

LOGIQUE & CALCUL

Au-delà de la loi de Moore ?

La loi de Moore est un cas de loi exponentielle empirique. Certains pensent que des lois de ce type déterminent notre avenir et sont une clef pour comprendre le passé.

Jean-Paul DELAHAYE

La formulation la plus courante de la loi de Moore indique que la puissance des dispositifs informatiques (pour un coût donné) double tous les 18 mois. Elle est assez bien vérifiée depuis plus de 40 ans, mais on annonce aujourd'hui que sa fin est proche. Depuis sa formulation dans les années 1960, il y a toujours eu des sceptiques pour mettre en doute sa validité. A-t-on aujourd'hui de meilleures raisons de penser qu'elle s'épuise et que nous vivons la fin d'une fantastique période qui a vu augmenter d'un facteur un million au moins les performances des technologies numériques (capacité de stockage, puissance de calcul, vitesse de transmission) ?

Les conséquences économiques d'une rupture de la loi seraient graves, tant l'industrie électronique est habituée à ce taux d'accroissement qui lui sert de base pour programmer son développement et décider de ses investissements, au point qu'on a dit de la loi de Moore qu'elle était une prophétie autoréalisatrice. Un ralentissement du progrès perturberait gravement le secteur, en faisant baisser le taux de renouvellement des PC et autres matériels électroniques – ce qui a peut-être déjà commencé.

On affirme aussi que la particularité de l'économie des ordinateurs n'est pas dans l'existence d'une telle loi exponentielle qui en décrit la dynamique, mais seulement dans le temps de doublement (18 mois) très court que l'on y observe. Une étude portant sur des dizaines de technologies montre que des variantes de la loi de Moore y ont

cours, avec des taux de croissance plus bas, mais des dynamiques exponentielles similaires. Il y aurait donc une loi générale affirmant l'existence d'une augmentation exponentielle des performances ou, ce qui revient au même, une décroissance exponentielle des coûts. Cette loi concernerait tous les secteurs économiques fondés sur des développements techniques.

Il est capital de mieux comprendre ces versions généralisées de la loi de Moore dont dépend l'avenir de notre monde. À côté des prévisions de science-fiction que certains tirent, une étrange variante de la loi de Moore apparaît à l'œuvre en biologie et concerne cette fois la complexité des êtres vivants. Même si elle est très discutée, les arguments présentés par Alexei Sharov en sa faveur sont assez soigneusement formulés. Ils conduisent à soutenir l'idée d'une croissance de type exponentiel de la complexité biologique et suggèrent une origine extraterrestre de la vie sur la Terre, idée ancienne déjà défendue par Francis Crick, l'un des découvreurs de la structure de l'ADN. Nous y reviendrons, mais auparavant, voyons ce que dit la loi de Moore et quelle a été sa validité passée.

En 1958, l'Américain Jack Kilby invente, ce qui lui vaudra de partager le prix Nobel de physique en 2000, le premier circuit intégré (on dit aussi puce électronique), qui est rapidement mis en production. Il s'agit d'un dispositif où sont reliés, sur une plaque, plusieurs composants élémentaires, essentiellement des transistors.

Le nombre de transistors qu'on réussit à placer sur une puce électronique augmente

avec régularité dès son invention. C'est en observant cette variation rapide et obstinée que Gordon Moore, l'un des fondateurs de la société *Intel*, formule l'énoncé initial de sa loi.

Diverses formulations

En 1965, il affirme qu'il y a chaque année un doublement de densité des transistors sur les puces et qu'il se prolongera au moins dix ans (voir la figure 1). En 1975, il ajuste sa loi et prévoit alors un doublement de densité tous les deux ans seulement, lui aussi annoncé comme devant se poursuivre une décennie. Traduite pour parler non plus de la densité des transistors sur une puce, dont l'utilisateur se moque un peu, la loi de Moore sous une forme plus vague, mais plus générale et donc plus intéressante, est l'affirmation suivante :

Tous les 18 mois, pour un prix donné, la puissance des dispositifs informatiques double.

La loi concerne aussi bien la capacité de stockage que la puissance de calcul ou la vitesse de transmission de l'information numérique. Puisqu'il y a 60 mois en cinq ans, le petit calcul $2^{60/18} = 10,08$ montre que cette affirmation générale équivaut à l'affirmation plus frappante encore que les performances informatiques sont multipliées par dix tous les cinq ans, et donc par 100 tous les dix ans, et par un million en 30 ans. Ces chiffres étonnants semblent irréels.

