



© Vlasja Kobelew/shutterstock

ci ont connu un essor important au cours des dernières décennies. Ils représentent aujourd'hui l'essentiel des supports de mémoire. Durant ces 20 dernières années, les performances de ces dispositifs ont aussi encouragé la production massive de données (voir la figure 1). La mise en mémoire étant devenue simple et peu chère, il n'y avait plus besoin de limiter le volume de données à stocker.

Les mémoires universelles : utopiques ?

Si l'accroissement du volume d'informations à stocker se poursuit au même rythme, les mémoires d'aujourd'hui seront-elles à la hauteur, en capacité de stockage comme en rapidité d'accès ? Il est probable que non. Les ingénieurs rêvent de mémoires dites universelles qui seraient notamment rapides d'accès, capables de stocker de grands volumes de données, fiables et économes en énergie. Malheureusement, un tel dispositif relève plus de l'idéal vers lequel tendre que d'une réalité concrète.

Les systèmes numériques d'aujourd'hui associent toujours plusieurs dispositifs : des mémoires dites de travail (ou « mémoires vives »), qui conservent de petits volumes de données mais qui ont des temps d'accès très courts, et des mémoires de stockage, qui peuvent enregistrer d'énormes quantités d'information, mais au prix d'un temps d'accès plus long. En associant ces deux types de mémoires, il est possible de construire un système qui peut exécuter des opérations en peu de temps et conserver un grand volume de données.

De nombreux concepts sont à l'étude et certains pourraient être opérationnels dans les années à venir. Pour comprendre l'intérêt de ces dispositifs, il faut commencer par s'intéresser aux mémoires utilisées aujourd'hui, avec leurs qualités et leurs limites. Je présenterai ensuite quelques solutions techniques qui pourraient succéder aux dispositifs actuels d'ici cinq à dix ans.

Comme je l'ai évoqué, nous ne savons pas fabriquer de mémoires universelles. La nature ne semble pas faire mieux dans ce domaine. Il existe, par exemple, différents modèles théoriques de la mémoire du cer-

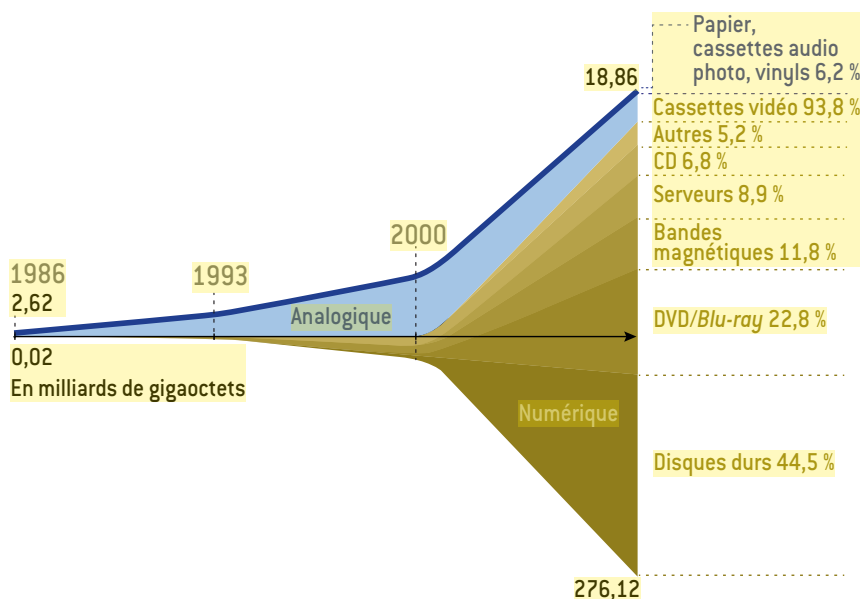
veau humain. Dans tous les cas, la mémoire est subdivisée en sous-systèmes spécialisés en différentes tâches. On observe ainsi de nombreuses analogies entre la mémoire biologique et les mémoires artificielles. Dans un modèle classique des années 1960, on distingue la mémoire sensorielle qui retient une grande quantité d'informations provenant des différents sens pendant un temps très bref et sans faire d'interprétation. Certaines informations sont ensuite transmises à la mémoire à court terme, dont la capacité est limitée et où les informations sont disponibles pendant plusieurs dizaines de secondes. Cette mémoire permet, entre autres, à un lecteur de ne pas oublier le début de la phrase pour pouvoir en comprendre le sens.

Enfin, une mémoire à long terme permet de conserver des informations longtemps et en quantités importantes. L'évolution a probablement conduit à cette structure qui permet par exemple de traiter rapidement une information perçue, comme quand nous sommes confrontés à un prédateur ou à un autre danger.

Depuis les années 1960, ce modèle de la mémoire humaine a subi des modifications et s'est enrichi de sous-catégories (voir l'article de Franck Chaillan dans ce numéro). Cependant, la conclusion est que le cerveau n'est pas une mémoire universelle et que la combinaison de mémoires spécialisées semble présenter des avantages.

