

SEANCE DU 08 décembre 2015.

Restitution de l'intervention de :  
Georges Linarès et Didier Josselin

Par l'équipe d'auditeurs : Barbara, Camille, Joëlle, Michèle, André, Gilles et Roland

TITRE : Informatique, mémoire et oubli.  
Première partie

Présentations de **Jean-Robert Alcaras** :

Nous accueillons ce soir Georges Linarès, vice-président de l'Université, en charge de la recherche, professeur d'informatique à l'Université d'Avignon, qui travaille sur l'intelligence artificielle, et Didier Josselin, qui était venu l'année où nous avons traité des frontières, directeur de recherche au CNRS, géographe, qui utilise l'informatique dans son travail. Ils se sont tous les deux associés ce soir pour nous proposer un certain nombre de réflexions sur « Informatique, mémoire et oubli », et je les remercie beaucoup d'être venus pour nous.

**Georges Linarès** : pour préciser un peu les choses, sur nos activités de recherche, comme chercheur je travaille en traitement automatique du langage naturel écrit et oral, traitement de la parole, reconnaissance de la parole, et ces dernières années pas mal sur tout ce qui est autour du web, c'est un peu décalé par rapport à ce que l'on appelle le traitement du langage naturel, qui sont les méthodes que vous employez lorsque vous envoyez une requête sur un moteur de recherche. En gros, mon activité de recherche tourne autour de ces questions-là. Je vais laisser la parole à Didier Josselin.

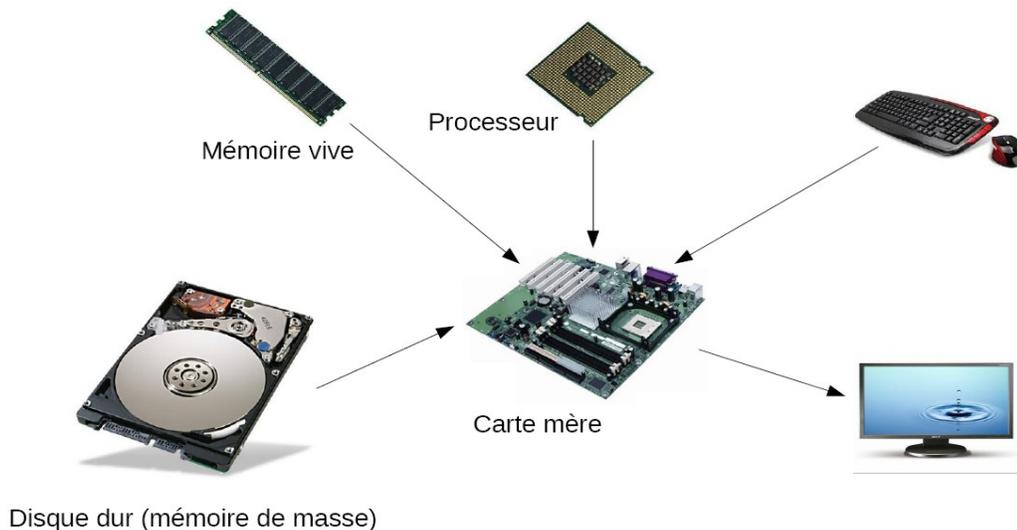
**Didier Josselin** : alors je suis géographe mais surtout géomaticien, mot dans lequel il y a « géographie » et « matique » qui est justement emprunté à l'informatique ; je suis si vous voulez un utilisateur de l'informatique au service de la géographie, et c'est pour cela que l'on fait un binôme pour cette présentation.

**G.L.** : nous allons venir à notre présentation, et juste préciser que l'on va vous présenter mémoire et oubli en vous parlant des aspects qui sont très liés au matériel, à l'architecture des ordinateurs, à la façon dont les ordinateurs sont capables de mémoriser l'information pour aller vers des problèmes qui sont un peu plus loin la machine, et qui sont plus près de l'usage que l'on peut faire de la machine. Concrètement, je vais vous présenter les divers aspects physiques de la mémoire d'un ordinateur. Je vais essayer de mettre cela dans une perspective historique ; l'informatique est un domaine qui évolue constamment, très rapidement, et il est important de voir où l'on en est, et dans quelle perspective cela s'inclut plus globalement.