Considérons d'abord le nombre de transistors présents dans un microprocesseur. Nous sommes passés de 2 300 transistors en 1971 pour la puce *Intel-4004* à

1. Le premier énoncé de la loi de Moore

Dans son article de 1965, Gordon Moore étudie l'évolution du coût par composant des puces dans la décennie qui précède. Il trace en particulier une courbe pour l'année 1962 et une autre pour l'année 1965 du coût par transistor en fonction du nombre de transistors sur une puce. Ce coût atteint un minimum pour 15 transistors en 1962, et pour 70 en 1965 (a). Par extrapolation, il propose le dessin de la courbe de coût qu'on peut espérer pour 1970. Il dessine aussi

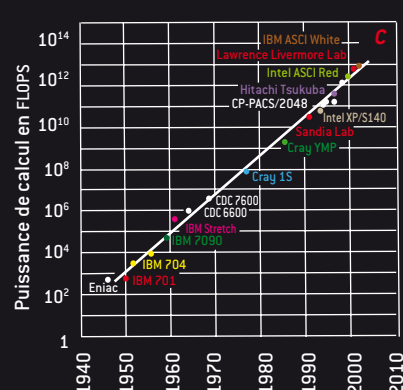
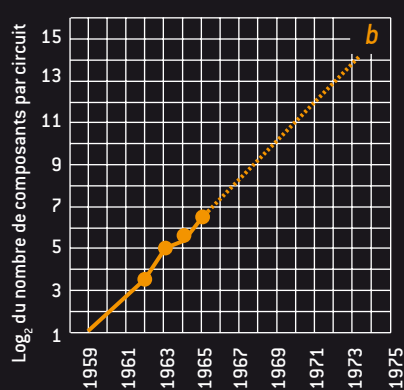
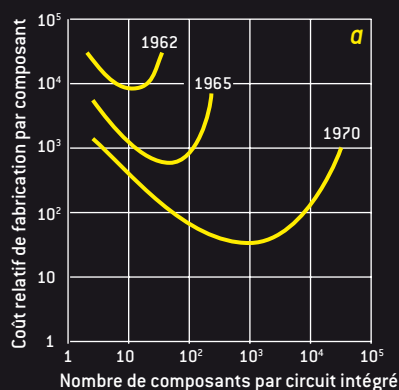
la courbe pour le nombre de composants par circuit, courbe qu'il prolonge en pointillés (b).

Dans son article, il écrit alors le premier énoncé de la loi de Moore : « La complexité pour le coût minimum s'est accrue d'environ un facteur 2 par an. À court terme, on a certainement raison d'espérer que cette vitesse se maintiendra et peut-être même augmentera. À plus long terme, ce rythme d'accroissement est moins certain, quoiqu'il n'y ait pas de raison de croire qu'il ne res-

tera pas à peu près constant pour au moins dix ans. D'ici 1975, le nombre de composants par circuit intégré de coût minimum sera de 65 000. »

C'était la naissance de la loi de Moore qui, dix ans plus tard, sera reformulée par G. Moore pour n'envisager qu'un doublement tous les deux ans. La loi est souvent présentée sous une forme moins précise qui annonce le doublement des performances informatiques tous les 18 mois. En voici un exemple. La courbe (c) représente la

performance maximale des ordinateurs les plus puissants entre 1950 et 2000 : durant ces 50 ans, la loi d'un doublement de puissance tous les 18 mois a été vérifiée. Nous sommes passés d'un peu moins de 1 000 opérations en virgule flottante par seconde (FLOPS) à 10^{13} FLOPS, soit un gain d'un facteur 10^{10} en 50 ans et donc un gain moyen d'un facteur 2 tous les 18 mois. Ces dix dernières années, un facteur supérieur à 100 a de nouveau été obtenu (non représenté sur la courbe).



