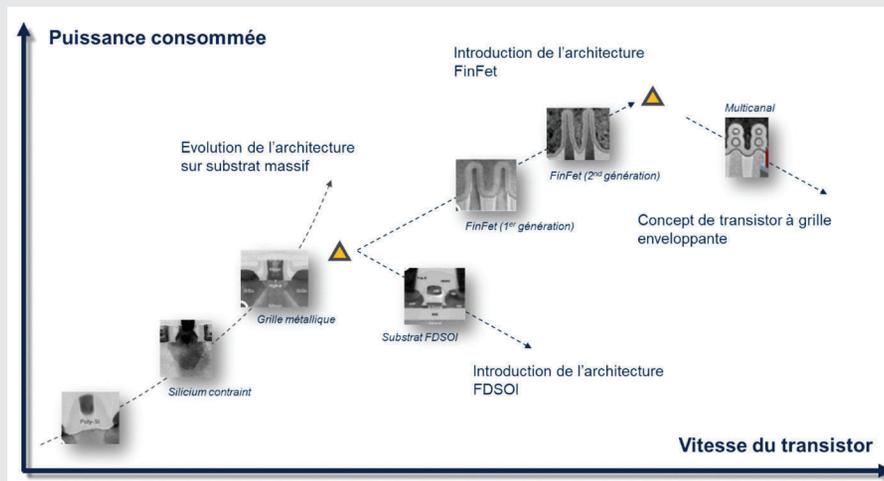


Figure 21: Illustration de l'évolution de l'architecture des transistors soutenant la feuille de route des technologies de l'industrie des semi-conducteurs. Source : ST (2020)

Notre avenir repose, d'une part, sur notre capacité à préserver notre position dominante en proposant des solutions intégrées monolithiques adaptées aux besoins de nos clients grâce à une force d'innovation continue, et d'autre part, sur notre aptitude à compléter et maîtriser ces nœuds technologiques avancés hybrides pour créer des produits différenciés ; c'est ainsi que nous resterons un acteur économique de premier plan sur la scène mondiale. Innover davantage et développer un outil de conception et de production selon ces deux axes stratégiques est la position forte choisie par STMicroelectronics.

UN NOUVEAU SOUFFLE POUR L'ARCHIVAGE DES BIG DATA

*François Képès et groupe de travail ADN :
Lire, écrire et stocker l'information*

Résumé

Les centres de données, « cloud » inclus, stockent les big data de l'humanité sur disques durs et bandes magnétiques dont la durée de vie limitée oblige à de dispendieuses recopies tous les cinq à sept ans ; ils représentent des gouffres pour les ressources en terrain, électricité, eau et matériaux rares. Ces centres ne parviennent plus à traiter le déluge des données. Malgré leur augmentation en nombre et en efficacité, le rapport de l'offre sur la demande en stockage va continuer de chuter dans le futur. En comparaison, l'archivage à l'échelle moléculaire sur un polymère tel que l'ADN (en dehors du vivant), permettrait une densité supérieure d'un facteur dix millions, une conservation prolongée d'un facteur dix mille sans recopie périodique, pour une consommation électrique divisée par environ mille. En effet, l'ADN est stable à température ordinaire durant plusieurs millénaires ; il peut, de plus, être aisément dupliqué ou volontairement détruit. Les technologies requises existent, et la faisabilité de cet archivage moléculaire est démontrée. Cependant, pour devenir économiquement viables, ces technologies nécessitent encore des progrès qui pourraient voir le jour sous cinq à vingt ans. Leur panoplie pourrait s'enrichir d'autres polymères plus performants que l'ADN avec un décalage temporel de quelques années. Le handicap principal de l'archivage moléculaire réside dans la lenteur des procédés d'écriture et d'édition, il est raisonnable de supposer que son usage se cantonnera d'ici 2030 à des marchés de niche comme l'archivage à long terme de données patrimoniales.

MOTIVATION

L'humanité accumule des données à un rythme jamais vu et qui va croissant. Les données considérées ici sont celles de nos connexions familiales, amicales et professionnelles, nos livres, vidéos, photos, achats, nos données médicales, celles de la recherche scientifique, de l'industrie, etc. On parle parfois de « big data » ou « mégadonnées »¹. Et bien plus est à venir : voitures autonomes, capteurs et autres objets connectés, télésurveillance, réalité virtuelle, déserts médicaux compensés par la téléconsultation, le diagnostic et même la chirurgie à distance. En 2025, il est estimé que les trois quarts d'entre nous seront connectés et que nous interagirons chacun avec des données toutes les 18 secondes en moyenne².

LA SPHÈRE GLOBALE DES DONNÉES

L'ensemble des données numériques créées par l'humanité, la « sphère globale des données » (SGD), contient environ autant d'octets que le nombre estimé de grains de sable sur la terre. Cette SGD était estimée en 2018 à 33 zettaoctets (Zo ; soit 33×10^{21} octets)³.

1 L'Académie des technologies s'est penchée à plusieurs reprises sur les big data, et deux contributions récentes en ont résulté, portant l'une sur les aspects technologico-stratégiques (*Big data : un changement de paradigme peut en cacher un autre*, EDP Sciences 2015), l'autre sur les aspects éthiques (*Big data - Questions éthiques*, Académie des technologies 2019).

2 HiPEAC Vision 2015 (Commission européenne, FP7, 2015).

3 Préfixes du Système International (SI) d'unités, et nombres correspondants :

K	kilo	10^3	1 000
M	mega	10^6	1 000 000
G	giga	10^9	1 000 000 000
T	tera	10^{12}	1 000 000 000 000
P	peta	10^{15}	1 000 000 000 000 000
E	exa	10^{18}	1 000 000 000 000 000 000
Z	zetta	10^{21}	1 000 000 000 000 000 000 000
Y	yotta	10^{24}	1 000 000 000 000 000 000 000 000

Elle double tous les 2 à 3 ans, et atteindra environ 175 Zo en 2025⁴. Par exemple, chaque minute dans le monde, environ 400 heures de vidéo (200 Go) sont ajoutées à la SGD. Autre exemple, pris dans le monde de la recherche, l'Organisation européenne pour la recherche nucléaire (CERN) a produit plus de 100 Po de données qui doivent être conservées pour les générations suivantes de physiciens. Au rythme actuel, la SGD dépassera 5 000 Zo en 2040. Une autre façon d'appréhender un nombre aussi élevé consiste à dire qu'il faudrait 50 millions d'années pour télécharger cette SGD avec une connexion internet de vitesse moyenne.

Une part majoritaire des données créées par l'humanité est ensuite stockée au long terme. Outre le stockage local (sur un ordinateur ou téléphone), dont la croissance ralentit actuellement, le stockage déporté est en rapide augmentation, en particulier dans les centres de données y compris dans le « cloud » ou « nuage ». Dans la suite de ce rapport, nous considérerons l'ensemble des centres de données, « nuage » inclus. Des hangars dédiés à ce stockage centralisé ne cessent de se développer dans le monde, souvent dans les pays froids, car ce stockage est grand consommateur d'électricité et de refroidissement.

Paradoxalement, les capacités disponibles de stockage en centres de données ne représentent en 2020 que 40 % des données numériques créées⁵. Ce pourcentage est en chute libre. Autrement dit, le modèle actuel de conservation des big data ne remplit pas son rôle, et le remplira de moins en moins.



Photographie aérienne du centre de données Google de Pryor creek Oklahoma

Source : Wikipédia - Xpda /Bob Webster

- 4 Reinsel D, Gantz J, Rydning J (2018). *The Digitization of the World - From Edge to Core* (International Data Corporation & SeaGate).
- 5 International Data Corporation digital universe study — <https://www.idc.com/>

GENTRE DE DONNÉES

Pour être plus concret, prenons le cas d'un gros centre de données d'un Eo. Il représente plusieurs milliards d'euros en investissement, plusieurs centaines de milliers de m², un million de serveurs. Il consomme un gigawatt (GW) d'électricité (dont 40-50 % environ pour le refroidissement), soit environ 10 TWh par an, soit plus qu'une ville française de 100 000 habitants. Bien entendu, ces centres sont reliés au reste du monde par d'importants réseaux de connexions, également consommateurs de diverses ressources dont l'électricité⁶. Au total, en incluant ceux des entreprises, il y avait 8,6 millions de centres de données dans le monde en 2017, occupant une surface totale de plus de 170 millions de m², soit l'équivalent de 25 000 terrains de football⁷. Cette surface représente environ un millionième des terres émergées de la planète puisque ces dernières couvrent approximativement 150 millions de km². Si le rythme actuel d'un doublement tous les deux ans se poursuivait, un millième des terres émergées serait occupé par ces centres avant 2040. Cependant, cette valeur est probablement surestimée, car l'efficacité énergétique des centres de données augmente continuellement⁸ et leur empreinte au sol diminue. Les centres de données et leurs réseaux de connexions consomment 2 à 4% de l'électricité dans les pays avancés. Ensemble, ils équivaldraient aujourd'hui au cinquième pays le plus consommateur d'électricité au monde, entre Inde et Japon. L'investissement mondial annuel pour construire de nouveaux centres de données est de l'ordre de plusieurs dizaines de milliards US\$⁹. Par exemple Google aurait à lui seul investi dix milliards US\$ par an sur la période 2015-2017.

PISTES

Pour le stockage et l'archivage des données, les centres reposent sur des solutions très sophistiquées au plan technique. Pourtant, sachant qu'en 2040 nous aurions cent à mille fois plus de données à archiver, le modèle actuel de conservation ne répondra pas aux besoins et deviendrait insupportable au plan environnemental.

6 Davey J (2019). *Powering the data revolution* (HSBC Global Research).

7 Reinsel D, Gantz J, Rydning J (2018). *The Digitization of the World - From Edge to Core* (International Data Corporation & SeaGate).

<http://hebergement-et-infrastructure.fr/actualites-et-innovations/8-6-millions-de-datacenters-dans-le-monde-en-2017>
<https://www.businesswire.com/news/home/2014110005018/en/IDC-Finds-Growth->

8 <https://www.datacenterknowledge.com/archives/2016/06/27/heres-how-much-energy-all-us-data-centers-consume>

9 Cook G (2012). *How clean is your cloud?* Greenpeace International : <https://www.greenpeace.org/archive-international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf>

Pour obtenir un volume d'archivage raisonnable et une consommation énergétique faible, on fait actuellement appel à la bande magnétique. Le volume des données de 1 E0, considéré précédemment, correspond à 50 000 cassettes standards dont la capacité approche aujourd'hui 20 To et continue à progresser. Le volume correspondant est de 10 m³ seulement. Leur temps d'accès de ces cassettes est important, mais adéquat pour l'archivage. Cependant, la mémorisation sur bande magnétique n'est pas pérenne et demande une lourde opération de rafraîchissement périodique (tous les cinq à sept ans). La continuité d'un tel rafraîchissement semble délicate à assurer à travers des siècles. Il y a donc un défi à relever pour trouver dans le futur un procédé de conservation permanente de la SGD, la bande magnétique assurant seulement une fonction de transition.

L'Académie des technologies s'est penchée sur cette question durant deux années (2018-2020). Son groupe de travail transdisciplinaire « ADN : lire, écrire, stocker l'information » a auditionné vingt-six spécialistes mondiaux venant de 10 pays. Il a publié son rapport en 2020¹⁰. Ce groupe de travail s'est focalisé sur les supports moléculaires porteurs d'information, tel que l'ADN utilisé ici comme agent chimique en dehors du vivant ou d'autres polymères¹¹ non-ADN très prometteurs¹².

En effet, de nouvelles méthodes de stockage d'information à l'échelle moléculaire sont envisagées afin d'augmenter la capacité de stockage et la durée de conservation des données, et de réduire la taille des supports. Les progrès récents des technologies d'écriture et de lecture de l'ADN ont amené les chercheurs à envisager ce polymère naturel comme support d'archives numériques. L'ADN permet optimalement de stocker 1 bit pour environ 50 atomes, alors que le stockage magnétique requiert environ un million d'atomes, support mécanique non compris.

BREF HISTORIQUE DU STOCKAGE DE DONNÉES SUR L'ADN

Richard Feynman en 1959 et Mikhail Neiman en 1964 ont été les premiers à envisager l'ADN comme support de stockage de l'information numérique. Mais c'est en 1977 qu'a été

10 Académie des technologies (2020). *Rapport Archiver les big data au-delà de 2040 : la piste de l'ADN* — <https://www.academie-technologies.fr/blog/categories/publications-de-l-academie/posts/archiver-les-megadonnees-au-dela-de-2040-la-piste-de-l-adn>

11 Un polymère est une grande molécule constituée de nombreuses sous-unités répétées, appelées monomères. Pour stocker de l'information, il faut un polymère constitué d'au moins deux sortes de monomères, représentant le '0' et le '1'. L'ADN par exemple a quatre monomères différents.

12 Chaput JC, Herdewijn P, Hollenstein M. Orthogonal Genetic Systems (2020). *Chembiochem* 21(10):1408-1411.
Hoshika H et al. (2019). Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 363:884-887.
Colquhoun H & Lutz JF (2014). Information-containing macromolecules. *Nature Chemistry* 6:455-456.

mise au point la première méthode de lecture de l'ADN et en 1983 une technique d'écriture de l'ADN. En 1988 et pour la première fois, Joe Davis¹³ a conçu et synthétisé un fragment d'ADN de 18 nucléotides contenant un message numérisé, qu'il a ensuite transféré chez une bactérie intestinale, le colibacille.

En 2012, l'équipe de George M. Church (Université de Harvard, États-Unis d'Amérique) a stocké 0,6 Mo d'information sur l'ADN, sous forme de fragments synthétiques¹⁴. En 2013, l'équipe de Nick Goldman (Institut européen de bioinformatique, Royaume-Uni) a converti quatre fichiers en séquence d'ADN, pour un total de 0,7 Mo¹⁵. L'information a été retranscrite sans erreurs.