L'industrie suit un schéma similaire, puisqu'elle associe différents types de mémoires. Si, à plusieurs reprises, de nouveaux dispositifs ont été présentés comme universels, des études approfondies les ont toujours mis en défaut. Quoi qu'il en soit, les fabricants se concentrent sur l'amélioration de leurs caractéristiques : la durée de vie, la fiabilité, etc. Une propriété importante est le temps d'accès, la durée nécessaire pour la lecture ou l'écriture d'une information sur une mémoire ; elle est comprise entre la nanoseconde et la milliseconde, suivant les dispositifs.

Le processeur d'un ordinateur, qui exécute des calculs, a besoin d'obtenir très rapidement les instructions et les données nécessaires à l'opération. Il doit donc être associé à des mémoires de travail à temps d'accès très court, pour ne pas ralentir les performances du processeur. De telles mémoires coûtent cher, et leur densité d'information stockée est faible : elles ne peuvent pas emmagasiner de grands



1. LES DONNÉES NUMÉRIQUES ont connu un essor spectaculaire. Le développement de supports numériques bon marché et simples d'utilisation marque un tournant au début des années 2000. Le numérique a supplanté les supports analogiques et le volume total de données enregistrées a augmenté rapidement. Cette tendance ne semble pas faiblir.

volumes de données. Pour archiver les informations, il faut donc ajouter au dispositif des mémoires de stockage.

Les deux types de mémoires sont complémentaires et cohabitent dans tous les dispositifs numériques (voir la figure 2) : ordinateurs, tablettes, téléphones portables, clés USB... On peut définir une hiérarchie pyramidale des mémoires. Au sommet se situent les mémoires de travail, aux temps d'accès très courts et aux capacités de stockage faibles. En descendant dans la pyramide, les volumes d'information qu'il est possible de stocker augmentent au détriment des temps d'accès. La base de la pyramide correspond aux mémoires de stockage.

Les mémoires de travail et de stockage se distinguent aussi par leur capacité à conserver l'information sans alimentation électrique. Les mémoires de travail sont dites volatiles : en l'absence d'une source d'énergie électrique, l'information qui y est enregistrée disparaît en quelques dizaines de millisecondes. Ainsi, quand un ordinateur effectue une opération et qu'une panne de courant survient, toutes les informations stockées dans la mémoire de travail sont perdues. Les mémoires de stockage sont, au contraire, non volatiles : l'information reste en mémoire. Si vous sauvegardez des fichiers sur une clé USB ou sur un disque dur et que vous éteignez l'ordinateur, l'information ne sera pas perdue et pourra être réutilisée ultérieurement.

Examinons le fonctionnement des différentes mémoires. Au plus près du processeur, on distingue deux types de mémoires de travail : les mémoires statiques (ou SRAM pour *Static Random Access Memory*) et les mémoires dynamiques (ou DRAM pour *Dynamic Random Access Memory*). Les premières occupent le sommet de la pyramide. Elles sont directement construites sur la puce de silicium du microprocesseur, car elles sont constituées de transistors, comme ce dernier. Leur temps d'accès est de quelques nanosecondes, mais le volume de stockage est très faible.

Les mémoires dynamiques, juste en dessous dans la pyramide des mémoires, ont une capacité de stockage plus importante ; elles sont en général placées sur des puces indépendantes, dont le temps d'accès est compris entre 10 et 100 nanosecondes. Ces mémoires de travail sont utilisées pour effectuer les opérations de calcul des microprocesseurs.

Des mémoires volatiles qui perdent les données

Dans une mémoire dynamique, le bit d'information est enregistré par un système composé d'un transistor et d'un condensateur. Ces composants sont gravés dans le substrat semi-conducteur en silicium de la puce de mémoire. Le transistor agit comme un interrupteur qui autorise, s'il est ouvert, la lecture du bit par un courant électrique. Physiquement, l'état du bit est défini par la charge électrique (l'accumulation d'électrons) du condensateur. Si le condensateur est chargé, le bit vaut 1, s'il est déchargé, le bit vaut 0.

Le problème des condensateurs est qu'ils retiennent mal les électrons, qui fuient dans le silicium environnant. Pour cette raison, la mémoire dynamique est volatile. Il est donc nécessaire de réécrire le bit d'information régulièrement, toutes les quelques dizaines de millisecondes, pour éviter que le condensateur ne se décharge trop et que le bit change d'état. Ainsi, sans alimentation électrique, l'information disparaît de la mémoire.

Dans les premières mémoires dynamiques des années 1960, les transistors et les condensateurs étaient accolés. La miniaturisation a conduit à réduire la taille des composants, mais aussi à changer leur placement relatif. Les condensateurs sont maintenant construits au-dessus des

L'AUTEUR



Luca PERNIOLA dirige le groupe des technologies de mémoires avancées au Laboratoire d'électronique et de technologie de l'information (LETI) au Commissariat à l'énergie atomique et aux énergies alternatives, à Grenoble.