Courbes c : d'après J. Koomey et al., IEEE Annals of the History of Computing, vol. 33(3), pp. 46-54, 2011

2 600 000 000 transistors pour le 10th core Intel Xeon Processor E7 en 2013. Durant cette période de 42 ans, la miniaturisation électronique a donc progressé d'un facteur 1 130 000. Cela correspond à une multiplication par 1,942 tous les deux ans, conforme, à 3 % près, à la prévision de Moore de 1975 qui est donc vraie non pas sur 10 ans, mais sur 40. Notons que cette augmentation du nombre de transistors entraîne un progrès supérieur pour la puissance de calcul, car la vitesse d'horloge des microprocesseurs a elle aussi été augmentée.

Considérons un autre dispositif important en informatique, mais très différent des puces : les disques durs magnétiques qui servent de mémoire de masse à la plupart des ordinateurs. Prenons comme repère l'énorme disque dur IBM-2310 de 1964 qui avait une capacité de un méga-octet, soit un million de paquets de huit bits d'information ou un million de caractères, ou encore un

assez gros livre. Les disques d'aujourd'hui ont une capacité qui atteint et dépasse un téraoctet : 1 000 milliards de paquets de huit bits d'information, le texte d'un million de livres !

Un facteur un million en 40 ans

Il y a donc eu un progrès d'un million sur une période de 40 ans, soit exactement un doublement tous les deux ans. Comme le prix a considérablement baissé entre les deux repères envisagés, on arrive au doublement de performance tous les 18 mois annoncé par la loi de Moore générale. Sur la base de 10 000 dollars pour le prix du disque dur de 1964, contre 100 dollars pour un disque dur d'un téraoctet aujourd'hui, on arrive à un gain de 100 millions sur 40 ans, soit une performance multipliée par $1,995 \approx 2$ tous les 18 mois.

La loi de Moore générale affirmant un doublement de la performance tous les 18 mois ou tous les deux ans n'est pas métronomique : d'un secteur à l'autre, d'une période à l'autre, les variations sont parfois importantes. La comparaison des mémoires flash (utilisées dans les clefs USB, les téléphones portables et les tablettes) et des tailles d'écran (mesurées en nombres de pixels) illustre l'irrégularité de la loi.

De 2004 à 2012, la densité des circuits des mémoires flash a doublé tous les ans, dépassant largement les prévisions de la loi de Moore de 1975, et c'est à ce très rapide progrès qu'on doit les remarquables capacités de stockage des smartphones ou des tablettes et la fluidité de leur affichage. Concernant le nombre de pixels d'un écran couleur d'ordinateur portable, en revanche, on observe une évolution qui va de 640×400 pixels pour un ordinateur portable à écran couleur en 1993 (par exemple

le *Mac Power Book 165c*) à $2\,880 \times 1\,800$ pixels pour les meilleurs écrans couleurs de 2013 (par exemple le *MacBook Pro* écran *Retina* 15 pouces). Le gain est d'un facteur 20 en nombre de pixels en 20 ans, soit un facteur 1,25 tous les 18 mois : pour les écrans, nous sommes loin des chiffres annoncés par la loi générale.

Consommation d'énergie et miniaturisation

La raison principale (mais pas la seule) de ces avancées est que d'une génération de puces à la suivante, la taille des dessins sur les microprocesseurs a été réduite de 30 %, conduisant à une réduction d'occupation de surface de 50 % pour un transistor (car $0,7 \times 0,7 \approx 0,5$). Cela a permis un accroissement de vitesse (la fréquence) de 40 %. De manière répétée pendant quatre décennies, pour une surface donnée, on a ainsi réussi l'in vraisemblable doublement du nombre de transistors et l'accélération des calculs. Malheureusement, tous les autres paramètres en jeu n'ont pas évolué de la même façon. En particulier, la consommation d'énergie et la dissipation se sont petit à petit accrues, créant à la longue un grave problème. Les temps d'accès à la mémoire de masse (les disques durs) n'ont pas toujours bien suivi, ce qui a limité la progression de certaines performances.

Pour lutter contre les problèmes de consommation d'énergie et de dissipation

de chaleur, une méthode est d'utiliser des puces à plusieurs cœurs. La multiplication de ces sous-unités du processeur travaillant simultanément en communiquant peu limite les échanges sur de longues distances dans le microprocesseur et donc l'échauffement électrique. Il existe des puces à dix cœurs et on envisage d'aller plus loin, au point qu'on évoque une formulation nouvelle de la loi de Moore annonçant un doublement du nombre de cœurs tous les deux ans. Malheureusement, cette stratégie ne donne un accroissement de puissance que pour le traitement en parallèle de plusieurs tâches, sans vraiment augmenter la vitesse de chacune. De ce fait, la vitesse de traitement des PC vendus ces trois dernières années s'est accrue non plus de 40 % par an en moyenne comme ce fut longtemps le cas, mais seulement de 10 à 20 % selon les applications.