- En 2018, Microsoft Corp. et l'université de Washington, aux États-Unis d'Amérique, ont stocké sur l'ADN 1 Go d'information venant de fichiers de types variés¹⁶. Ils détiennent depuis le record ayant fait l'objet d'une publication validée.
- En 2024, il est projeté d'archiver 1 To (équivalent à environ 1 000 films) en vingt-quatre heures pour un coût de 1 000 US\$¹⁷.

L'ADN

Pour rappel, l'acide désoxyribonucléique (ADN) est formé de deux brins antiparallèles enroulés l'un autour de l'autre pour former une structure en double hélice. Chaque brin d'ADN est un polymère linéaire (non branché) composé d'un assemblage de nucléotides. Chaque nucléotide est composé d'une des quatre bases azotées, adénine (A), guanine (G), thymine (T), cytosine (C), liée à un sucre désoxyribose, lui-même lié à un groupe phosphate. Les bases nucléiques d'un brin d'ADN peuvent interagir avec les bases nucléiques de l'autre brin d'ADN à travers des liaisons hydrogène en respectant des règles d'appariement. Ainsi, A et T s'apparient avec deux liaisons hydrogène, tandis que G et C s'apparient avec trois liaisons hydrogène.

Dans le monde vivant, à l'exception de quelques virus, l'ADN est le support de l'information héréditaire. L'ADN peut, depuis 1869 (Friedrich Miescher), être manipulé *in vitro*,

////////////////////////////////////

13 [https://en.wikipedia.org/wiki/Joe_Davis_\(artist\)](https://en.wikipedia.org/wiki/Joe_Davis_(artist))

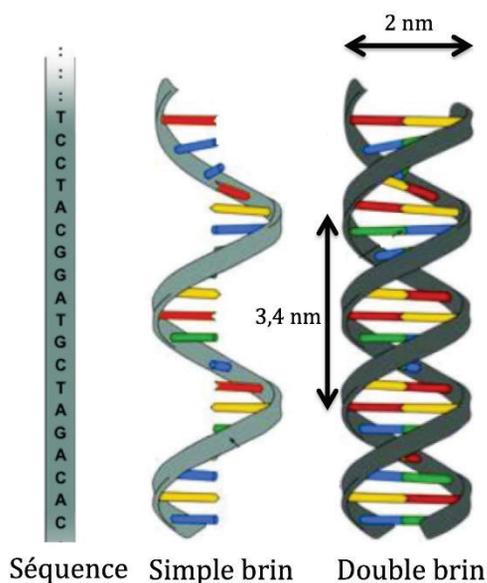
14 Church GM, Gao Y, Kosuri S (2012). Next-Generation Digital Information Storage in DNA. *Science* 337:1628.

15 Goldman N, Bertone P, Chen S, Dessimoz C, Leproust EM (2013). Toward practical, high-capacity, low maintenance information storage in synthesized DNA. *Nature* 494:77-90

16 Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA (2019). *Nat Rev Genet* 20: 456–466.

17 Intelligence Advanced Research Projects Activity [IARPA] (2020). <https://www.dni.gov/index.php/newsroom/press-releases/item/2086-iarpa-announces-launch-of-the-molecular-information-storage-program>

en dehors des cellules. C'est principalement *in vitro* qu'il a été envisagé de l'utiliser pour stocker des données numériques. Comparé aux systèmes traditionnels, il présente à ce titre de nombreux avantages.



Représentations de l'ADN. À droite, hélice d'ADN double brin (montrant les appariements AT et CG) ; au milieu, ADN en simple brin ; à gauche, séquence du même simple brin, maintenant déroulé pour une représentation linéaire.

Crédit : adapté par François Képès.

AVANTAGES DU STOCKAGE D'INFORMATION SUR L'ADN

La densité informationnelle

La densité informationnelle de l'ADN est environ 10 millions de fois supérieure à celle des meilleurs systèmes traditionnels. L'ADN peut en principe stocker un demi Zo d'information par gramme (g). Ainsi, les chercheurs estiment que la SGD de l'humanité entière tiendrait actuellement dans moins de 100 g d'ADN.

- Cependant, en pratique, ce n'est pas une seule molécule d'ADN qui est synthétisée pour coder un fichier, mais de nombreux exemplaires identiques (entre 10 et 10^{15}).
- En outre, des zones de cet ADN devront porter des signaux de contrôle qualité et d'indexation, en sus des données.
- Enfin, l'ADN doit être préservé dans des conteneurs macroscopiques¹⁸.

18 Organick L et al. (2018). Random access in large-scale DNA data storage. *Nature Biotechnology* 36(3):242-248.

Tenant compte de ces pertes de densité, on peut estimer que la SGD de 2018 stockée sur ADN tiendrait plus réalistement dans une fourgonnette.

La consommation

Le stockage de l'ADN à température ordinaire n'implique aucune consommation de ressources, et les opérations sur l'ADN sont beaucoup moins énergivores qu'en électronique ; un gain potentiel d'un facteur mille a été évoqué¹⁹.

La longévité

La longévité de l'ADN est environ dix mille fois supérieure à celle des supports traditionnels. Des molécules d'ADN vieilles de plus de 560 000 ans ont été analysées à partir d'échantillons historiques²⁰. En laboratoire, une demi-vie de 52 000 ans a été démontrée en accélérant artificiellement son vieillissement²¹.

L'obsolescence

L'obsolescence du support ADN ne se produira pas tant que l'homme disposera des technologies nécessaires à l'écriture et à la lecture de molécules d'ADN, qui font partie intégrante de la médecine moderne. Reste la question du codage/décodage, qui serait résolue si un standard émergeait.

La copie ou multiplication

Le phénomène naturel de réplication est reproduit *in vitro* par la « réaction en chaîne de la polymérase » (PCR). Ainsi, un seul fragment d'ADN peut être dupliqué en chaîne par des thermocycleurs de paillasse, engendrant par ce processus exponentiel plusieurs milliards de copies en quelques heures pour une fraction d'euro.

La destruction volontaire

La destruction totale volontaire de l'ADN est réalisable aisément et rapidement, soit de manière sophistiquée avec des enzymes (des catalyseurs protéiques) disponibles dans le commerce (ADNases), soit par des traitements physiques comme la haute température.

19 <https://www.iarpa.gov/index.php/research-programs/mist>

20 Orlando L et al. (2006). Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Current Biology* 16, R400-402.

Orlando L et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74-78.

21 Bonnet J et al. (2010). Chain and conformation stability of solid-state DNA: implications for room temperature storage. *Nucleic Acids Res.* 38(5):1531-46. <http://www.imagine.fr/dnashell-rnashell/dnashell/>

Calculs

Les propriétés physico-chimiques de l'ADN se prêtent à y implémenter directement certains calculs. Le principe d'un tel calcul est de coder un problème combinatoire avec des brins d'ADN que l'on fait synthétiser sur mesure, de manipuler ces brins par les outils de la biologie moléculaire pour simuler les opérations qui isolent la solution, puis de lire cette dernière par séquençage²².

PROCESSUS DU STOCKAGE D'INFORMATION SUR ADN

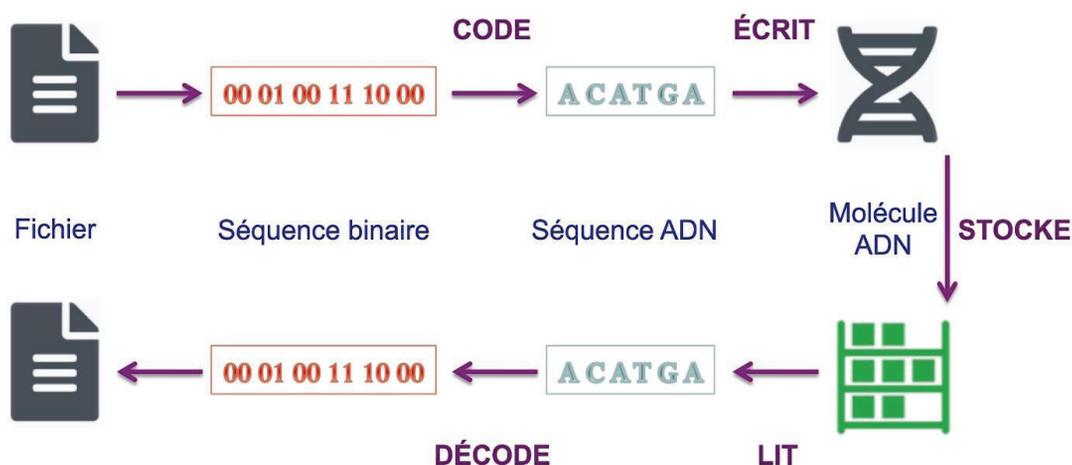
Le processus enchaîne cinq étapes : codage des données, écriture sur l'ADN, stockage de l'ADN, lecture de l'ADN et décodage. Peuvent s'y greffer : modification, amplification ou destruction de l'ADN.

Codage

Pour stocker des données dans l'ADN, le concept est similaire à l'approche informatique, mais le processus est différent. Plutôt que de créer des séquences de 0 et de 1, comme pour les données numériques, le stockage de données sur l'ADN utilise des séquences de nucléotides. L'idée générale est d'attribuer des valeurs numériques aux nucléotides d'ADN. Par exemple, le couple de bits '00' pourrait être équivalent au nucléotide 'A', '01' à 'C', '10' à 'G' et '11' à 'T'. Ainsi, un nouveau code est inventé, où les bits sont convertis en nucléotides pour former un fragment d'ADN, ensuite synthétisé *in vitro*. Cependant, des méthodes de codage plus élaborées commencent à apparaître.

La technologie actuelle de synthèse d'ADN est limitée à des fragments de l'ordre de 200 nucléotides au maximum, donc très courts au regard des fichiers informatiques. On peut donc dresser une analogie entre les « paquets de nucléotides » (les courts fragments d'ADN) et les paquets d'octets qui sont expédiés lors d'une transaction internet, par exemple l'envoi d'un courriel. Dans les deux cas, le message complet sera correctement reconstitué à l'arrivée à partir des paquets grâce à leurs signaux d'indexation, adressage et contrôle qualité. Le fichier informatique est découpé en fragments d'une vingtaine d'octets. Ces fragments sont convertis par un ordinateur en une suite de nucléotides d'ADN. Chaque segment d'ADN possède environ 200 nucléotides et contient la charge utile (*payload*) et l'étiquette. Cette dernière permet de regrouper les fragments d'ADN contenant l'information issue des mêmes fichiers numériques. Elle est utilisée pour l'accès sélectif à l'information, selon un principe d'indexation. Une fois le processus de codage d'information numérique sur l'ADN effectué, les fragments d'ADN sont synthétisés.

22 Adelman, LM (1994). Molecular computation of solutions to combinatorial problems, *Science* 266, 5187, 1021-1024.



Étapes du processus de stockage des big data numériques sur l'ADN. Ici sont représentés pour exemple 12 bits successifs extraits du fichier numérique. Ces 12 bits sont codés sous la forme de 6 nucléotides qui sont écrits en succession dans une molécule d'ADN. Cet ADN est ensuite stocké. Puis il est lu, et la séquence de nucléotides ainsi obtenue est décodée pour reconstituer le fichier numérique d'origine.

Crédit : François Képès.

Écriture de l'ADN

La synthèse d'ADN permet d'écrire des fragments d'acides nucléiques relativement courts, possédant un enchaînement défini de nucléotides. Il existe deux voies de synthèse, chimique ou enzymatique. En cas de modification apportée au fichier stocké, la séquence de l'ADN synthétisé peut ensuite être modifiée par mutagenèse.

Depuis 1983, il existe une méthode de synthèse de l'ADN par voie chimique. C'est aujourd'hui la seule méthode commerciale d'écriture de l'ADN. Son principe basé sur la chimie des phosphoramidites a très peu évolué depuis. Lors de la synthèse d'« un » fragment d'ADN, ce n'est en réalité pas une unique molécule qui est synthétisée, mais une population de copies identiques : typiquement 10^{12} à 10^{15} copies, qu'il est possible de réduire en conditions contrôlées à une dizaine de copies afin d'augmenter la densité informationnelle²³. L'état de l'art, en particulier chez la société Twist Bioscience (États-Unis)²⁴, permet l'écriture d'ADN dans des puits gravés dans une micropuce de silicium, synthétisant simultanément 10 000 fragments d'ADN différents comprenant jusqu'à 200 nucléotides. Elle atteindra rapidement le million de puits. Le taux d'erreur pour chaque nucléotide ajouté est d'environ 0,1 %, ce qui explique qu'en pratique la longueur du fragment utilisable est limitée à 200 nucléotides. Pour écrire des morceaux d'ADN bien plus longs, la méthode usuelle est d'assembler bout

23 Organick L et al. (2020). *Nat Commun.* 11(1):616.

24 <https://www.twistbioscience.com/technology>

à bout de nombreux fragments d'environ 200 nucléotides, par exemple en jouant sur le chevauchement de leurs extrémités prévues pour être complémentaires deux à deux afin d'éviter toute erreur d'assemblage. Un autre inconvénient de la synthèse par voie chimique est que la chimie des phosphoramidites est polluante.

En raison de ces limites, une méthode alternative de synthèse de l'ADN par voie enzymatique a été inventée au début des années 2010. Cette synthèse fait usage d'une ADN-polymérase spéciale qui permet d'allonger l'ADN avec le seul nucléotide désiré en fournissant uniquement celui-ci à une étape donnée. Cette approche a plusieurs avantages par rapport à la traditionnelle voie chimique : une vitesse de synthèse élevée, un taux d'erreur présentement similaire, mais avec un potentiel supérieur de diminution, une hydrosolubilité du nucléotide rendant l'approche enzymatique moins polluante que la voie chimique qui recourt à des solvants organiques. Une demi-douzaine de sociétés se sont lancées dans cette nouvelle approche, parmi lesquelles la société française DNA Script²⁵ est en pointe. Il est trop tôt pour estimer à quel horizon cette approche biologique deviendra commercialement attractive.

