

ENSEIGNEMENT DE PROMOTION SOCIALE

Cours de

STRUCTURE DES ORDINATEURS

- Stockage de l'information -

H. Schyns

Janvier 2004

Sommaire

4. LE STOCKAGE DE L'INFORMATION
--

4.1. Position du problème

4.2. Technologies

- 4.2.1. Le stockage mécanique
- 4.2.2. Le stockage magnétique
- 4.2.3. Les stockage électronique
- 4.2.4. Le stockage optique
- 4.2.5. Les autres techniques

4.3. Permanence de l'information

4.4. Types d'accès

- 4.4.1. L'accès séquentiel
- 4.4.2. L'accès direct
- 4.4.3. L'accès aléatoire
- 4.4.4. L'accès associatif

4.5. Hiérarchie de la mémoire

- 4.5.1. La localisation temporelle.
 - 4.5.2. La localisation spatiale.
 - 4.5.3. L'augmentation du coût spécifique
 - 4.5.4. La pyramide des mémoires
 - 4.5.5. La collaboration dans la pyramide
-

4. Le stockage de l'information

4.1. Position du problème

Revenons un instant à la définition de l'ordinateur :

Ensemble de systèmes électroniques utilisés pour l'acquisition, la conservation, le traitement et la restitution des données.

L'ordinateur doit disposer d'une mémoire puisqu'il a la charge de la conservation et la restitution des données. De plus, nous avons vu que l'ordinateur doit aussi mémoriser les procédures de traitement. Ceci pose le problème de la conception et de l'organisation des systèmes de stockage de l'information dans un ordinateur.

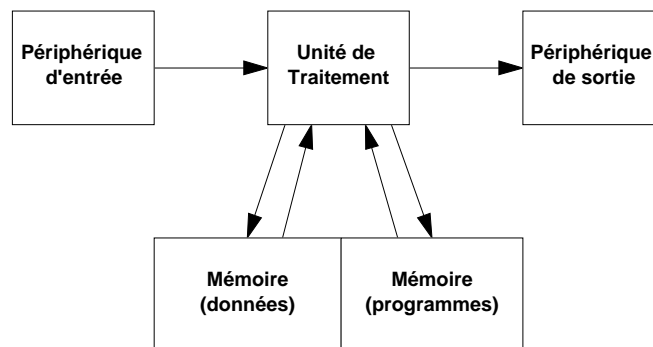


Figure 4.1 L'ordinateur doit mémoriser les données et les procédures de traitement

Un être humain dispose de plusieurs systèmes dans lesquels il stocke l'information. Le choix se fait en fonction des besoins et de la durée de conservation souhaitée. On peut citer, entre autres :

- les livres, les cartes routières, les plans, les cassettes audio ou vidéo, etc.
L'information y est volumineuse, conservée pendant de longues périodes mais sa fréquence d'utilisation est relativement faible;
- les notes, les bouts de papier, les post-it, les agendas
L'information y est succincte, conservée pendant peu de temps mais l'individu y accède fréquemment;
- les neurones de son cerveau, sa mémoire biologique
Le volume d'information est important, la durée de conservation varie en fonction de l'usage qui en est fait mais surtout, les *procédures* qui y sont stockées sont utilisées en permanence (entre autres les processus de reconnaissance visuelle, auditive et sensorielle).

Techniquement parlant, la mémoire idéale - d'un homme ou d'un ordinateur - doit :

- stocker un grand volume d'information,
- être fiable et rapidement accessible,
- prendre peu de place,
- consommer peu d'énergie,
- être bon marché.

Malheureusement, en pratique, une telle mémoire n'existe pas. Il faut donc accepter des compromis. De ce fait, l'ordinateur, à l'image de l'être humain, utilisera plusieurs formes de mémoire en fonction de l'usage requis (¹).