### **Plan**

- *Architecture d'un ordinateur*
- *Historique*
- *Codage et représentations*
- *Optimiser la mémoire : vitesse d'accès, compression, robustesse.*
- *Mémoire et oubli : la mémoire comme un processus*
- *Mémoire et oubli : la mémoire comme fonction globale*
- *Mémoire distribuée : le découplage support/information*
- *Usages de la mémoire numérique*
- *Informatique et libertés*

Je pense que vous avez tous vu un ordinateur, qui est une boîte fermée, avec un clavier, une souris, un écran, et à l'intérieur, il y a un ensemble de composants que l'on retrouve à peu près dans tous les ordinateurs, qui peuvent être plus ou moins intégrés c'est-à-dire collés les uns aux autres. Nous avons, en gros, une carte mère, qui est la partie centrale de l'ordinateur, qui supporte presque tous les autres périphériques.



Il y a sur la carte mère :

- le processeur, c'est lui qui fait le traitement ; en réalité, il y a d'autres parties du circuit qui font aussi le traitement, mais l'essentiel est fait par le processeur.
- la mémoire vive, sur laquelle je reviendrai, qui est une mémoire assez rapide, c'est elle qui interagit le plus avec le processeur.

En dehors de la carte-mère, il y a le disque dur, qui peut avoir la forme représentée. Dans cette boîte, il y a des disques empilés et des têtes de lecture : cela ressemble un peu aux platines de lecture des disques microsillons que l'on a connus. Cela fonctionne un peu différemment, puisqu'il y a un empilement de disques, et pas un seul, il y a une logique pour les utiliser, mais en gros, on a le même principe de disques qui supportent l'information.

Pour finir, les périphériques entrées-sorties : écran, clavier, souris, et potentiellement plein d'autres choses.

Ce sont les éléments principaux de l'ordinateur. Pour ce qui est de la mémoire de l'ordinateur, ce sont ces barrettes vertes avec des circuits soudés ; c'est une mémoire qui est assez rapide d'accès pour le processeur, c'est aussi une mémoire qui est volatile. C'est un concept important lorsque l'on parle de mémoire : lorsque vous éteignez l'ordinateur, tout ce qui se trouve dans la mémoire vive s'efface. C'est une mémoire que l'on peut réinscrire en permanence, c'est un support, on peut mémoriser de l'information et ensuite changer à un endroit, puis écraser cette information (qui est stockée à un endroit) par une autre, et lorsque l'on éteint l'ordinateur, on perd tout ce qui est écrit dans cette mémoire.

C'est une mémoire qui est relativement chère, en tout cas, le prix relatif à l'octet (l'unité d'information que l'on peut mémoriser) est relativement élevé. C'est une mémoire qui est très rapide. Typiquement dans les ordinateurs modernes, on peut considérer que l'on a entre un et huit gigaoctets de mémoire, soit 8 000 000 000 d'octets, mémorisables sur ces supports-là. On peut avoir des machines qui ont beaucoup moins de mémoire que cela, il y a des ordinateurs partout, dans vos voitures, dans vos GPS, ce sont des ordinateurs qui ont quelques kilos octets seulement de mémoire. Cela veut dire quelques milliers d'octets, l'octet est le mot élémentaire, on peut considérer que c'est, disons, un caractère (on peut dire aussi que un octet est composé de huit interrupteurs qui ont soit la valeur « zéro », soit la valeur « un », et qui à travers leurs combinaisons donnent une information numérique unique, et qui ont une importance relative suivant leur position dans la lecture).

Le disque dur est connecté à la carte mère, il constitue une mémoire de masse, dont la capacité est beaucoup plus élevée, typiquement un disque dur sur une machine actuelle va de 100 mégaoctets jusqu'à deux téraoctets. On n'est plus du tout sur la même échelle que ce que l'on peut avoir en mémoire vive ; c'est une mémoire qui est plus lente, moins chère et persistante, c'est-à-dire que ce que vous enregistrez sur votre disque dur reste présent sur le disque dur, même si vous coupez l'alimentation électrique. Vous retrouverez tout lorsque vous remettrez votre ordinateur sous tension. Ce sont des mémoires qui, au niveau capacité et prix ne sont pas du tout du même niveau, mais en même temps, elles n'ont pas du tout la même fonction.