Disposer d'un accès évolutif aux données offre une flexibilité appréciable dans le monde de l'informatique, car cela permet la ré-écriture d'une partie d'un fichier numérique sans entraîner la coûteuse nécessité de tout ré-écrire. De même serait-il intéressant de ré-écrire certaines parties de l'ADN sans toucher aux autres zones. De nombreuses méthodes de ré-écriture ont été développées depuis les débuts du génie génétique en 1973²⁶. Sur ce point cependant, l'approche moléculaire est lente et complexe, au regard des méthodes informatiques.

Stockage au long terme de l'ADN

Affranchir du froid la conservation de l'ADN représente un avantage technologique, économique et écologique considérable pour le stockage moléculaire. L'ADN synthétisé peut ensuite être stocké par différentes méthodes, dont deux sont particulièrement adéquates au vu de l'objectif : l'une physique, l'autre chimique.

Dans le système de stockage chimique, développé par Robert Grass²⁷ (ETH Zürich, Suisse), l'ADN synthétisé est encapsulé dans des nanobilles de silice, ensuite référencées et réparties dans des plaques à micropuits. Avant d'être lu, l'ADN stocké dans les nanobilles doit être extrait par un réactif chimique capable de dissoudre les silices tout en

25 <http://www.dnascript.com/>

26 Esvelt KM, Wang HH. (2013). Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol.* 9:641.

27 Chen WD et al. (2019). Combining Data Longevity with High Storage Capacity—Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. *Advanced Functional Materials* 1901672.

préservant l'ADN. La durée de conservation de l'ADN à température ambiante est estimée à quelques décennies.

L'approche physique utilise des capsules conçues par la société Imagene²⁸ (France). L'extérieur de ces capsules est en acier inoxydable, l'intérieur en verre, et elles ont la taille d'une pile bouton. Chaque capsule contient jusqu'à 0,8 g d'ADN (potentiellement 1,4 Eo de données en tenant compte de la redondance). Elle est à usage unique. Son ouverture permet de récupérer l'ADN. La durée de conservation de l'ADN à température ambiante est estimée à quelques dizaines de millénaires dans ces systèmes de stockage qui protègent l'ADN de ses trois ennemis que sont l'eau, l'oxygène et la lumière.

Lecture de l'ADN

Pour retranscrire l'information stockée sur l'ADN, il faut la séquencer. Le séquençage de l'ADN consiste à déterminer l'enchaînement des nucléotides A, T, G et C au sein de ce polymère linéaire. Il fait appel à la PCR ou à des nanopores.

En effet, la PCR permet non seulement l'amplification avec accès sélectif (indexation), mais aussi le séquençage. L'ADN est fragmenté, l'ordre des nucléotides est déterminé dans chaque fragment, puis les fragments sont ordonnés par alignement de leurs séquences qui se recouvrent partiellement. Du fait du grand nombre de copies disponibles dans l'échantillon, la lecture de l'ADN n'est pas destructrice.

Les séquenceurs de troisième génération développés par Oxford Nanopore Technologies²⁹ font passer l'ADN ou autre polymère par des nanopores en enregistrant au passage sa composition en nucléotides ou autres monomères, qui sont différenciés par leurs traces électriques. Ils permettent de lire de longues séquences d'ADN d'une traite (record à 2,2 millions de nucléotides) afin d'éviter les étapes de fragmentation et d'alignement des séquences et d'analyser les données en temps réel. Malgré un taux d'erreur sur chaque brin d'ADN supérieur à 10 %, le séquençage en parallèle de nombreuses copies de l'ADN à travers différents pores permet une correction quasi parfaite de la séquence par recoupement.

Décodage de l'information

Les séquences des fragments d'ADN, issues du même fichier numérique, sont regroupées informatiquement par leurs étiquettes puis par leurs parties communes. Les séquences d'ADN sont retranscrites en chaînes d'octets. Au sein d'un même fichier numérique, les chaînes d'octets sont ordonnées, grâce à leurs identifiants, afin de reconstituer le fichier numérique complet.

28 <http://www.imagene.fr/>

29 <https://nanoporetech.com/>

Les opérations de codage et décodage sont donc d'ordre purement calculatoire et peuvent être réalisées massivement avec de faibles moyens informatiques. Les opérations d'écriture et lecture de l'ADN relèvent de la chimie ou biochimie et représentent les étapes limitantes, comme il sera discuté dans les perspectives.

Chimies alternatives

En principe, tout polymère comportant au moins deux monomères différents pourrait être utilisé pour stocker l'information numérique. En pratique, il faut que ce polymère puisse être écrit selon une séquence arbitraire, donc par chimie itérative en phase solide. Il faut aussi qu'il existe des méthodes pour le conserver longtemps et le lire aisément. Le projet académique porté par Jean-François Lutz (Institut Charles Sadron, université de Strasbourg, France) utilise des co-polymères non-ADN pour stocker de l'information numérique³⁰. Ces polymères de synthèse ont un potentiel considérable pour le stockage moléculaire. Ils permettent une plus grande densité d'information par l'élargissement de l'alphabet, et une meilleure conservation des données que les supports de stockage électroniques actuels. Pour lire l'information, les polymères sont « séquencés » par spectrométrie de masse, technique qui permet de détecter, identifier et caractériser les molécules d'intérêt. Le résultat, appelé spectre de masse, est décodé informatiquement pour reconstituer le message en bits, et donc l'information initiale.

Plusieurs approches moins radicales, conservant la structure générale de l'ADN, ont aussi été proposées. Elles modifient soit le sucre³¹, soit les bases azotées³², de l'ADN.

Enfin, des approches hétérodoxes ont aussi été proposées, par exemple basées sur la formation et la reconnaissance de structures secondaires de l'ADN de type tige-boucle (de deux longueurs nettement différentes représentant '0' et '1'), avec l'avantage d'un taux d'erreur abaissé, et l'inconvénient d'une perte de densité informationnelle³³.

30 - Colquhoun H & Lutz JF (2014). Information-containing macromolecules. *Nature Chemistry* 6:455-456.

- Al Ouahabi A, Charles L, Lutz JF (2015). Synthesis of Non-Natural Sequence-Encoded Polymers using Phosphoramidite Chemistry. *JACS* 137:5629-5635.

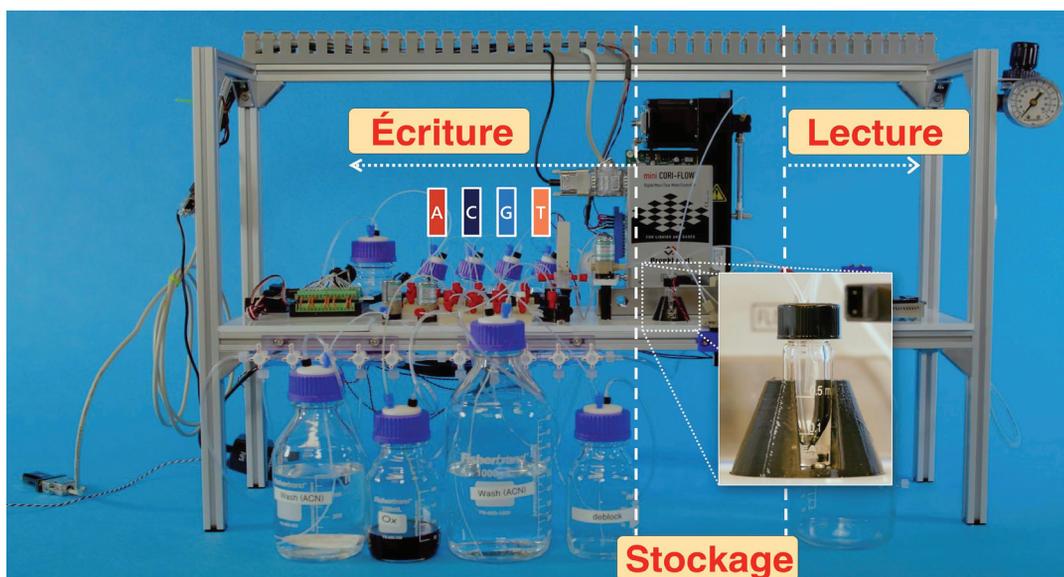
- Roy RK, Meszynska A, Laure C, Charles L, Verchin C, Lutz JF (2015). Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nat Comm* 6 :7237.

31 Chaput JC, Herdewijn P, Hollenstein M (2020). Orthogonal Genetic Systems. *Chembiochem* 21(10):1408-1411.

32 Hoshika H et al. (2019). Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 363:884-887.

33 Chen K, Kong J, Zhu J, Ermann N, Predki P, Keyser UF (2019). Digital Data Storage Using DNA Nanostructures and Solid-State Nanopores. *Nano Lett.* 19(2):1210-1215.

PERSPECTIVES



Premier prototype entièrement automatisé de stockage de données sur l'ADN

Crédit : Microsoft Corp. / Université de Washington, États-Unis.

Le projet le plus abouti à ce jour dans le domaine du stockage ou de l'archivage de données numériques sur l'ADN est coordonné par Karin Strauss (Microsoft Corp. et université de Washington, États-Unis). Son équipe a conçu en 2019 un prototype de paillasse utilisant l'ADN comme support pour stocker l'information³⁴. L'appareil est entièrement automatisé et autonome. Il est composé de trois parties : le synthétiseur, le système de stockage et le séquenceur. Le synthétiseur est capable de convertir l'information numérique en séquence d'ADN et d'écrire cette dernière. Le système de stockage encapsule l'ADN dans des nanobilles. L'ADN est ensuite extrait des nanobilles et séquencé via le dispositif d'Oxford Nanopore Technologies. Ce prototype, quoique lent, est fonctionnel et a déjà permis de stocker et retrouver 1 Go de données³⁵. Les chercheurs l'optimisent pour qu'il devienne plus compact et plus rapide. Ils développent notamment de nouveaux systèmes basés sur la microfluidique pour transporter des gouttes de réactifs sur un support électronique³⁶.

34 Takahashi CN, Nguyen BH, Strauss K, Ceze L (2019). Demonstration of End-to-End Automation of DNA Data Storage. *Scientific Reports* 9(1):4998.

35 Ceze L, Nivala J, Strauss K (2019). Molecular digital data using DNA. *Nat Rev Genet* 456:466.

36 Newman S et al. (2019). High density data storage library via dehybridization with digital microfluidic retrieval. *Nat Commun* 10(1):1706.

PERSPECTIVES TECHNO-SCIENTIFIQUES

Au-delà des approches théoriques, empiriques ou éducatives visant à limiter la quantité d'information engendrée par l'humanité, son archivage moléculaire constitue un enjeu majeur et stratégique à horizon proche.

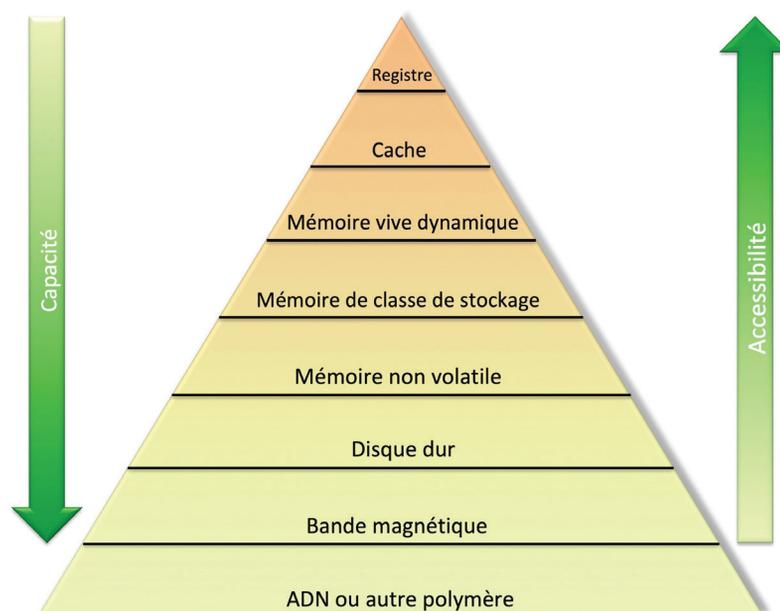
La faisabilité de l'archivage de données numériques sur l'ADN hors du vivant (*in vitro*) est établie. Plusieurs études ont montré que cet archivage peut prendre en charge l'accès sélectif et évolutif aux données, ainsi que le stockage et la restitution d'information sans erreur. Cependant, des défis techniques subsistent pour que ce procédé devienne viable économiquement pour un large spectre de données. Ils concernent l'amélioration des coûts, de la vitesse, du temps de latence et de l'efficacité des technologies d'écriture, édition et — dans une moindre mesure — lecture de l'ADN ou autres polymères.

Concernant l'écriture, plusieurs acteurs du domaine placent leurs espoirs dans la synthèse d'ADN par voie enzymatique, dont le potentiel de développement semble supérieur à celui de la voie chimique traditionnelle. Quant à la lecture, l'usage de nanopores offre un bon potentiel, car cette approche restitue de longues séquences d'un seul tenant, est intrinsèquement parallélisable, et montre une grande versatilité en s'adaptant à la croissante diversité chimique des hétéro-polymères ou co-polymères à applications numériques. Notons aussi que, quoique les vitesses d'écriture et de lecture de l'ADN soient limitantes, cet inconvénient est pallié dans certaines applications par la possibilité de parallélisation massive. Concrètement, d'ici 2024, une seule machine pourrait écrire et lire 1 To par jour.