D'après certains analystes, il s'agit des premiers effets de la rencontre avec le mur des réalités physiques qui va mettre fin à la loi de Moore. Cette baisse de puissance serait d'ailleurs l'explication principale de la baisse sensible de la vente des PC constatée depuis peu. Pouvant garder plus longtemps leurs machines qui ne subissent plus un processus d'obsolescence aussi rapide qu'auparavant, les consommateurs les renouvellent moins... d'autant qu'ils préfèrent parfois s'acheter des téléphones ou des tablettes dont les marchés cannibalisent celui des PC, aggravant le phénomène et risquant à terme de déstabiliser une partie

importante du secteur de l'électronique et de l'informatique.

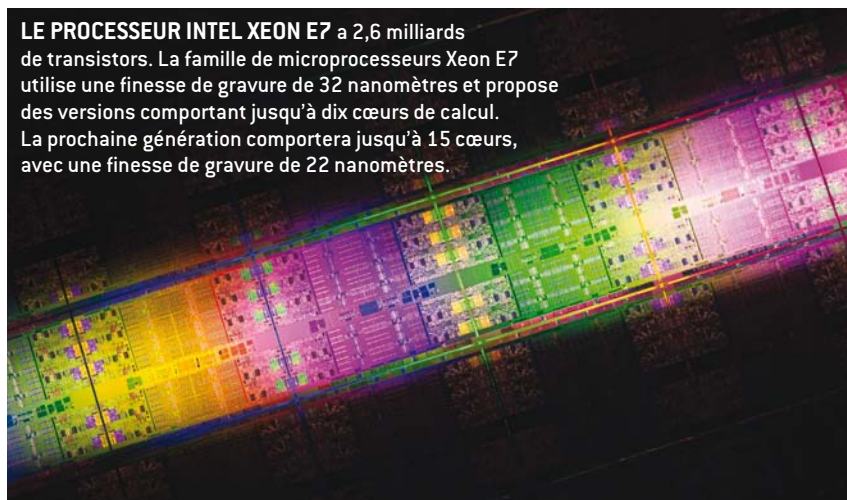
Nul ne conteste que la loi de Moore finira par se heurter à une barrière physique. Sans évoquer les limites extrêmes que les théoriciens de la mécanique quantique calculent, il est peu probable qu'on puisse stocker plus d'un bit d'information par atome de matière. Des limites économiques sont aussi à craindre : le coût des investissements nécessaires pour maintenir le progrès augmente exponentiellement (loi de Rock) et nous conduira à un point impossible à dépasser puisque l'économie ne peut consacrer la totalité de ses investissements à la conception et à la production de la prochaine génération de puces électroniques. La question est : sommes-nous proches du moment où la loi de Moore cessera d'être vraie et où il faudra affronter toutes les conséquences de l'arrêt de ce stimulant économique ?

Dès 2005, G. Moore pronostiquait que la fin était en vue. « Cela ne peut durer indéfiniment. Le propre d'une exponentielle est de vous entraîner hors des limites raisonnables jusqu'à un éventuel désastre. La taille d'un transistor approche celle d'un atome, qui est une barrière fondamentale. Nous l'atteindrons d'ici deux ou trois générations [de microprocesseurs]. »

Murs infranchissables

La réalisation d'un transistor à un seul atome en 2012 par une équipe de l'Université des Nouvelles-Galles-du-Sud, en Australie, nous place devant le mur. Même si l'exploit de laboratoire ne concernera pas nos machines avant quelques années, il fixe un point indépassable des techniques de gravure microélectronique.