Les différentes formes de mémoire sont caractérisées par :

- la technologie employée,
- le volume d'information mémorisé,
- le temps d'accès à l'information,
- la manière d'accéder à l'information,
- la durée de conservation de l'information,
- le coût de fabrication du support de l'information (coût du matériel),
- le coût d'utilisation du support de l'information (coût de la puissance électrique),
- l'encombrement du support (densité de l'information).

Ces différentes caractéristiques sont autant de critères de choix et de classement des systèmes de mémoires.

4.2. Technologies

Au fil du temps, les constructeurs ont exploité les différentes propriétés de la matière pour stocker l'information. Plusieurs technologies sont toujours présentes au sein d'un ordinateur.

4.2.1. Le stockage mécanique

L'information est stockée dans l'apparence du support. C'est le règne de la **carte** et du **ruban perforé** (*ang.: punched card / tape*) où l'information est codée sous la forme de trous.

Le support est simple à fabriquer, donc économique, relativement stable dans le temps mais fragile à l'usage (cartes déchirées, rubans cassés).

Cette technologie obsolète n'est plus utilisée, sauf dans quelques cas particuliers d'environnement industriel hostile : champs magnétiques intenses, rayonnements, poussières métalliques, ...

4.2.2. Le stockage magnétique

L'information est stockée dans des cristallites magnétiques dont la taille devient de plus en plus petite avec le perfectionnement de la technologie.

La **bande magnétique** (*ang.: tape*), déjà utilisée dans les domaines audio et vidéo, bénéficie d'une technologie bien maîtrisée et est bon marché. La capacité de stockage est importante et la stabilité dans le temps est satisfaisante moyennant quelques précautions. Le principal inconvénient sa lenteur : le temps d'accès à une information donnée peut atteindre plusieurs secondes.

La **disquette** (*ang.: floppy disk*) et le **disque dur** (*ang.: hard disk*), sont des innovations majeures dans l'histoire de l'informatique. Les têtes de lecture/écriture mobiles rendent l'information directement accessible mais exigent des conditions de

1 C'est le dilemme du consommateur qui recherche un produit qui soit à la fois bon, bon marché et moderne :

Good AND Recent = NOT Cheap
Recent AND Cheap = NOT Good
Good AND Cheap = NOT Recent

fonctionnement relativement stables. La principale faiblesse du disque dur est son entraînement mécanique qui est sujet à l'usure. Les disques durs sont dès lors caractérisés par un "temps moyen avant panne" (*ang.: mean time before failure* ou *MTBF*) ce qui signifie que l'utilisateur doit s'attendre à les voir tomber inévitablement en panne au bout "d'un certain temps".

Bien que beaucoup plus rapides que les bandes, les disques restent relativement lents. Le temps d'accès à une information est de l'ordre de la milliseconde. Cette lenteur provient, d'une part, du déplacement mécanique des dispositifs de lecture/écriture et, d'autre part, de l'aspect séquentiel du processus de lecture/écriture qui s'effectue bit par bit.

4.2.3. Les stockage électronique

L'information est stockée au moyen d'une tension aux bornes d'un transistor ou d'une charge électrique dans un condensateur.

Cette technologie permet d'accéder une information quelconque en un temps extrêmement court : une dizaine de nanosecondes pour une mémoire **cache SRAM**, une centaine de nanosecondes pour une **barrette de mémoire DRAM**. La mémoire RAM (*ang.: Random Access Memory* ou *Mémoire à Accès Aléatoire*) stocke un volume d'information relativement important mais elle exige une alimentation électrique permanente. Dès que l'ordinateur est éteint ou redémarré, son contenu est détruit.

4.2.4. Le stockage optique

L'information est stockée sous la forme d'une séquence d'alvéoles microscopiques plus ou moins longues gravées dans un support synthétique ⁽¹⁾. La lecture est assurée par un faisceau lumineux qui est réfléchi ou non par les alvéoles.

Les avantages du **CD-ROM** et du **DVD** sont leur encombrement réduit par rapport au grand volume d'information que l'on peut y stocker. La technologie est fiable, le prix de revient est relativement bon marché et la durée de conservation est de l'ordre du siècle. Par contre le temps d'accès à l'information est plus long que dans le cas du disque dur.