Pourtant, plusieurs ordres de grandeur manquent actuellement pour la pleine adoption de la solution d'archivage des big data dans l'ADN : un facteur environ mille pour le coût de lecture et cent millions pour celui d'écriture. Ces facteurs peuvent sembler faramineux. Ce serait oublier la célérité des progrès des technologies associées à l'ADN. Ainsi, les coûts de lecture et d'écriture de l'ADN ont chuté en 10 ans d'un facteur supérieur au million, soit de moitié tous les six mois. Dans les domaines électronique et informatique, la société Seagate a signalé qu'elle avait fait diminuer en 29 ans le coût du stockage unitaire de données sur disque d'un facteur 1,3 millions. On voit donc que, selon ce critère, les technologies de l'ADN évoluent bien plus vite que celles de l'information et du stockage.

Les performances du procédé global d'archivage pourraient aussi être améliorées par des alphabets étendus ou composites. Enfin, plusieurs lignes de recherche sont parties du constat que l'ADN n'est pas nécessairement le polymère le plus performant : soit que son alphabet soit trop limité (à quatre lettres), soit que sa physico-chimie pose des problèmes pratiques. Ce constat a donné lieu à des approches alternatives s'éloignant plus ou moins de l'ADN, pour envisager d'autres hétéro-polymères ou co-polymères linéaires présentant

des avantages théoriques. Lorsque leurs performances de lecture, d'écriture et d'édition rejoindront celles de l'ADN, ce qui pourrait prendre 5-10 ans, ces polymères très prometteurs feront probablement irruption sur le marché de l'archivage de l'information numérique.



Pyramide des types de mémoires dans les systèmes informatiques actuels. Du sommet vers le bas, les mémoires augmentent en capacité et en temps d'accès. En bas de la pyramide a été ajouté à titre hypothétique l'usage de l'ADN ou d'un autre hétéropolymère.

Crédit : François Képès et Carlo Reita.

PERSPECTIVES ÉCONOMIQUES

À l'heure actuelle, le stockage d'information dans l'ADN reste expérimental. Un certain consensus s'est dessiné parmi quelques acteurs du domaine pour considérer que la viabilité économique du stockage moléculaire d'information pourrait être atteinte sous cinq à dix ans pour des marchés de niche s'apparentant au « stockage froid » (*cold storage*), autrement dit la conservation de données immuables consultées très rarement. Citons à titre d'exemple l'archivage à long terme du patrimoine culturel (par exemple films, livres, monuments, archives), scientifique (expériences de physique nucléaire) et d'informations sensibles (renseignement) ou imposées par la réglementation (banques). Ces applications valorisent certains atouts maîtres du stockage moléculaire : densité informationnelle, longévité, durabilité, économie en énergie et autres ressources, facilité à multiplier l'ADN et instantanéité de sa destruction volontaire. Elles ne souffrent que peu des actuels handicaps de cette solution puisque l'écriture se produit une fois et les modifications jamais ; la

lecture — qui est très peu coûteuse comparée à l'écriture — se produit occasionnellement dans des circonstances qui ne souffrent pas de l'important temps de latence imposé par cette approche.

Pour entrer en compétition avec les marchés plus globaux de l'archivage de big data, en incluant les coûts des machines requises, il faudra peut-être 10 à 20 ans. En ce cas, il est plausible que le stockage sur l'ADN entre en complémentarité puis en compétition avec la bande magnétique, actuellement la solution de choix pour l'archivage à long terme, au bas de la pyramide des types de mémoire (cf. figure).

PERSPECTIVES NATIONALES ET EUROPÉENNES

Des équipes travaillent sur ce sujet en Allemagne, Suisse, Irlande, Israël, dans un laboratoire européen (EBI) localisé au Royaume-Uni, et en France.

Le contexte français est d'une grande richesse en gros centres de données³⁷. Aujourd'hui, hors entreprises, existent en France 155 gros centres plus 22 dédiés au « nuage »³⁸.

En France, au moins un laboratoire universitaire et deux petites entreprises ont des positions originales et fortes dans les segments-clés très prometteurs que sont les polymères non-ADN (équipe de Jean-François Lutz, Strasbourg), la synthèse enzymatique de l'ADN (DNA Script, Paris) et son stockage de très longue durée (Imagene, Bordeaux/Évry). Citons aussi une très jeune société positionnée exactement sur ce créneau : Biomemory (Sorbonne Université, Paris)³⁹. Au-delà, existe en France un gisement de compétences pertinentes en biologie, chimie, informatique et sciences de l'ingénieur, qui pourraient être mobilisées dans une nécessaire synergie entre secteurs public et privé. Un exemple de gisement de compétences est le trio de laboratoires de Sophia-Antipolis (I3S, IPMC, Eurecom) contribuant à un consortium européen sur l'intégration de l'ADN dans la hiérarchie de stockage des bases de données. Un autre exemple est celui d'une équipe d'informaticiens à Rennes (IRISA).

Il est notable que les deux petites entreprises mentionnées ci-dessus sont mobilisées pour des collaborations de fond par des entités états-uniennes ; l'une d'entre elles, DNA Script, a même été richement dotée en 2020 par une agence états-unienne

37 <https://www.journaldunet.com/solutions/cloud-computing/1141294-data-center-la-france-quatrieme-pays-le-mieux-equipe-au-monde/>

38 <https://www.datacentermap.com>

39 <https://www.biomemory-labs.com>

(IARPA, projet MIST⁴⁰). En novembre 2020, la « DNA data storage Alliance⁴¹ » a été fondée par quatre grosses entreprises : Microsoft, Western Digital (plus grand fabricant de mémoires), Illumina (dont les machines réalisent 96% des lectures d'ADN au monde) et Twist Bioscience (compagnie la plus avancée techniquement pour l'écriture de l'ADN). Cette Alliance états-unienne a vite été rejointe par Quantum. Plus significatif, de nombreuses entités européennes des secteurs public et privé l'ont rejointe aussi. Le risque est donc réel que les idées européennes soient développées aux États-Unis, qui en outre offrent des débouchés commerciaux. *A minima*, il semble donc indispensable de faire de la veille active dans le domaine et d'encourager les initiatives ponctuelles prises dans ce domaine par diverses entités pionnières sur sol français. Cependant, comme ce domaine émergent constitue probablement la clé de l'archivage soutenable des big data, il convient de réfléchir à un positionnement plus actif.

Face aux limites physiques qu'atteignent les centres de données, la technologie moléculaire d'archivage des big data a en effet le potentiel de devenir économiquement viable entre 2025 et 2040, progressant de marchés de niche vers des marchés plus globaux. Il s'agit donc bien d'un enjeu majeur et stratégique à horizon proche. Aussi, il serait souhaitable de capitaliser sur les laboratoires et petites sociétés qui ont été identifiés ainsi que sur le gisement plus large de compétences, afin de devenir des acteurs significatifs et d'ouvrir une perspective européenne qui agisse comme un contre-feu par rapport à la *DNA data storage Alliance* états-unienne déjà mentionnée.

L'objectif global serait de prendre ce train à son quasi-démarrage, puisqu'aussi bien la fabrication industrielle de mémoires conventionnelles est un train — de 57 milliards de dollars en augmentation de 2-3 % par an — qu'il est trop tard pour attraper en Europe.

40 <https://www.iarpa.gov/index.php/research-programs/mist>

41 <https://www.nature.com/articles/d41586-021-00534-w>

LE CALCUL QUANTIQUE CONCEPTS ET ÉTAT DES LIEUX

Boris Bourdoncle

Résumé

Le calcul quantique s'appuie des phénomènes propres à la physique quantique tels que l'intrication et la superposition d'états et constitue en cela un nouveau paradigme de calcul. En théorie, un ordinateur quantique peut effectuer certaines tâches de manière bien plus rapide que les ordinateurs dont nous disposons actuellement. Les premières réalisations physiques de ces processeurs d'un genre nouveau commencent à voir le jour, mais elles sont imparfaites et très limitées en puissance de calcul. Elles pourraient toutefois donner lieu à des applications à court ou moyen terme. En revanche, de nombreux progrès théoriques et expérimentaux restent nécessaires avant qu'un ordinateur quantique universel ne soit construit et que le potentiel du calcul quantique ne soit pleinement réalisé.

INTRODUCTION

Le concept d'ordinateur quantique s'est développé dans les années 1980. Paul Benioff fut le premier à proposer une réalisation physique de machine de Turing basée sur un système quantique¹. L'information et le calcul sont encore classiques, mais, pour la première fois, le support physique de ce calcul est quantique (ici et dans le reste de ce texte, « classique » est employé par opposition à « quantique »). Une idée similaire est formulée à la même époque par Yuri Manin en Russie. En 1981, la conférence *Physics of computation* rassemble physiciens et informaticiens au MIT. Au cours de sa leçon inaugurale intitulée « *Simulating physics with computers* »², Richard Feynman introduit l'expression « *quantum computer* », et avance l'idée suivante : simuler des phénomènes quantiques sur un ordinateur classique est inefficace, voire impossible, car modéliser le comportement d'un petit nombre de particules quantiques requiert une puissance de calcul inaccessible ; pourtant, le système à simuler et l'ordinateur utilisés sont eux-mêmes des dispositifs physiques composés de particules ; un simulateur/ordinateur se comportant lui-même de manière quantique serait bien plus efficace. Dans un texte publié en 1985, David Deutsch propose une généralisation quantique de la machine de Turing, qu'il nomme « ordinateur quantique universel »³. Cette machine d'un genre nouveau, avance-t-il, pourrait accomplir certaines tâches bien plus rapidement que n'importe quel ordinateur classique.

Dans les années 1990 sont conçus les premiers algorithmes quantiques qui permettent des gains de complexité quadratiques ou exponentiels par rapport aux meilleurs algorithmes classiques connus. Leurs premières exécutions sur des machines physiques ne tardent pas. En 1992, Deutsch et Richard Josza proposent un algorithme permettant de

-
- 1 P. Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines, *Journal of Statistical Physics* 22, 563–591 (1980).
 - 2 R. Feynman. Simulating physics with computers, *International Journal of Theoretical Physics* 21, 467–488 (1982).
 - 3 D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer, *Proceedings of the Royal Society of London A* 400, 97-117 (1985).

déterminer si une fonction est constante ou équilibrée⁴. Il sera exécuté pour la première fois via résonance magnétique nucléaire (RMN) par une collaboration IBM-Stanford-MIT en 1998⁵. En 1994, Peter Shor conçoit un algorithme permettant de factoriser un nombre en temps polynomial⁶, ce qui remet en cause la sécurité de nombreux protocoles cryptographiques ; le nombre 15 est factorisé par cet algorithme en 2001, là aussi par une collaboration IBM-Stanford utilisant la RMN⁷. En 1996, Lov Grover élabore un algorithme quantique de recherche⁸, dont la première exécution par une collaboration IBM-MIT-Berkeley date de 1998⁹.

Dans les années 2000-2010, les avancées théoriques, notamment sur la correction d'erreurs, et expérimentales, via la diversification des réalisations physiques d'ordinateurs quantiques, se sont accélérées. La recherche, jusqu'ici cantonnée au milieu universitaire et à quelques grandes firmes, se développe désormais également dans nombre de grands groupes du numérique et dans un nombre croissant de start-up. Un avantage quantique, c'est-à-dire, la résolution d'un problème par un ordinateur quantique de manière plus rapide que par n'importe quel ordinateur classique, a été mis en évidence à plusieurs reprises. Cet avantage est pour le moment conceptuel, c'est-à-dire qu'il n'a pas d'application pratique. Bien que l'ordinateur quantique universel demeure une perspective lointaine, de telles applications, rendues possibles par des machines intermédiaires imparfaites, pourraient voir le jour dans les années à venir, ce qui explique l'excitation industrielle et médiatique autour du domaine et la forte croissance des fonds publics et privés qui lui sont alloués.

L'objectif de ce texte est de faire le point sur la situation actuelle de l'informatique quantique, en se concentrant sur les aspects industriels et technologiques. Pour ce faire, nous revenons en premier lieu sur un ensemble de concepts et de notions utiles pour comprendre le domaine, puis nous dressons un état des lieux de la recherche industrielle.

-
- 4 D. Deutsch, R. Jozsa. *Rapid solution of problems by quantum computation*, Proceedings of the Royal Society of London A 439, 553–558 (1992).
 - 5 I. Chuang, L. Vandersypen, X. Zhou et al. Experimental realization of a quantum algorithm, *Nature* 393, 143–146 (1998).
 - 6 P. Shor. *Algorithms for quantum computation: discrete logarithms and factoring*, Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 124-134 (1994).
 - 7 L. Vandersypen, M. Steffen, G. Breyta et al. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance, *Nature* 4141, 883–887 (2001).
 - 8 L. Grover. A fast quantum mechanical algorithm for database search , Proceedings of the 28th Annual ACM Symposium on the Theory of Computing, 212-219 (1996).
 - 9 Chuang, N. Gershenfeld, M. Kubinec. Experimental Implementation of Fast Quantum Searching, *Physical Review Letters* 80, 3408 (1998).

CONCEPTS

Technologies quantiques, révolution (s) quantique (s), informatique quantique, simulation quantique... La physique quantique et les technologies de l'information et de la communication sont étroitement liées depuis les années 1950. Cependant, des concepts fondamentalement nouveaux sont apparus dans les années 1980. Leur fort potentiel de transformation est à l'origine des investissements colossaux observés aujourd'hui dans ce domaine, que l'on pourrait nommer les technologies de l'information quantique. Nous rappelons le sens de ce terme et les technologies qu'il recouvre. Nous nous concentrons ensuite sur le cas de l'ordinateur quantique, qui constitue l'une de ces technologies, et nous examinons les différents modèles de calcul envisagés et les notions qui permettent de les décrire.

PREMIÈRE ET DEUXIÈME RÉVOLUTIONS QUANTIQUES

Le développement de la physique quantique au début du XXe siècle a permis une compréhension très précise du comportement de la matière et de l'énergie à l'échelle microscopique. Les propriétés contre-intuitives et longtemps contestées qui découlent de cette théorie, telles que l'intrication et la superposition d'états, ont été expérimentalement observées et vérifiées à de nombreuses reprises depuis.