Dès 2012, le physicien américain Michio Kaku formulait un pronostic définitif : d'ici dix ans, il en sera fini de la loi de Moore ; d'ici 2022 au plus tard, les progrès de l'industrie du silicium auront pris fin. L'utilisation de transistors en trois dimensions (depuis peu proposés par Intel) et bientôt peut-être de puces elles-mêmes en trois dimensions permettra sans doute de maintenir un certain rythme de progrès pendant



LE PROCESSEUR INTEL XEON E7 a 2,6 milliards de transistors. La famille de microprocesseurs Xeon E7 utilise une finesse de gravure de 32 nanomètres et propose des versions comportant jusqu'à dix cœurs de calcul. La prochaine génération comportera jusqu'à 15 cœurs, avec une finesse de gravure de 22 nanomètres.

2. L'importance de l'énergie : la loi de Koomey

Une loi plus précise que la loi de Moore a été énoncée dans un article récent de Jonathan Koomey, de l'Université Stanford, et ses collègues (*IEEE Annals of the History of Computing*, vol. 33(3), pp. 46-54, 2011).

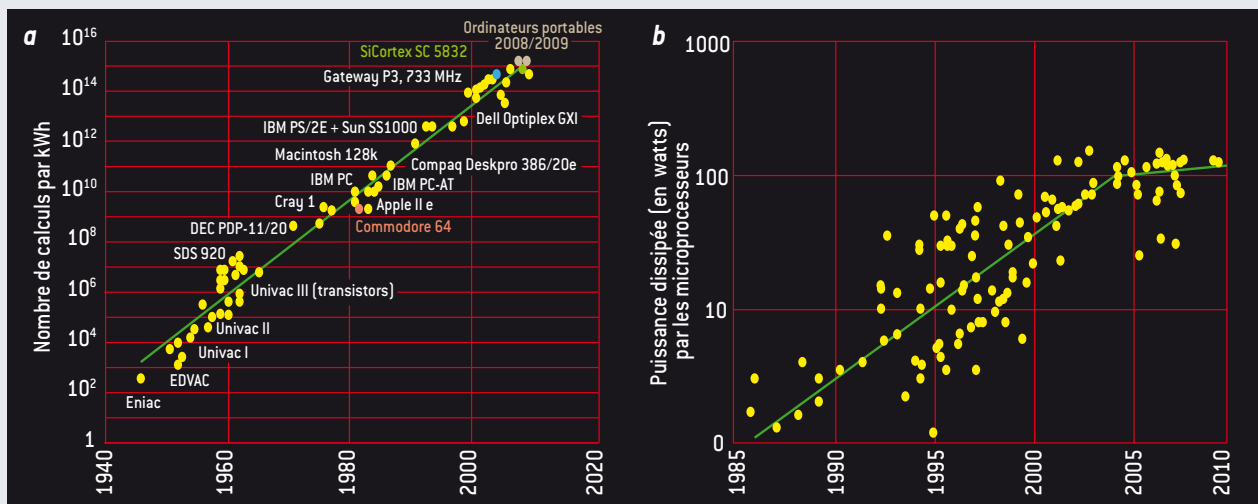
La loi de Koomey se réfère au nombre de calculs faits par unité d'énergie dissipée, nombre qui double à peu près tous les 1,57 ans depuis 1950 (a). Cet énoncé centré sur l'énergie et donc aussi sur l'échauffement produit par

les calculs est une reformulation pertinente, puisque plus régulière et précise, et puisqu'une part importante des problèmes de l'industrie des microprocesseurs est liée à des questions énergétiques. La vitesse de calcul des microprocesseurs est entravée par l'échauffement dû à l'effet Joule.

Par ailleurs, la commodité des dispositifs informatiques mobiles (ordinateurs portables, téléphones, tablettes) dépend des batteries, qui re-

présentent une part importante de leur prix et dont la capacité est toujours jugée insuffisante par les utilisateurs. Enfin, les microprocesseurs sont toujours plus gourmands en énergie (b). Les gros centres de calcul dénommés *datacenters* (comme ceux des firmes *Google*, *Facebook*, etc.) sont confrontés à des problèmes de consommation énergétique si importants qu'on choisit de les installer dans des régions froides ou bien là

où l'énergie est moins chère. Selon l'Université Stanford, la consommation des *datacenters* a augmenté de 56 % de 2005 à 2010 et elle représente aujourd'hui près de 1,5 % de la consommation mondiale d'électricité, soit l'équivalent de 30 centrales électronucléaires. Aux États-Unis, les *datacenters* consommeraient autant d'énergie que l'industrie automobile, d'après un article du *New York Times* (J. Glanz, 22 septembre 2012.)



quelques années. Cependant, dit-il, « si je devais engager un pari, je dirais qu'après, nous devons passer aux ordinateurs moléculaires et, peut-être beaucoup plus loin dans le cours du XXI^e siècle, aux ordinateurs quantiques ». Or il s'agit de technologies qui ne semblent pouvoir aujourd'hui assurer la relève de l'électronique à silicium sur laquelle, depuis un demi-siècle, tous nos efforts ont été concentrés, sans vraiment préparer d'alternatives.