4.2.5. Les autres techniques

Plusieurs technologies faisant appel à la physique de l'état solide sont à l'étude.

Citons par exemple certaines mémoires ferromagnétiques qui utilisent le mouvement d'un atome dans un microcristal ou le sens de rotation (*ang.: spin*) d'un électron sur lui-même pendant qu'il orbite autour du noyau atomique ⁽²⁾. De telles mémoires seraient capables de stocker un grand nombre d'informations, pendant de longues durées et sans apport d'énergie... On peut rêver.

-
- 1 Les puristes diront - avec raison - que la gravure est un procédé mécanique : un CD peut être obtenu par pressage sans intervention quelconque d'un phénomène optique. En toute rigueur, un système purement optique devrait utiliser les propriétés de la lumière : couleur, phase, plan de polarisation.
 - 2 Les électrons tournent sur eux-mêmes pendant qu'ils tournent autour du noyau, à l'image de la Terre qui tourne sur elle-même en 24 h alors qu'elle tourne autour du soleil en 365 jours. Le jour où le soleil se lèvera à l'Ouest nous pourrions dire que le spin de la Terre a changé et est passé de l'état 0 à l'état 1.

4.3. Permanence de l'information

Une mémoire est dite **volatile** quand la conservation de l'information exige un apport continu ou intermittent d'énergie. Cette catégorie comprend la mémoire RAM (SRAM, DRAM et variantes) qui est toujours **volatiles**.

Une mémoire est dite **permanente** ou **non volatile** dans le cas contraire. Cette catégorie comprend les systèmes mécaniques (cartes et rubans), magnétiques (bandes et disques), optiques (CD et DVD) ainsi que certaines mémoires électroniques (ROM, EPROM et variantes, FLASH).

Notez que le processus d'écriture ou de lecture de l'information est *toujours* consommateur d'énergie, qu'il s'agisse d'une mémoire volatile ou non.

Un ordinateur combine toujours ces deux types de mémoire :

- les **disques**, **disquettes** et **CD-ROM** sont des systèmes non volatils capables de conserver une quantité considérable d'information, même quand l'ordinateur est arrêté.
- une **mémoire RAM**, appelée mémoire vive, volatile mais très rapide, dans laquelle l'ordinateur transfère les programmes et données dont il a besoin pur accomplir une tâche.

4.4. Types d'accès

4.4.1. L'accès séquentiel

C'est le cas du stockage sur **bande magnétique**. Le processus est comparable au joueur qui recherche une carte donnée dans un jeu bien mélangé : il a devant lui un tas unique contenant toutes les cartes et ne voit qu'une carte à la fois.

Le processus de recherche commence à une extrémité du stockage et lit tous les enregistrements l'un après l'autre, suivant un ordre prédéfini, jusqu'à ce qu'il trouve celui qui l'intéresse. Pour accéder à un enregistrement donné, il faut donc lire tous ceux qui le précèdent.

Dès lors, le temps d'accès à une information dépend de sa localisation et de la localisation de la lecture précédente. En effet, pour éviter de recommencer chaque fois au début de la bande, le processus repart de l'endroit où il s'était arrêté lors de la recherche précédente. Si le processus atteint la fin de la bande sans avoir trouvé l'information, il la rembobine et recommence au début.

4.4.2. L'accès direct

C'est le cas du stockage sur **disque dur**. Le processus est comparable au joueur qui recherche une carte donnée dans un jeu dans lequel les cartes de même couleur ont été regroupées mais non classées. Il a devant lui quatre petits tas qui correspondent aux quatre couleurs et voit la première carte de chaque tas.

L'information est structurée en blocs et chaque bloc a une adresse unique. Le processus de recherche peut accéder directement au bloc qui contient l'information. Ensuite, il explore ce bloc de manière séquentielle.