Cette compréhension fine du monde microscopique a abouti, en quelques décennies, à de nombreuses applications, dont les matériaux semi-conducteurs, bases des transistors et donc de l'électronique, ou les lasers, dont les usages médicaux et industriels sont considérables. Cette vague d'innovations a été qualifiée a posteriori de première révolution quantique.

Les possibilités techniques de l'électronique, combinées aux avancées théoriques de l'informatique et de la théorie de l'information, ont été le moteur de changements profonds et systémiques de notre société, tant et si bien que l'on appelle parfois la période débutant au milieu du XXe siècle 'ère de l'information'. Cette information est habituellement encodée de la manière la plus simple qui soit, c'est-à-dire en un nombre binaire, ou binary digit — bit —, valant '0' ou '1'. Le support physique qui encode cette information est un système qui peut prendre deux valeurs bien distinctes, tel que deux valeurs de tension dans un circuit électrique ou les deux positions d'un interrupteur.

Pour comparaison, le bit classique peut prendre les deux valeurs 0 et 1. Appelons « bit probabiliste » un dispositif qui prend la valeur 0 avec probabilité p et 1 avec la probabilité q ,

où $p + q = 1$. Un bit quantique, ou qubit (pour *quantum binary digit*), est une superposition $\alpha|0\rangle + \beta|1\rangle$ selon la notation bra-ket de Paul Dirac, où α et β sont des nombres complexes tels que $|\alpha|^2 + |\beta|^2 = 1$. Toutes les opérations quantiques se font sur l'état quantique, c'est-à-dire simultanément sur l'ensemble des éléments de la superposition. En revanche, l'état quantique ne peut être observé directement. Lorsque l'on mesure l'état pour connaître la valeur du qubit, celui-ci se comporte comme un bit probabiliste, ou $p = |\alpha|^2$ et $q = |\beta|^2$, de sorte que le résultat de la lecture est un bit classique. La lecture du qubit est en outre une opération irréversible : après la mesure, le qubit est projeté sur l'état mesuré, de sorte que la superposition est perdue. Toutefois, pour atteindre des gains de performance avec le calcul quantique, il faut réussir à tirer parti de la superposition d'états malgré la nature probabiliste de la mesure finale.

Les possibilités théoriques de l'encodage quantique de l'information en matière de technologies de l'information et de la communication ont été entrevues à partir de la fin des années 1970 et se sont développées dans les années 1980, comme illustré en introduction par l'exemple de l'ordinateur quantique. Les premières réalisations expérimentales sont concomitantes : l'une des expériences pionnières est celle d'Alain Aspect à l'Institut d'Optique entre 1980 et 1982¹⁰, qui a mis en évidence la violation d'une inégalité de Bell.

Le saut conceptuel d'une information classique à une information quantique constitue la seconde révolution quantique.

LES TECHNOLOGIES DE L'INFORMATION QUANTIQUE

Le domaine des technologies de l'information quantique est généralement divisé en trois ou quatre secteurs.

Le domaine des capteurs et de la métrologie quantiques correspond à l'utilisation d'un système quantique pour mesurer une quantité physique¹¹. L'exploitation de phénomènes quantiques permet une précision de mesure dépassant les limites classiques. Les applications sont nombreuses : systèmes de positionnement, imagerie... Les défis sont moindres que pour les autres technologies quantiques, et le niveau de maturité est plus avancé. Cela s'explique en partie par le fait que des systèmes quantiques, même de petite taille, donc relativement simples à manier, peuvent être utiles.

10 A. Aspect, P. Grangier, G. Roger. Experimental Tests of Realistic Local Theories via Bell's Theorem, *Physical Review Letters* 47, 460 (1981).

11 C. Degen, F. Reinhard, P. Cappellaro. Quantum sensing, *Reviews of Modern Physics* 89, 035002 (2017).

La deuxième catégorie de technologies quantiques est celle des communications. L'utilisation de phénomènes propres à la physique quantique, tels que le caractère aléatoire du résultat d'une mesure ou l'intrication, permet de concevoir des primitives cryptographiques inédites, et ainsi d'imaginer des systèmes de communication, et en particulier de communications sécurisées, d'un genre nouveau. Parmi les applications les plus connues, on peut citer la génération de nombres aléatoires et la distribution quantum de clef (Quantum Key Distribution — QKD)¹², dont le premier exemple est le protocole BB84, conçu par Gilles Brassard et Charles Bennett en 1984¹³.

La cryptographie dans son ensemble est doublement concernée par les technologies quantiques : non seulement les communications quantiques peuvent modifier et enrichir les standards cryptographiques en place, mais, par ailleurs, ces derniers sont menacés par l'avancée des calculateurs quantiques. En effet, ceux-ci pourraient résoudre certains problèmes bien plus rapidement que leurs pendants classiques, tels que la factorisation (cf. algorithme de Shor évoqué ci-dessus). Or, la sécurité de nombreux protocoles cryptographiques repose sur la difficulté supposée à résoudre ces mêmes problèmes. Face à cette menace, le recours à la cryptographie quantique, ou bien à d'autres protocoles de cryptographie classique dont la sécurité est garantie même face à la puissance de calcul des ordinateurs quantiques, dénommée 'cryptographie post-quantique', s'avérera indispensable, et ce, même avant que des ordinateurs quantiques universels ne voient le jour, si l'on souhaite que des informations échangées aujourd'hui soient protégées pour plusieurs années à venir¹⁴.

Le troisième domaine est l'objet de ce chapitre : il s'agit justement du calcul (ou de l'informatique) quantique. On place parfois la simulation quantique dans une catégorie séparée, selon qu'on adopte une définition plus ou moins restreinte de ce qu'est un calcul ou un ordinateur. En effet, un simulateur quantique n'a pas pour objet d'exécuter un programme conformément à un algorithme ; un simulateur quantique est un système physique qui est préparé et manipulé de manière à fournir de l'information sur une ou des propriétés d'un autre système plus complexe¹⁵. Cela correspond à l'intuition originelle de Feynman : un dispositif quantique simplifié pour modéliser, simuler et ainsi comprendre un système quantique plus élaboré. Dans ce texte, nous traitons la simulation quantique comme une catégorie restreinte du calcul quantique.

////////////////////////////////////

- 12 R. Alléaume et al. Using quantum key distribution for cryptographic purposes: a survey, *Theoretical Computer Science* 560, 62-81 (2014).
- 13 C. Bennett, G. Brassard. *Quantum cryptography: Public key distribution and coin tossing*, Proceedings of IEEE International Conference on Computers, Systems and Signal Processing (1984).
- 14 M. Mosca, M. Piani. *Quantum Threat Timeline*, Global Risk Institute (2019).
- 15 Antonio Acín et al. The quantum technologies roadmap: a European community view, *New Journal of Physics* 20, 080201 (2018).

CALCULABILITÉ ET COMPLEXITÉ

L'ordinateur quantique est une innovation de rupture qui pourrait modifier en profondeur de nombreux usages de notre société de l'information, tels que la cryptographie, l'optimisation ou la modélisation. Cependant, les modèles de calcul quantique ne définissent pas une nouvelle forme de calculabilité : les problèmes que peut résoudre un ordinateur classique peuvent être résolus par un ordinateur quantique, et inversement. En d'autres termes, une machine de Turing classique peut simuler un ordinateur quantique, et le calcul quantique ne remet pas en cause la thèse de Church-Turing.

En revanche, il existe des raisons de penser que le calcul quantique invalide la thèse de Church-Turing étendue, c'est-à-dire que certains problèmes peuvent être résolus de manière exponentiellement plus rapide par un ordinateur quantique que par une machine de Turing classique¹⁶. Cela signifierait que les classes de complexité des ordinateurs quantiques diffèrent des classes de complexité des machines classiques : la classe 'BQP' (*'bounded-error quantum polynomial time'*) des problèmes de décision qu'un ordinateur quantique peut résoudre en un temps polynomial avec probabilité d'erreur inférieure à 1/3 serait donc plus grande que son pendant classique, la classe BPP (*'bounded-error probabilistic polynomial time'*).

C'est là que réside le potentiel de rupture de l'informatique quantique, comme l'illustre l'algorithme de Shor : la factorisation d'un grand nombre peut être accomplie par un ordinateur classique ou quantique ; cependant, elle peut l'être en temps polynomial grâce à l'algorithme de Shor, alors que les meilleurs algorithmes classiques connus sont sous-exponentiels. Toutefois, un ordinateur quantique ne permettrait pas de résoudre tous les problèmes de manière exponentiellement plus rapide qu'un ordinateur classique, tant s'en faut. Pour certains problèmes, le calcul quantique ne permet même aucun gain de temps. Les gains de complexité possibles selon les différents types de problèmes considérés ne sont pas toujours bien connus, mais l'on dispose de résultats partiels dans ce domaine¹⁷.

Lorsque ce gain de complexité potentiel est effectivement réalisé par une machine quantique pour un problème donné, on parle de suprématie quantique¹⁸ ou, plus récemment, d'avantage quantique. Le fait que le problème traité soit utile ou non n'a ici pas d'importance : il s'agit d'une étape conceptuelle dans le développement de l'informatique quantique, qui

16 E. Bernstein, U. Vazirani. Quantum Complexity Theory, *Siam Journal on Computing* 26(5), 1411-1473 (1997).

17 F. Magniez. *Limites du calcul quantique : liens entre complexité classique et quantique*, cours au Collège de France (2021).

18 J. Preskill. *Quantum computing and the entanglement frontier*, 25th Solvay Conference on Physics (2011).

mêle progrès théoriques sur l’algorithmique quantique et progrès expérimentaux sur la réalisation physique d’un ordinateur quantique, et qui est par ailleurs une cible en mouvement, puisque le gain de temps doit se mesurer face aux meilleurs superordinateurs et aux meilleurs algorithmes classiques du moment.

Les paradigmes de calcul quantique

Il existe plusieurs modèles de calcul quantique¹⁹. Le plus répandu est le modèle dit ‘par circuit’ ou ‘par porte’ (*circuit-based* ou *gate-based*) : un calcul correspond à un ensemble de portes quantiques, analogues des portes logiques classiques issues de l’algèbre de Boole, appliqué à un état quantique généralement initialisé dans l’état $|0\dots 0\rangle$, et à l’issue duquel on effectue une mesure. Contrairement aux portes logiques classiques, les portes quantiques doivent être réversibles. Elles correspondent à des matrices unitaires. Un ensemble donné de portes, qui correspond, lorsqu’on y ajoute la mesure finale, à un calcul quantique, est donc une transformation unitaire de l’état quantique initial. Toutes les transformations peuvent être approchées de manière aussi précise que voulu par une séquence de portes logiques appartenant à un ensemble fini de quelques portes élémentaires appelé ‘ensemble de portes universel’(théorème de Solovay-Kitaev).

Un autre modèle assez répandu est le calcul quantique adiabatique : un opérateur hamiltonien, qui représente l’énergie totale d’un système, est donné ; un système quantique est préparé dans un état initial correspondant à l’état fondamental de ce même hamiltonien ; on fait évoluer cet hamiltonien initial vers un hamiltonien final, dont on sait que l’état fondamental représente la solution à notre problème (en quelque sorte, l’hamiltonien final encode le problème à résoudre et son état fondamental représente la solution du problème). D’après le théorème adiabatique, si cette évolution est suffisamment lente et si l’écart entre l’état fondamental et le premier état excité est suffisamment grand, l’état initial évolue en restant l’état fondamental du hamiltonien, jusqu’à devenir l’état fondamental de l’hamiltonien final : l’état du système est donc devenu la solution du problème. Bien que la formulation soit radicalement différente, les deux paradigmes de calcul, calcul par circuit et calcul adiabatique sont équivalents²⁰. Le calcul adiabatique est bien adapté à des problèmes d’optimisation, car ceux-ci sont souvent facilement représentables sous une forme hamiltonienne. Toutefois, contrairement au calcul par circuit, on ne sait pas vraiment traiter les problèmes liés au bruit et aux imperfections dans le cas du calcul adiabatique (cf. ci-dessous). Il s’agit donc d’un modèle utile, mais il est peu probable qu’il constitue la base d’un ordinateur quantique universel de grande taille.

////////////////////////////////////

19 S. Jordan. *Quantum Computation Beyond the Circuit Model*, Thèse de doctorat, Massachusetts Institute of Technology (2008).

20 D. Aharonov et al. Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation, *SIAM Journal of Computing* 37, 166-194 (2007).

Un autre type de calcul, qui ne constitue pas un paradigme de calcul quantique à part entière, est le recuit simulé quantique (*'quantum annealing'*), inspiré du recuit simulé classique. Il s'agit d'une métaheuristique permettant de résoudre certains problèmes d'optimisation de manière approchée. C'est une version bruitée du calcul adiabatique, dans laquelle l'évolution de l'état initial ne se fait pas nécessairement de manière suffisamment lente pour être adiabatique. Les solutions obtenues peuvent être intéressantes, mais il y a peu de résultats formels sur l'efficacité et l'exactitude de ces calculs.

Enfin, il existe d'autres modèles, comme le *'measurement-based quantum computing'* ou le calcul quantique topologique. Tous ces modèles sont équivalents en matière de calculabilité, mais ils sont plus ou moins adaptés quand il s'agit de réaliser physiquement un ordinateur quantique universel de grande taille (c'est-à-dire, comptant un grand nombre de qubits). La scalabilité est en effet liée à la capacité à fonctionner avec des réalisations imparfaites, et donc des erreurs, qui varie d'un modèle à l'autre. Le modèle par circuit est aujourd'hui souvent considéré comme le meilleur candidat à moyen terme. Le modèle topologique est naissant, mais son potentiel est important, car les réalisations physiques de ces qubits pourraient être très robustes au bruit et aux erreurs.

ERREURS ET BRUITS

Tel qu'introduit ci-dessus, un qubit est une unité d'information abstraite, comme l'est un bit en informatique classique. Pour construire une machine quantique physique, il faut concevoir un système quantique matériel à deux niveaux, qui corresponde à la réalisation de cette unité abstraite. C'est la différence entre qubit logique (abstrait) et qubit physique (matériel).