Il y a d'autres dispositifs qui permettent d'enregistrer des informations dans les ordinateurs : vous avez les DVD, on peut écrire un DVD, on peut le lire, c'est encore plus lent qu'un disque dur, ce n'est pas très cher, du coup on arrive à avoir de grandes quantités d'informations stockées à bon marché sur des DVD, et on a une durée de vie qui est plus élevée que celle des disques durs.

La durée de vie correspond au nombre d'accès que peut supporter la mémoire sans qu'elle finisse par tomber en panne, c'est-à-dire qu'un disque dur est une machine, et qu'il y a un nombre d'utilisations prévues que peut supporter la machine, par exemple un certain nombre d'écritures sur le disque. Au-delà de cela, on va avoir un risque de panne : ce nombre d'accès ou cette durée de vie moyenne du support fait partie de l'équation lorsque l'on cherche quel est le type de support que l'on doit utiliser (disque dur, bande magnétique,...).

Un autre aspect qui est important est que la mémoire est indexée. Qu'est-ce que cela veut dire ? Lorsque l'on a de la mémoire vive, la mémoire vive est organisée dans l'ordinateur par adresse. Cela veut dire que l'on accède à une mémoire en connaissant l'adresse de la mémoire. C'est la métaphore de la commode à tiroirs : chaque case contient un élément, et ajouter de la mémoire, c'est ajouter des tiroirs, et l'on accède au contenu de ce qui est dans les tiroirs par la position du tiroir.

Cela correspond à l'accès standard en informatique mais nous ne fonctionnons pas comme cela. Imaginons que vous essayez de vous rappeler une scène de votre enfance, ou le

numéro d'immatriculation de votre voiture, vous ne dites pas « *cela je l'ai stocké dans le nord au numéro xxxx* », cela ne fonctionne pas de cette façon. Accéder au numéro d'immatriculation de votre voiture en essayant de mémoriser, mais vous avez un doute sur les deux lettres centrales par exemple, connaissant les chiffres, vous faites un accès par le contenu. Vous avez une idée de l'information que vous voulez retrouver, mais c'est une idée partielle. Et c'est cette idée partielle qui vous aide à retrouver ce que vous cherchez ; cela peut être très indirect, si l'on prend l'exemple de la Madeleine de Proust, c'est une odeur qui lui rappelle une scène de son enfance, si je me souviens bien. Ceci étant dit, on est toujours sur du contenu, on n'est pas sur une sorte de numérotation de la mémoire, et c'est comme cela que les machines fonctionnent.

Ce que l'on va voir, c'est que cette idée d'indexation, par des contenus ou des adresses, est un point qui est très fondamental et notamment plus on va vers des idées de représenter l'information de façon complexe, plus on va vers de l'accès par le contenu plutôt que par les adresses. Ceci dit, on manipule quand même, dans la vie quotidienne, l'accès par adresse.

Nous passons au stade suivant : l'évolution du prix au mégaoctets (un mégaoctets= 1 million d'octets, on peut considérer que c'est 1 million de lettres) :

### Evolution des prix (disque dur):

#### Un mégaoctet (1 chanson MP3, un livre)

- 1956 : 10 000 \$
- 1980 : 700 \$
- 1990 : 10 \$
- 2000 : 0.02 \$
- 2010 : 0.001 \$

Cela veut dire que, plutôt que d'avoir des disques qui font 10 mégaoctets, on aura des disques qui font un ou deux téraoctets pour le même prix. La comparaison précédente s'applique aux mémoires de masse des ordinateurs quel que soit leur type, le disque dur n'est apparu que dans les années 80 effectivement.

**Question d'un auditeur** : A quel système correspondent les clés USB actuelles ?

**Réponse** : C'est une bonne question parce que c'est une espèce d'intermédiaire entre la mémoire vive et le disque dur. Au niveau fonctionnel, c'est comme un disque dur : c'est une mémoire persistante, lorsque vous enlevez votre clé vous ne perdez pas les informations, sur les capacités, on va de 2 à 64 gigaoctets, il y en a peut-être de plus grosses, je ne sais pas. C'est un peu ambigu parce que cela fonctionne comme un disque dur amovible, mais par contre au niveau technologique, ce qu'il y a dans la clé USB, cela ressemble à de la mémoire vive, mais pas exactement. C'est une technologie avec un circuit intégré qui enregistre l'information, on est sur une sorte d'intermédiaire, et maintenant il y a des disques durs qui sont basés sur cette technologie-là : ils sont plus chers, mais cela n'est pas hors de prix, et ils sont très rapides d'accès parce que basés sur une technologie de circuits intégrés, et non sur une technologie mécanique à tête de lecture. C'est plus compact et l'on peut accéder plus rapidement à l'information.