Le temps d'accès dépend de la localisation de l'information au sein d'un bloc et de la localisation du bloc qui a été sélectionné lors de la lecture précédente. En effet, si le bloc à rechercher est voisin du bloc trouvé précédemment, les têtes de lecture ne devront presque pas se déplacer. Par contre, si les têtes doivent passer du centre à la périphérie, le temps d'accès sera plus long.

4.4.3. L'accès aléatoire

C'est le cas du stockage en **mémoire RAM**. Pour notre joueur, tout se passe comme si les cartes étaient classées dans l'ordre et étalées sur plusieurs lignes. Chaque ligne contient les cartes d'une même couleur et chaque colonne contient les cartes de même valeur.

Chaque information est localisée très précisément par une adresse de type (ligne, colonne). Il suffit de connaître l'adresse pour accéder directement à l'information désirée.

Le temps d'accès est constant, quelle que soit la localisation de la donnée recherchée et quelle que soit la localisation de la lecture précédente.

4.4.4. L'accès associatif

C'est le cas de la **mémoire cache**. Dans notre comparaison, les cartes sont disposées en lignes et colonnes comme dans le cas précédent mais elles sont mélangées et posées à l'envers. Chaque fois qu'il a tiré une carte, le joueur note sur une petite liste la valeur de la carte ainsi que ses coordonnées. Cette petite liste constitue sa mémoire cache : quand on lui demande de trouver une carte donnée, le joueur consulte d'abord sa liste pour voir s'il s'agit d'une carte dont il connaît déjà l'emplacement.

La mémoire cache n'est pas constituée d'une suite de données mais d'une suite de couples (adresse, donnée). La recherche se fait sur l'adresse pour accéder ensuite à la donnée.

L'accès à une des informations déjà repérée dans le cache est beaucoup plus rapide que l'accès à une autre information. Pour les données du cache, le temps d'accès est indépendant de la localisation. Le cache fonctionne comme une pile FIFO : la donnée la plus récente remplace la plus ancienne. Le temps d'accès aux données récentes (présentes dans le cache) est plus court que le temps d'accès aux données anciennes (vidées du cache).

4.5. Hiérarchie de la mémoire

L'organisation du système de mémoires tient compte de trois principes :

- la localisation temporelle
- la localisation spatiale
- l'augmentation du coût spécifique

4.5.1. La localisation temporelle.

La loi de Pareto s'applique aussi aux logiciels : l'exécution d'un programme passe 80% de son temps dans 20% du code. Ainsi, quand on utilise un traitement de texte, c'est principalement la fonction de saisie des caractères qui est active. La gestion des fontes, des styles ou des options ne sont utilisées qu'occasionnellement.

Dès lors, une donnée ou une instruction utilisée récemment sera probablement *réutilisée* prochainement. Dès lors, l'ordinateur a intérêt à retenir son adresse dans un cache.

4.5.2. La localisation spatiale.

Un logiciel traite rarement *une seule* instruction et *une seule* donnée. En réalité, un programme est une suite d'instructions et les données sont très souvent organisées en tableau.

Par conséquent, Il y a de fortes chances que la prochaine donnée à laquelle le programme accédera soit proche d'une donnée à laquelle il a accédé récemment. En d'autres mots, un programme utilise généralement des données qui sont voisines. Il en va de même pour les instructions des programmes.

Dès lors, au lieu de charger *une* instruction ou *une* donnée, l'ordinateur a intérêt à rechercher directement *une page* d'instructions ou *une plage* de données.

4.5.3. L'augmentation du coût spécifique

Pour augmenter la vitesse des mémoires, on doit faire appel à des technologies de plus en plus poussées. Il en va de même, pour réduire leur encombrement.

Il en découle que les mémoires les plus performantes sont aussi celles qui présentent le coût spécifique le plus élevé. Par coût spécifique on entend le prix de revient du bit stocké. Ainsi, un disque dur de 20 Gb coûte proportionnellement moins cher par bit stocké qu'une barrette de SDRAM de 256 Mb.