Or, les réalisations physiques de qubits sont très délicates : lorsque ces qubits physiques sont manipulés pour exécuter un calcul, les erreurs peuvent être nombreuses. À cette difficulté s'ajoute le problème de la décohérence, qui entraîne la perte des particularités quantiques d'un système lorsque celui-ci interagit avec son environnement. Les spécificités quantiques, qui font la richesse du calcul quantique, constituent ainsi également un obstacle à son exécution physique :

- le théorème d'impossibilité de clonage quantique indique qu'il est impossible de cloner l'information quantique, rendant les techniques classiques de redondance inapplicables,
- la réduction du paquet d'ondes empêche de mesurer un état quantique sans perdre ce même état,
- le caractère continu d'un qubit $\alpha|0\rangle + \beta|1\rangle$ implique que les erreurs de transmission et de calcul sont elle-mêmes continues.

Toutefois, une vaste théorie dite des codes correcteurs d'erreurs quantiques (*'quantum error correction code'*), par analogie avec les codes correcteurs classiques, s'est développée afin de concevoir des outils théoriques pour protéger l'information quantique des erreurs et de la décohérence²¹. L'idée est qu'un qubit logique soit encodé à l'aide de plusieurs qubits physiques, de telle sorte que les erreurs qui affectent les qubits physiques puissent être corrigées et que le qubit logique soit préservé.

Lorsque les qubits sont encodés via un code correcteur, les opérations à effectuer sur les qubits physiques pour réaliser un calcul doivent elles-mêmes être redéfinies. La théorie correspondante est le calcul quantique tolérant aux fautes (*'fault-tolerant quantum computing'*)²². Le théorème du seuil quantique indique que, lorsque le taux d'erreurs par opération physique se situe en dessous d'un certain seuil, il est possible de corriger les erreurs logiques avec une précision arbitraire, et ainsi d'accomplir tout calcul²³. En d'autres mots, ce théorème garantit que le calcul quantique est physiquement réalisable avec une précision arbitraire en dépit des erreurs physiques, du moment que celles-ci sont bornées.

ORDINATEUR QUANTIQUE UNIVERSEL ET CALCULATEUR SPÉCIALISÉ

Une machine quantique capable d'effectuer tout calcul quantique est dite universelle. Elle est programmable, en ce sens qu'il existe des paramètres de contrôle que l'on peut faire varier pour effectuer le calcul souhaité. Une telle machine restera hors de portée à court terme parce qu'on ne sait actuellement pas manipuler les très grands nombres de qubits nécessaires au calcul tolérant aux fautes. Précisons que le terme 'universel' s'entend ici au sens de « modèle de calcul universel ». Même si une machine quantique universelle voyait le jour à moyen terme, cela ne signifierait pas qu'elle serait utilisée universellement, c'est-à-dire pour toutes les applications envisageables pour un dispositif de traitement de l'information.

Quoi qu'il en soit, la réalisation d'un tel calculateur quantique reposera conjointement sur des progrès théoriques et expérimentaux : les avancées théoriques sur la correction d'erreur et la tolérance aux fautes réduiront les exigences sur les systèmes physiques ; les progrès expérimentaux permettront de créer et de manipuler des systèmes quantiques de manière toujours plus précise. En outre, des progrès sur toutes les couches intermédiaires des

21 D. Gottesman. *An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation*, Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics 68, 13-58 (2010).

22 Ibid.

23 E. Knill, R. Laflamme, W. Zurek. Resilient Quantum Computation, *Science* 279, 342-345 (1998).

pires logicielles et matérielles seront nécessaires : langages de programmation, système opératoire, gestion des entrées et sorties, compilateur...

Parallèlement à ces avancées, il est possible de concevoir des machines spécialisées, qui mettent de côté les objectifs de scalabilité et de fonctionnement idéal pour tirer profit à court terme des particularités du calcul quantique dans des scénarios plus restreints, avec des systèmes imparfaits et de taille modeste.

Nous avons déjà cité l'exemple des simulateurs quantiques : un système quantique peut être simulé sur un ordinateur quantique universel à circuit, mais une simulation peut aussi être effectuée à l'aide d'un autre système quantique, plus simple à manipuler et dont certaines propriétés se comportent de manière similaire au système à simuler. On appelle parfois ce second type plus restreint de machine un simulateur quantique analogique, tandis que le premier cas correspond à une simulation quantique numérique²⁴ (voir ci-dessous).

Un autre exemple de machine spécialisée que nous avons évoqué est le recuit simulé quantique. Les *quantum annealers* sont aujourd'hui utilisés pour résoudre des problèmes d'optimisation combinatoire²⁵, bien qu'il n'y ait pas de consensus sur le gain qu'ils permettent réellement par rapport à des méthodes classiques²⁶. Les *annealers* pourraient par ailleurs s'avérer utiles pour d'autres tâches que l'optimisation, comme la simulation²⁷.

Enfin, une classe de machine désignée par un terme proposé en 2018 par John Preskill²⁸ reçoit beaucoup d'attention depuis deux ans : il s'agit des technologies NISQ, pour '*Noisy Intermediate-Scale Quantum*'. Ce terme désigne des calculateurs composés de 50 à quelques centaines de qubits physiques manipulés de manière imparfaite, c'est-à-dire avec des taux d'erreurs non nuls. En principe, il ne s'agit pas de calculateurs spécialisés, mais d'une étape vers le calculateur quantique universel de grande taille évoqué plus haut. Toutefois, ils méritent une attention particulière, d'une part car ils correspondent à l'état de l'art de la recherche actuelle, et d'autre part, car ils permettent d'ores et déjà d'atteindre l'avantage quantique (cf. partie suivante). Ils laissent donc entrevoir des applications utiles dans les années à venir, sous la forme de machines hybrides quantiques-classiques, composées d'un petit « processeur quantique » de quelques dizaines de qubits fonctionnant

24 I. Georgescu, S. Ashhab, F. Nori. Quantum simulation, *Review of Modern Physics* 86, 153 (2014).

25 P. Hauke et al. Perspectives of quantum annealing: methods and implementations, *Reports on Progress in Physics* 83, 054401 (2019).

26 T. Rønnow et al. Defining and detecting quantum speedup, *Science* 354, 420-424 (2014).

27 Andrew King et al. Observation of topological phenomena in a programmable lattice of 1,800 qubits, *Nature* 560, 456-460 (2018).

28 J. Preskill. Quantum Computing in the NISQ era and beyond, *Quantum* 2, 79 (2018).

en relation avec un ordinateur classique. Ce dernier réalise des opérations d’optimisation et d’apprentissage automatique et met à jour les paramètres du processeur quantique en conséquence.

CALCUL ANALOGIQUE ET CALCUL NUMÉRIQUE

Comme pour les calculateurs classiques, on peut établir une distinction, mentionnée dans le paragraphe précédent, entre calculateur quantique analogique (*‘analog’*) et calculateur quantique numérique (*‘digital’*). L’usage de ces termes est cependant changeant selon les auteurs : on qualifie parfois de ‘numérique’ le modèle par circuit, tandis que les autres, en particulier le calcul adiabatique et sa version bruitée le recuit simulé quantique, sont alors qualifiés de ‘calcul quantique analogique’; parfois, on réserve cette différence terminologique à la simulation, pour distinguer les simulations effectuées grâce à un calculateur quantique universel et les simulations via des systèmes physiques spécialisés.

Pour lever l’ambiguïté, on peut définir un calcul (classique ou quantique) numérique de la manière suivante²⁹ : un modèle de calcul est dit numérique si toutes les entrées et toutes les sorties possibles appartiennent à un ensemble d’états discret, et si toutes les fonctions applicables, elles-mêmes spécifiées par un ensemble discret de paramètres, peuvent être réalisées à l’aide d’un ensemble discret d’opérations.

Selon cette définition, le calcul adiabatique et les simulations basées sur l’évolution continue d’un système physique constituent donc bien des calculs analogiques, tandis que le modèle de calcul par circuit ou par porte relève du calcul numérique, mais uniquement lorsque les fonctions applicables sont bâties à partir d’un ensemble discret de portes élémentaires, tel qu’un ensemble de portes universel.

Les machines NISQ, qui fonctionnent sans correction d’erreurs et avec des opérations imparfaites, sont, en ce sens, analogiques. Elles ne permettront donc pas d’exécuter des programmes conformément aux algorithmes emblématiques cités plus haut, mais la recherche d’algorithmes spécialement conçus pour ces machines rudimentaires, qui permettraient de donner lieu à des applications utiles à court terme, est un champ de recherche dynamique³⁰.

29 I. Deutsch, Harnessing the Power of the Second Quantum Revolution, *PRX Quantum* 1, 020101 (2020).

30 K. Bharti et al. Noisy intermediate-scale quantum (NISQ) algorithms, *arXiv:2101.08448* (2021).

ÉTAT DES LIEUX

La recherche dans le domaine du calcul quantique s'intensifie, comme en témoignent le nombre croissant de publications et de dépôts de brevets. Elle est désormais multiple : académique et industrielle ; théorique et expérimentale ; fondamentale et appliquée ; matérielle et logicielle... Les investissements importants et l'attention médiatique donnent une forte résonance à certaines avancées majeures ou présentées comme telles. Les effets d'annonce coexistent avec des percées véritables. Nous donnons ici une vue d'ensemble de l'état de la recherche industrielle dans le domaine.

LES RÉALISATIONS PHYSIQUES DE QUBITS

Le qubit a été initialement défini comme une unité d'information, mais désormais on utilise également ce terme pour désigner un système quantique à deux niveaux qui peut être utilisé comme support de cette information. Dans ce second sens, il existe de nombreux candidats qubits possibles, caractérisés par des propriétés différentes concernant leur initialisation, leur manipulation, leur durée de vie... Le choix d'un qubit est déterminant dans la construction d'un ordinateur quantique, comme en témoignent les critères de Divincenzo. David Divincenzo a formulé ces cinq critères en 2000, comme autant de conditions que doit remplir un système physique pour qu'il puisse être utilisé pour réaliser un ordinateur quantique³¹ :

1. Un système quantique extensible/scalable avec un qubit bien caractérisé ;
2. La possibilité d'initialiser l'état des qubits dans un état standard simple ;
3. Un temps de cohérence suffisamment long ;
4. Un ensemble de portes quantiques universel ;
5. La possibilité de mesurer l'état final des qubits.

Il existe par ailleurs deux critères supplémentaires concernant la communication.

Nous citons ici six technologies particulièrement bien établies^{32,33,34,35}.

31 D. Divincenzo. The Physical Implementation of Quantum Computation, *Fortschritte der Physik* 48, 771-783 (2000).

32 Antonio Acín et al., *ibid.*

33 *L'essentiel sur le calcul et l'ordinateur quantique* CEA. <https://www.cea.fr/comprendre/Pages/nouvelles-technologies/essentiel-sur-ordinateur-quantique.aspx> (publié : 2 avril 2021 ; consulté : avril 2021).

34 Qubit implementations, *The Quantum Daily*, <https://thequantumdaily.com/qubit-implementations/> (consulté : avril 2021)

35 US National Academies of Sciences, Engineering, and Medicine. *Quantum Computing: Progress*

Supraconducteur. Deux matériaux supraconducteurs sont séparés par une fine barrière. Grâce à l'effet Josephson, un courant peut circuler au travers de cette barrière. Le support de l'information peut être : l'énergie (qubit de phase), et les deux niveaux sont alors l'état fondamental et le premier état excité ; le courant (qubit de flux), et les deux niveaux sont donnés par son sens de circulation ; la charge (qubit de charge), nulle ou non nulle. Leur fabrication peut s'appuyer en partie sur l'industrie des semi-conducteurs, et les opérations sur ces qubits peuvent être très rapides, mais leur temps de cohérence est bref et ils doivent être maintenus à très basse température (inférieure à 100 mK).

Ion piégé. Un ion est piégé par un champ électromagnétique, et l'information correspond à deux niveaux d'énergie différents. Ils ont des temps de cohérence plus longs que toutes les autres technologies, mais les opérations sur ces qubits sont lentes. En outre, plusieurs obstacles technologiques demeurent avant que la scalabilité ne puisse être garantie.

Photon. Le qubit est encodé par un photon, et plusieurs paramètres peuvent constituer le support de l'information, tels que sa polarisation (horizontale ou verticale) ou sa présence (absence/présence). Ces qubits peuvent fonctionner avec des dispositifs photoniques sur silicium, compatibles avec l'industrie des semi-conducteurs existante. La scalabilité semble prometteuse, mais n'est pas garantie, et les opérations nécessitant des interactions entre photons sont encore difficiles à accomplir.

Semi-conducteur/silicium. Le spin d'un électron est utilisé comme support d'information. Ces qubits ont un temps de cohérence important, et leur production pourrait s'appuyer assez directement sur l'industrie des semi-conducteurs actuelle. Néanmoins, accomplir des opérations impliquant plusieurs de ces qubits avec un faible taux d'erreur reste un défi.

Atome froid. L'approche est similaire à celle des ions piégés, mais un atome remplace l'ion, d'où le nom alternatif d'atome neutre. Les avantages et les inconvénients sont donc semblables, mais ces techniques sont un peu moins matures.

Centre NV. Un centre azote-lacune (*'Nitrogen-vacancy center'*) est un défaut ponctuel dans la structure du diamant, et le spin des électrons localisés autour de ces centres peut être utilisé comme qubit. Ces qubits pourraient être très stables, mais le niveau de maturité est bas et plusieurs défis demeurent, tels que la génération efficace de ces défauts et la scalabilité.

Sur ces six technologies, les supraconducteurs et les ions piégés occupent les première et deuxième places en niveau de maturité, car elles ont fait l'objet d'investissements plus

and Prospects, The National Academies Press (2019).

précoces et plus importants. Elles constituent les deux technologies privilégiées par la R&D, mais les autres possibilités gagnent du terrain.