Linus Torvalds, l'acteur principal du développement de *Linux*, pense lui aussi que nous allons rapidement nous retrouver dans une situation totalement nouvelle en informatique quand la loi de Moore cessera de se vérifier et qu'alors nous devons apprendre à vivre autrement.

John Gustafson, architecte chargé de la conception et du dessin des produits de

la société AMD (seconde entreprise mondiale de production de microprocesseurs juste derrière *Intel*) est lui aussi formel et annonce que « vous assistez au début de la fin de la loi de Moore ».

Tout le monde n'est pas d'accord. Certains, comme Hans Moravec (roboticien, professeur à l'Université de Carnegie-Mellon), Raymond Kurzweil (inventeur et écrivain) ou John Smart (philosophe, futurologue et directeur de l'*Acceleration Study Foundation*) défendent qu'il y a toutes sortes de raisons de penser que la loi de Moore sous une forme généralisée restera vraie encore longtemps.

Ils évoquent les nombreuses pistes technologiques permettant de contourner les obstacles, comme déjà tant d'autres ont été contournés. Certes, une transition technologique sera prochainement

nécessaire. Ce qui semblait devoir ne durer que dix ans a duré plus de 40 ans et l'on peut envisager d'introduire des éléments moléculaires, optiques ou quantiques et les combiner et les intégrer aux éléments de silicium. N'y a-t-il pas encore de quoi maintenir le rythme plusieurs décennies ? Pourquoi ne pas croire que ce miracle se prolongera ? Les transitions technologiques passées, que ce soit le passage des calculateurs mécaniques aux calculateurs à relais, ou celui des calculateurs à relais aux calculateurs électroniques, ont à chaque fois accéléré le rythme des gains. Si nous tirons les leçons de l'histoire de l'informatique, nous devons donc non pas craindre la prochaine transition technologique, mais, au contraire, compter sur elle pour qu'elle maintienne, voire accélère, la dynamique du progrès.

3. La croissance de la complexité du vivant

Alexei Sharov a proposé une méthode de mesure de la complexité (entendue comme richesse en fonctions et en organisation) à l'aide de l'évaluation de ce qu'il dénomme l'ADN fonctionnel non répétitif (en gros, les gènes et les séquences promotrices des gènes). Il obtient une courbe qui, en échelle logarithmique, est constituée de points assez bien alignés. Cette courbe, si on la considère comme repré-

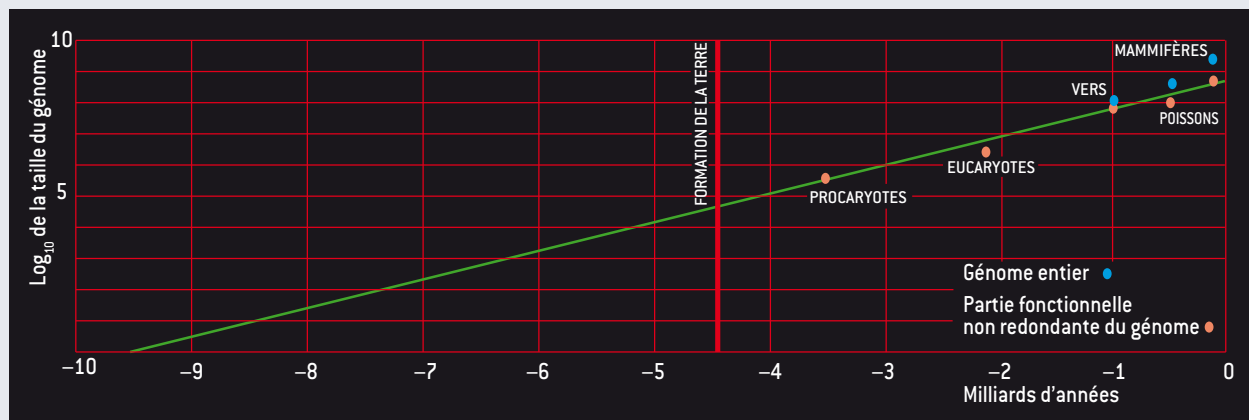
sentative d'une évolution de la complexité du vivant, signifie que celle-ci double tous les 350 millions d'années environ. C'est une sorte de loi de Moore pour la vie.