4.5.4. La pyramide des mémoires

Les mémoires de masse (disques) permettent de stocker beaucoup d'informations mais elles sont très lentes par rapport aux performances du processeur. Par contre, une mémoire cache L1 est extrêmement rapide mais elle n'autorise que le stockage de quelques octets.

On arrive ainsi à une hiérarchie en pyramide où, à mesure que l'on monte :

- la mémoire est de plus en plus proche du processeur,
- la mémoire est de plus en plus rapide.

Le temps d'accès passe de quelques millisecondes pour le disque dur à une nanoseconde pour les registres, soit un ratio de 1 à 1.000.000.

- la capacité de la mémoire diminue,
- le prix du byte stocké augmente.

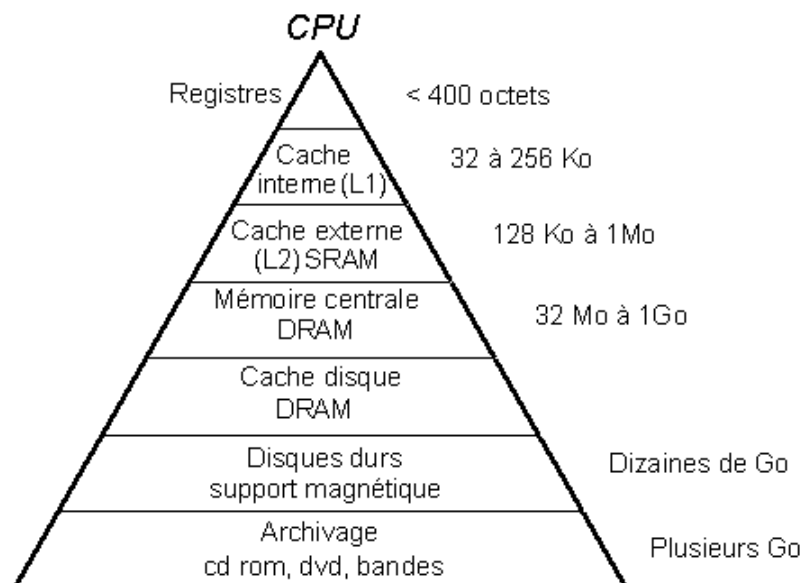


Figure 4.2 La pyramide de la mémoire

4.5.5. La collaboration dans la pyramide

La pyramide agit comme une pile dans laquelle les données (ou les instructions) montent progressivement en fonction des besoins.

Supposons par exemple qu'un utilisateur veuille modifier une information archivée sur un CD-ROM.

La première opération consiste à recopier le fichier sur le disque dur. Ensuite, l'utilisateur lance l'exécution d'un éditeur de texte (genre Notepad.exe) et ouvre le fichier en question. Ceci provoque le transfert d'une partie du fichier vers la mémoire centrale (RAM) au travers du cache disque. Le secteur dans lequel se trouve la donnée à corriger est copié dans le cache externe L2. Un petit fragment de ce secteur est copié à son tour dans le cache interne L1. Enfin, le CPU appelle dans le registre le ou les quelques caractères à modifier.

En réalité, les choses peuvent se passer un peu différemment mais le principe reste valable. Ceci explique pourquoi l'ouverture d'un fichier n'est pas une opération instantanée.

Lors de l'enregistrement, les différents étages sont mis à jour en partant du haut. A nouveau, les choses peuvent être un peu différentes dans la réalité mais c'est le principe général que nous tentons d'illustrer.

Le registre modifié est recopié dans le cache L1, lui-même est recopié dans le cache L2. Celui-ci rafraîchit la zone de la RAM dont il provient. Le contenu du fichier présent dans la RAM est ensuite transféré vers le disque, page par page, au travers du cache disque.

Tout ce processus de purge (*ang.: flush*) en cascade des caches prend un certain temps. C'est pour cette raison que, lors de l'arrêt d'un PC, Windows affiche le message "Veuillez patienter quelques instants...". C'est pour la même raison que des données sont perdues lors d'un arrêt intempestif du PC : les données présentes dans les caches n'ont pas eu le temps d'être recopiées sur le disque.