Il est possible qu'une technologie s'impose petit à petit et devienne la technologie de référence pour les calculateurs quantiques, à l'instar des transistors pour le calcul classique. Toutefois, il est bien trop tôt pour en choisir une. En outre, il est possible que différentes technologies se développent conjointement pour des usages complémentaires.

Signalons que, dans les synthèses qui recensent les acteurs de l'informatique quantique et les classent par technologies, on trouve parfois, aux côtés de technologies citées ci-dessus, les qubits topologiques ou adiabatiques. Comme décrit dans la partie précédente, il s'agit là de modèles de calcul différents, et non seulement d'une alternative technologique parmi d'autres pour construire un ordinateur quantique par circuit. Il est pertinent de recenser les acteurs qui font ces choix aux côtés des autres, mais ces catégories doivent être vues de manière différente. Par exemple, les qubits topologiques peuvent être réalisés physiquement par des fermions de Majorana ou par des anyons.

Enfin, ces technologies ont des besoins énergétiques différents, mais toujours importants, dus par exemple à la cryogénie nécessaire au fonctionnement de certains composants. Cependant, il est bien trop tôt pour estimer la consommation énergétique à venir du secteur, puisque les usages sont loin d'être clairement identifiés et que les progrès technologiques sont très rapides. Il est néanmoins possible de commencer à tenir compte de ces considérations dès maintenant³⁶.

LES MÉTRIQUES DE PERFORMANCE

Alors que les annonces concernant les avancées technologiques se multiplient, des métriques standards pour réaliser des analyses comparatives objectives deviennent essentielles. Or, ces métriques sont aujourd'hui encore l'objet de recherche et de discussion, en raison d'une maturité technologique insuffisante du domaine.

Pour une technologie de qubit donnée, il existe plusieurs paramètres pertinents, tels que le temps de cohérence ou le taux d'erreur sur les portes logiques. La qualité d'une machine quantique est ensuite souvent exprimée en nombre de qubits physiques atteints, dont le record varie de la dizaine à 127³⁷ selon les technologies choisies (bien plus pour le recuit

36 A. Auffèves, Quantum technologies need a quantum energy initiative, *arXiv:2111.09241* (2021).

37 P. Ball, First quantum computer to pack 100 qubits enters crowded race, *Nature*, <https://www.>

simulé quantique, mais rappelons qu'il s'agit d'un modèle de calcul restreint et non d'un modèle équivalent au calcul par circuit, pour lequel la notion même de qubit revêt un sens différent). Toutefois, cette valeur ne donne qu'une indication vague, puisque les performances réelles dépendent fortement de la qualité des qubits. Rapporter l'ensemble des données sur la qualité ainsi que le nombre total de qubits fournit également une quantité d'information limitée, puisque les performances rendues possibles par ces valeurs dépendent de l'architecture et donc intrinsèquement des technologies choisies. Ces informations ne permettent donc pas, ou mal, de comparer les technologies entre elles.

Des propositions de mesures et de protocoles intégrant un ensemble de valeurs pertinentes se développent³⁸, basées sur le volume quantique ou l'analyse randomisée ('randomized benchmarking'), et qui sont universelles en ce sens qu'elles permettent de comparer les performances des différentes technologies. Ces métriques peuvent être centrées sur les composants ou bien sur les applications³⁹.

Enfin, rappelons que les performances avancées en nombre de qubits par les grands acteurs du secteur se réfèrent jusqu'à présent au nombre de qubits physiques, et non logiques. Le nombre de qubits physiques nécessaires pour encoder des qubits logiques et réaliser des opérations sans erreurs dépend du modèle de calcul, du code correcteur utilisé, de la fidélité avec laquelle les qubits sont créés et manipulés et de leur nombre total, mais il est probable que 100 à 1000 qubits physiques au moins soient nécessaires pour réaliser un seul qubit logique.

LES ENTREPRISES DE L'INFORMATIQUE QUANTIQUE

La recherche actuelle en informatique quantique se conduit dans des universités et des centres de recherche publics (qui ne sont pas l'objet de ce texte), mais également au sein d'entreprises en nombre croissant. On peut regrouper les acteurs privés du calcul quantique en trois grandes catégories : les acteurs de la recherche sur le matériel (*hardware*), les acteurs du logiciel (*software*) et de l'algorithmique, et les utilisateurs potentiels du calcul quantique.

[nature.com/articles/d41586-021-03476-5](https://www.nature.com/articles/d41586-021-03476-5) (publié : 19 novembre 2021, consulté : décembre 2021)

38 R. Blume-Kohout, K. Young. A volumetric framework for quantum computer benchmarks, *Quantum* 4, 362 (2020).

39 T. Lubinski et al. Application-oriented performance benchmarks for quantum computing, *arXiv:2110.03137* (2021).

Matériel

On trouve différents types de sociétés dans la première catégorie, dont la plupart sont recensées en Figure 1, aux côtés des acteurs académiques. Les grandes entreprises du numérique, grâce à leur pouvoir d'investissement important, y sont bien représentées. IBM y fait figure de pionnier. Outre ses prototypes de processeurs quantiques à qubit supraconducteur, la firme a développé l'IBM Quantum Experience, un service de calcul quantique en cloud qui permet à tous d'accéder à un processeur de 16 qubits. Les processeurs quantiques développés par Google sont bien connus, particulièrement depuis qu'un processeur de 53 qubits supraconducteurs dénommé Sycomore a permis de mettre en évidence le premier avantage quantique⁴⁰. L'écart entre le temps mis par Sycomore pour accomplir la tâche et le temps supposément nécessaire au meilleur supercalculateur a depuis été remis en question et revu à la baisse, mais l'expérience n'en reste pas moins un succès majeur. Depuis, deux autres expériences d'avantage quantique ont été rapportées en Chine^{41,42}. Parmi les géants du numérique, Microsoft est le seul à avoir fait le choix de l'approche topologique. De grands groupes hors du secteur du numérique développent leur propre recherche en calcul quantique, tels que Honeywell qui a fait le choix des ions piégés. À cela, s'ajoutent des entreprises nouvelles entièrement consacrées au calcul quantique, des 'pure players', dont les plus connus sont la canadienne D-Wave systems (*quantum annealers*), la canadienne Xanadu et l'américaine PsiQuantum (photons), l'américaine IonQ (ions piégés), l'américaine Rigetti et la finlandaise IQM (supraconducteurs). L'écosystème européen est également dynamique, avec des startup récemment issues de l'université telles qu'*Alpine quantum technologies* (Autriche, ions piégés), *Oxford quantum circuits* (Royaume-Uni, supraconducteurs), ou, en France, Pasqal (atomes froids) et Quandela (photons).

Logiciel et algorithmique

Les grandes firmes du numérique sont souvent également présentes du côté logiciel et algorithmique. Elles développent par exemple des kits de développement logiciel et des langages de programmation (Q# chez Microsoft, Amazon Braket chez AWS, Qiskit chez IBM...). Le nouveau *AWS Center for Quantum Computing* s'est récemment illustré en proposant une architecture nouvelle de calcul tolérant aux fautes⁴³. En France, Atos est la seule grande entreprise active dans ce domaine, notamment via la Quantum Learning

40 F. Arute et al. Quantum supremacy using a programmable superconducting processor, *Nature* 574, 505–510 (2019).

41 H.-S. Zhong et al. Phase-programmable gaussian boson sampling using stimulated squeezed light, *Physical Review Letters* 127, 180502 (2021)

42 Y. Wu et al. Strong quantum computational advantage using a superconducting quantum processor, *Physical Review Letters* 127, 180501 (2021).

43 C. Chamberland et al. Building a fault-tolerant quantum computer using concatenated cat codes, arXiv:2012.04108 (2020).

Machine, un supercalculateur classique capable de simuler plusieurs dizaines de qubits, et qui constitue donc une aide pour développer et tester des programmes quantiques. Enfin, comme pour le matériel, on trouve ici des *pure players* qui conçoivent des plateformes de développement ou des aides au développement d’algorithmes quantiques, tels que *Zapata computing*, *Classiq* ou *Cambridge Quantum Computing*.

Partenariats industriels

À ces initiatives s’ajoutent l’investissement récent des groupes industriels utilisateurs potentiels du calcul quantique, qui entrevoient avec l’arrivée des machines hybrides NISQ des cas d’usages viables à court terme, et nouent des partenariats avec des groupes de recherche ou des entreprises du secteur pour les examiner. Les domaines concernés sont multiples :

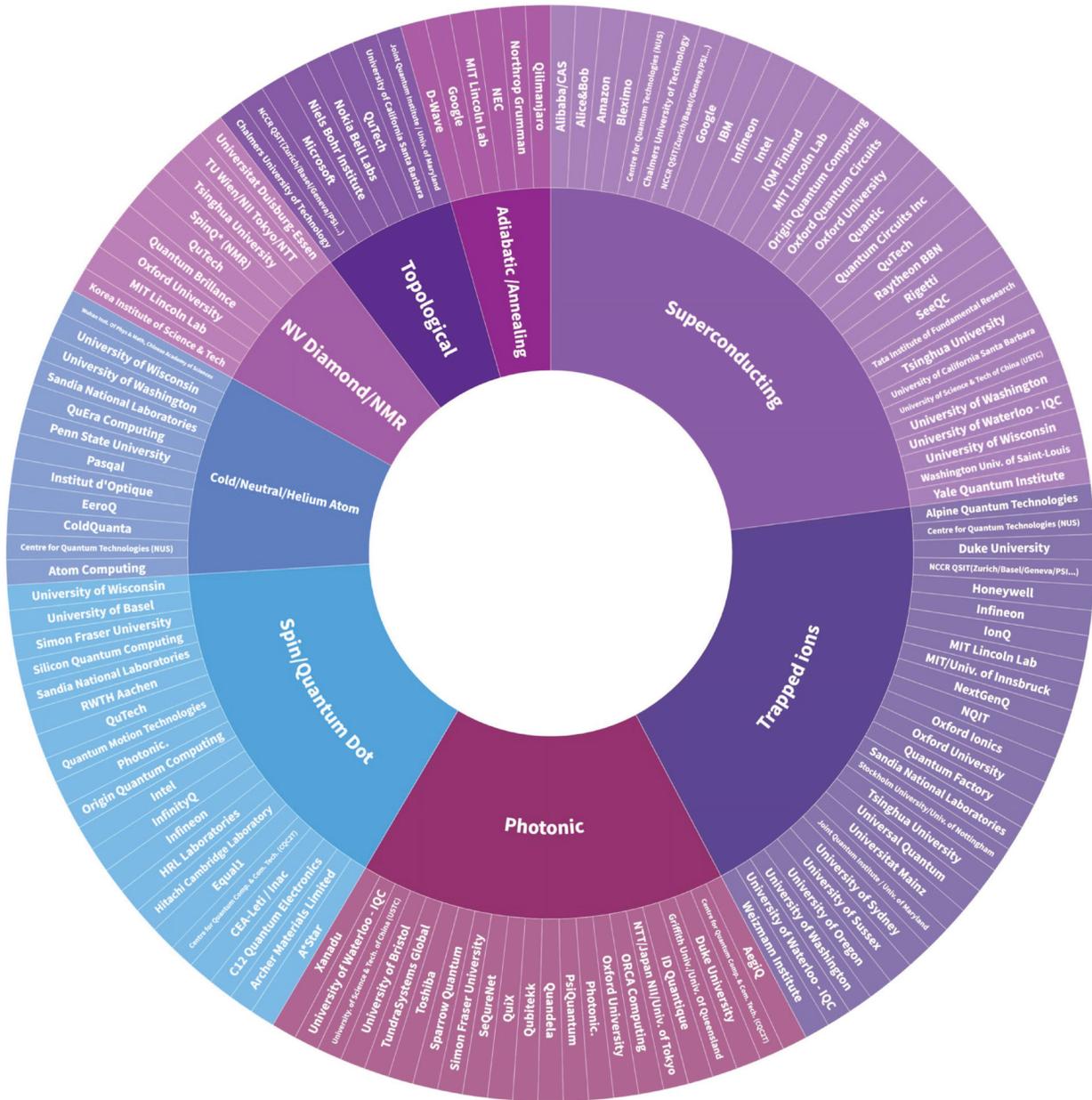
- simulation moléculaire pour le développement de nouveaux médicaments (pharmacie), de nouveaux matériaux (construction, batteries), de nouveaux catalyseurs (capture et utilisation du carbone) ;
- dynamique des fluides, avec des applications dans l’aéronautique et l’automobile ;
- optimisation de réseaux de transport, d’énergie, de télécommunication ;
- simulation de marché pour la gestion de portefeuilles et la gestion de risques,

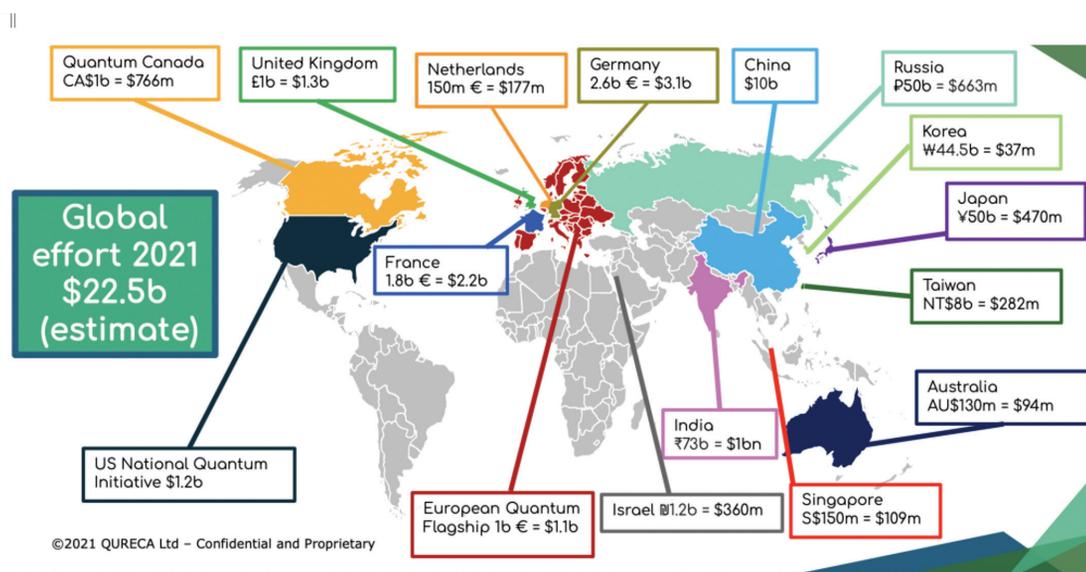
pour ne citer que quelques exemples. Les possibilités de bénéfices à court terme ne sont pas confirmées : certains résultats récents concernant l’optimisation sont plutôt négatifs⁴⁴, tandis que d’autres concernant l’apprentissage automatique quantique (‘quantum machine learning’) sont encourageants⁴⁵. Quoi qu’il en soit, les importants gains potentiels expliquent l’engouement des acteurs industriels, tels que Lockheed Martin avec l’Université de Californie du Sud et D-Wave, Total avec Cambridge Quantum Computing ou Covestro avec Google. Le consortium NEASQC (Next Applications of Quantum Computing) rassemble à cette fin industriels et académiques en Europe : on y trouve par exemple EDF, HSBC, AstraZeneca, le Irish Center for High-end Computing, l’Université de Lorraine, HQS Quantum simulations...