En prolongeant la courbe vers la gauche, elle touche l'axe des abscisses en un point correspondant environ à -10 milliards d'années, ce qui est plus ancien que l'âge de la Terre. On pourrait en conclure qu'il ne faut pas

prolonger la courbe et qu'il y a sans doute eu une phase où la croissance de la complexité a été plus rapide. A. Sharov préfère une autre explication de ce point d'intersection, selon laquelle la vie est arrivée sur la Terre avec une certaine complexité après son apparition ailleurs. Pour lui, le transport d'un point à l'autre de l'espace n'est pas le fait d'extraterrestres qui auraient ensemencé volontairement

la Terre (théorie de la panspermie dirigée), mais le résultat de déplacements – rares, mais possibles – de matière portant les structures moléculaires propres à la vie, par exemple sous la forme de spores résistantes.

Inutile de dire que les idées de A. Sharov ne font pas l'unanimité et que les spécialistes de l'origine de la vie y voient au contraire de nombreuses objections.



L'idée que l'avenir nous réserve une accélération du progrès produisant des doublages de plus en plus rapprochés des diverses mesures de performance de nos dispositifs numériques nous plonge dans des visions de science-fiction. On évoque la naissance de machines super-intelligentes contribuant elles-mêmes à la conception de nouvelles machines encore plus puissantes, etc.

De la science-fiction ?

Une volumineuse littérature futuriste s'est développée qui se plaît à discuter une loi « des retours accélérés » (due à R. Kurzweil), sorte de généralisation un peu imprécise de la loi de Moore à tous les domaines. Elle conduirait à un instant nommé « la singularité » (idée introduite par Vernor Vinge, professeur d'Informatique à l'Université de San Diego) et pronostiqué avant la fin du XXI^e siècle, où l'intelligence humaine aura engendré des machines qui la dépasseront.

Heureusement (?), selon certains, nous pourrions alors transférer, télécharger, nos esprits dans ces machines, ce qui nous permettrait d'accéder à une forme d'immortalité. Une question vient naturellement à l'esprit : pourquoi ces machines nous accepteraient-elles en leur sein si nous leur sommes inférieurs et que nous ne servons plus à rien ? Bien d'autres questions sont abordées et débattues par ceux qui se nomment eux-mêmes « transhumanistes » ou « extropiens ».

Une idée développée par Hugo de Garis (chercheur en intelligence artificielle) évoque la survenue prochaine d'une inévitable et terrible guerre entre ceux qui s'opposeraient aux développements de cette intelligence nouvelle, qui condamne à terme les êtres humains, et ceux qui y verraient simplement la continuation de l'humanité sous d'autres formes matérielles. La question de savoir si les cyborgs, hybrides d'humains et de machines, seraient en mesure d'ouvrir une troisième voie est aussi posée et donne lieu

à de rudes controverses entre ces futurologues, parfois qualifiés de technoprophètes.

Parmi les pistes de réflexion, John Smart défend une idée inattendue. Non seulement les performances technologiques en général continueront de progresser, mais elles s'accompagneront d'une réduction des tailles, qui finira à long terme par rendre si petits les lieux où la vie se poursuivra dans ses formes nouvelles qu'elle disparaîtra du monde physique macroscopique. Cette idée résoudrait le paradoxe de Fermi : « Nous ne voyons pas les civilisations intelligentes qui devraient être apparues avant la nôtre et dont l'existence est pourtant probable, puisque la nôtre est assez récente en regard des durées astronomiques. »

Il existe aussi des spécialistes de l'histoire de l'économie et des techniques qui contestent que le progrès va en s'accélération. Ils tentent de montrer que nous sommes en train d'observer une baisse du rythme de l'évolution technique. Les difficultés rencontrées par l'industrie des microprocesseurs