Enfin, mentionnons le regroupement d’acteurs européens issus de ces trois groupes au sein du *Quantum Industry Consortium (QuIC)*, dont l’assemblée fondatrice s’est tenue le 4 février 2021, et dont les membres fondateurs sont Airbus, Alpine Quantum Technologies, Atos, BBVA, Bosch, ID Quantique, IQM, Quantonation, Qu&Co, Quside Technologies, SAP, Thales, Toptica Photonics et Zurich Instruments.

44 D. Stilck França et al. Limitations of optimisation algorithms on noisy quantum devices *Nature Physics* (2021).

45 H.-Y. Huang et al. Quantum advantage in learning from experiments, arXiv:2112.00778 (2021).





Initiatives nationales et européennes pour les technologies quantiques

© Qureca, reproduit ici avec l'aimable autorisation d'Araceli Venegas-Gomez⁴⁶

LES PLANS NATIONAUX ET LES INVESTISSEMENTS

Les investissements publics et privés consacrés au calcul quantique ont augmenté de manière considérable au cours des années 2010, via les fonds publics désormais souvent coordonnés au sein de plans nationaux, la R&D des grandes firmes du secteur et le capital-investissement qui financent les nouvelles entreprises spécialisées.

De nombreux pays ont annoncé des plans d'investissements importants. Ils couvrent en général toutes les technologies de l'information quantique, et pas seulement le calcul quantique, et regroupent des fonds préexistants et des investissements nouveaux, et incluent parfois des fonds privés aux côtés de la dépense publique. Ils permettent aux États de montrer leur volontarisme et de coordonner les efforts de recherche nationaux. La Figure 2 recense les principaux montants annoncés. L'initiative européenne d'un milliard d'euros dénommée Quantum flagship a été lancée en 2018, et est conçue et gérée en lien étroit avec la communauté académique européenne⁴⁷. Le dernier-né des plans nationaux est le plan

46 <https://www.quireca.com/overview-on-quantum-initiatives-worldwide/>

47 M. Riedel et al. Europe's Quantum Flagship initiative, *Quantum Science and Technology* 4, 020501 (2019).

français, annoncé par Emmanuel Macron le 21 janvier 2021. Sur le budget de 1,8 milliard d'euros sur cinq ans (1,05 milliard de fonds publics nationaux, 550 millions de fonds privés et 200 millions de fonds européens), 780 millions sont consacrés au calcul quantique⁴⁸.

Il est difficile de chiffrer les investissements des grands groupes tels que Google, IBM, ou, plus récemment, AWS, Alibaba ou Baidu, mais, au vu de leurs accomplissements récents (machines de plusieurs dizaines de qubits, recrutement de chercheurs de renom, centres de recherche entièrement dédiés au calcul quantique...), il est certain qu'une part significative de leur budget de R&D annuel multimilliardaire est consacré à l'informatique quantique.

Enfin, les start-ups du domaine se financent également auprès d'investisseurs privés⁴⁹, via des fonds d'investissement spécialisés tels le français Quantonation et des levées de fonds. On peut citer par exemple IQM (levée de 39 millions d'euros en novembre 2020), Rigetti (79 millions en août 2020), ou PsiQuantum (150 millions en avril 2020). IonQ est devenu en mars 2021 le premier *pure player* de l'informatique quantique à être introduit en bourse⁵⁰.

CONCLUSION

Le calcul quantique constitue un changement profond de paradigme par rapport à l'informatique classique. Il ne remet pas en question la notion de calculabilité, mais il pourrait bouleverser la complexité algorithmique. Les ordinateurs quantiques pourraient nous permettre de résoudre certains problèmes de manière considérablement plus rapide que les meilleurs ordinateurs classiques, et ainsi de répondre à des questions hors de la portée du calcul classique.

Cependant, les réalisations de processeurs quantiques sont aujourd'hui embryonnaires. Diverses architectures et systèmes physiques sont à l'étude et forment de bons candidats à la construction de calculateurs quantiques universels de grande taille, mais des obstacles

48 K. Poireault. Le plan quantique français est enfin lancé, *Industrie & Technologies* <https://www.industrie-techno.com/article/le-plan-quantique-francais-est-enfin-lance-tous-les-details-de-cette-strategie-a-1-8-milliard-d-euros-sur-5-ans.63594> (publié : 21 janvier 2021, consulté : avril 2021).

49 E. Gibney. Quantum gold rush: the private funding pouring into quantum start-ups, *Nature* **574**, 22-24 (2019).

50 P. Smith-Goodson. IonQ Takes Quantum Computing Public With A \$2 Billion Deal, *Forbes*, <https://www.forbes.com/sites/moorinsights/2021/03/23/ionq-takes-quantum-computing-public-with-a-2-billion-deal/?sh=58d533d85d06> (publié : 23 mars 2021 ; consulté : avril 2021).

technologiques et théoriques de taille demeurent et de telles machines, si elles voient le jour, ne seront pas concrétisées avant de nombreuses années...

En revanche, des processeurs imparfaits de quelques centaines de qubits seront disponibles dans les années à venir, et donneront un avantage quantique pour résoudre certains problèmes. Il est possible que ces problèmes aient des applications, dans des domaines aussi variés que la chimie, la science des matériaux, la gestion des flux ou la finance. Ces applications pourraient assurer une rentabilité à court terme et éventuellement servir de fondement à de nouvelles perspectives par un cercle vertueux investissements — bénéfiques analogue à celui observé dans l'industrie des semi-conducteurs⁵¹.

On assiste actuellement à une créativité foisonnante, à l'émergence de nombreuses start-up et à un fort investissement tant public que de grandes entreprises privées. On peut se réjouir de cette richesse d'activité induite par l'indéniable attractivité du sujet. Il faut cependant être attentif à ce que l'on n'assiste pas à un monopole déraisonnable sur les moyens financiers et l'embauche de jeunes talents.

En tout état de cause, même si le futur industriel de ce domaine n'est certainement pas assuré et qu'il est possible qu'au boom actuel succède un hiver semblable à ceux qu'a connus l'intelligence artificielle⁵², la recherche académique devrait perdurer grâce au financement public, tant le domaine est, indépendamment de ses applications possibles, scientifiquement fécond.

Remerciements. L'auteur remercie Flavio Baccari et Ivan Šupić pour leurs commentaires pertinents.

Déclaration d'intérêts. L'auteur a écrit ce texte alors qu'il était chargé d'études pour l'Académie des technologies. Il est désormais chargé de recherche chez Quandela.

51 US National Academies of Sciences, Engineering, and Medicine, *ibid*.

52 J. Horgan. Will Quantum Computing Ever Live Up to Its Hype? *Scientific American*, <https://www.scientificamerican.com/article/will-quantum-computing-ever-live-up-to-its-hype/> (publié : 20 avril 2021, consulté : avril 2021).

CONCLUSION

***V*ERS UNE SPÉCIALISATION DES DISPOSITIFS MATÉRIELS DE TRAITEMENT NUMÉRIQUE**

La loi de Moore, qui a rythmé pendant de nombreuses années et qui rythme encore l'amélioration de performances des systèmes numériques, a consacré d'une certaine manière un principe d'universalité des dispositifs de traitement de l'information numérisée. À programme constant et quel que soit le programme considéré, sa durée d'exécution pouvait être réduite en augmentant la fréquence d'horloge du processeur, ou en améliorant l'architecture des processeurs par des mécanismes génériques (« *lookahead processor* », processeurs VLIW...)

Le premier chapitre montre que ce principe est en voie d'être battu en brèche. L'intégration au sein d'un même circuit (Soc) ou de plusieurs circuits au sein d'un même packaging (Sip), de diverses fonctions d'un système donné va être à même de fournir des accélérations de performances significatives avec une consommation énergétique réduite. Il en est de même avec différents modèles d'intégration de nombreux cœurs de processeurs dans un même circuit selon le type de parallélisme souhaité, ou encore la spécialisation d'un circuit pour l'exécution exclusive d'une opération complète comme celle de l'inférence neuronale en apprentissage automatique.

Le calcul quantique est souvent vu comme un mécanisme de traitement générique appelé à fournir des améliorations de performances qui dépassent même celles promises par la Loi de Moore. Boris Bourdoncle dans son texte met en doute l'avènement, à court ou à moyen terme, d'un tel dispositif universel. Par contre il souligne qu'il est raisonnable de penser que des coprocesseurs quantiques spécialisés verront le jour dans les années à venir et seront intégrés dans des architectures d'ordinateurs ou de supercalculateurs plus

classiques pour résoudre certains problèmes d'optimisation et de simulation.

L'usage de polymères pour miniaturiser l'archivage de très grandes quantités d'information avec un coût énergétique réduit est séduisant. Toutefois la vitesse d'écriture sur de tels supports va rester limitée. Néanmoins, comme le souligne François Kepes, il sera rapidement possible d'utiliser ce type de technologie dans le cadre de certaines applications comme l'archivage dit « à froid » de très gros volumes de données patrimoniales.

Une des conséquences de l'utilisation à grande échelle de dispositifs génériques comme les processeurs à vocation universelle pour calculer a été de provoquer au niveau mondial une très grande concentration du marché qui s'est vu limité à quelques acteurs seulement, qu'il s'agisse de la conception de microprocesseurs ou de leur fabrication.

Un autre effet a été de pouvoir dissocier totalement le développement de logiciels, de l'architecture matérielle censée les exécuter. Il est clair que cette dissociation a été bénéfique pour les utilisateurs et la protection de leur patrimoine applicatif. Elle a aussi permis d'augmenter de manière très significative la productivité du logiciel et démultiplier ainsi la quantité d'applications du numérique et conduire notamment à la création d'une industrie des logiciels et des services florissante.

En matière d'acquisition de compétences dans le numérique, cette dissociation a, dans de nombreux cas, simplifié la formation en la limitant le plus souvent au logiciel. De la même manière, la recherche dans le domaine des logiciels et celle dans celui du matériel se sont poursuivies, pour l'essentiel, indépendamment l'une de l'autre.

Soulignons par ailleurs que l'universalité des dispositifs matériels de traitement a eu tendance à augmenter le besoin en énergie nécessaire à leur fonctionnement.

La tendance vers une spécialisation des dispositifs de calcul et de stockage va bien sûr entraîner une transformation profonde du marché mondial, une rupture dans la répartition des rôles entre les différents acteurs ainsi qu'une remise en cause des modèles économiques actuels. Cette spécialisation impliquera en outre l'obsolescence d'une partie du parc logiciel existant.

Sans entrer ici dans le détail de ces transformations, soulignons qu'un des effets possibles pour les utilisateurs est d'avoir à élaborer pour eux-mêmes des dispositifs de type super-calculateur dédié, leur permettant d'acquérir des éléments de différenciation majeurs par rapport à leurs concurrents. Une autre possibilité pourrait être que ces utilisateurs conçoivent

CONCLUSION

eux-mêmes les circuits intégrés nécessaires aux objets connectés utiles à leur activité.

Dans tous les cas la nécessaire imbrication du logiciel et du matériel pour obtenir des systèmes numériques performants va créer de nombreuses opportunités d'innovation à la portée de l'Europe. Il importe toutefois que les entreprises utilisatrices ou productrices de technologies ou services numériques, ainsi que les États prennent en compte rapidement et avec toute l'ampleur nécessaire les changements de paradigmes évoqués dans ce rapport.

« Le développement de dispositifs de traitement numérique spécifiques aux applications constitue une voie prometteuse pour pallier le ralentissement de la miniaturisation des circuits intégrés tout en procurant à l'Europe de nouvelles opportunités d'innovation et d'indépendance ».

C'est la réponse des experts de l'Académie des Technologies à des questions de réalisabilité industrielle à moyen terme, de trois types de technologies matérielles : les circuits intégrés sur Silicium, le calcul quantique, l'archivage de données sur polymères de type ADN.

L'abandon de la polyvalence des dispositifs numériques matériels pourrait alors conduire à rompre le principe d'indépendance matériel/logiciel sur lequel s'est fondé l'énorme développement du numérique au cours des 50 dernières années.

Académie des technologies
Le Ponant – Bâtiment A
19, rue Leblanc
75015 PARIS
+33(0)1 53 85 44 44
secretariat@academie-technologies.fr
www.academie-technologies.fr

©Académie des technologies 2022
ISBN : 979-10-97579-38-8