### Question d'un auditeur : stockage sur de l'ADN

**Réponse :** La question de la permanence : tout dépend de ce qu'on veut en faire. Il y a des disques externes garantis dix ans. La technologie évolue très vite et donc repose la question de la durabilité. C'est pour cela aussi que la sauvegarde du support matériel sera revue à la fin. On est très concentré sur la technologie et cela soulève le problème de quel service cela rend, que veut-on en faire ?

À propos de la technologie des mémoires, l'évolution est tellement rapide que les systèmes deviennent obsolètes avant la fin de leur durée de vie. Par exemple, si l'on avait encore des machines des années 50 encore en état opérationnel, d'accord on aurait fait un très bel objet qui continue à fonctionner, sauf que l'on ne pourrait rien en faire, puisqu'il serait dépassé au point de vue technologique. Donc cette question de la durabilité du dispositif, il faut aussi la poser en ces termes là (c'est aussi le problème des disquettes que l'on a connu qui sont impossibles à lire actuellement, dont une énorme quantité tiendrait sur une clé USB actuelle, et qui contiendraient des erreurs en très grande quantité). Il est important de bien dissocier la question du support matériel et celle de l'information que l'on veut stocker : on va de plus en plus vers cela. Je ne veux pas anticiper sur la fin de la présentation, mais lorsque l'on parle Internet, que l'on parle du « cloud », c'est exactement cela, peu importe comment est stockée l'information, il faut voir en terme fonctionnel : quels services cela rend, ce que l'on veut en termes de service, car cela peut être contre-productif. Si l'on oublie la technologie, un moment, on arrive à trouver la solution technologique qui est vraiment adaptée à l'usage, à ce que l'on veut en faire. En général la technologie n'est pas un objet qui se suffit à lui-même, c'est un objet qui est destiné à rendre un service, et c'est parfois quelque chose que l'on a tendance à oublier.

Sur les capacités, évolution historique :

	1980	1990	2000	2010	2020 (prévision)
<b>Mémoire RAM</b>	16 Ko	1 Mo	128 Mo	4 Go	256 Go
<b>Disque dur</b>	5 Mo	40 Mo	20 Go	1 To	50 To
<b>Clé USB</b>	/	/	32 Mo	64 Go	16 To
<b>CD, DVD</b>	/	650 Mo (CD)	4,7 Go (DVD)	25 Go (Blu-Ray)	200 Go

La mémoire RAM correspond à la mémoire vive :

(1 Ko=1000 octets, 1 Mo= 1000Ko=1 000 000 octets, 1 Go= 1000Mo=1 000 000 000 octets, 1 To= 1000 Go=1 000 000Mo=1.000.000 000Ko=1 000 000 000 000)

Il faut se méfier des prévisions : en général, c'est souvent ce qui n'arrive pas, parce que ce n'est pas souvent très régulier au niveau des courbes qui permettent d'anticiper sur la façon dont les choses évoluent, mais on se trouve dans des systèmes qui sont soumis à de multiples paramètres qui n'évoluent pas tous de la façon prévue. Pour les 256 Go de mémoire vive en 2020, on a au laboratoire une machine qui a 512 Go de mémoire depuis deux ans ; donc les 256 Go de mémoire vive c'est plutôt réservé à l'ordinateur que vous avez à la maison. Ce sont des choses que l'on peut faire actuellement sans problème, et donc la prévision est plutôt pessimiste.

Pour le disque dur, actuellement passer de 1 à 2 To, correspond à une toute petite somme, il n'y a pas d'écart significatif pour doubler la capacité du disque dur.

Pour les clefs USB, je trouve que les prévisions sont un peu optimistes, en particulier les 64 Go de 2010.

Pour les disques, CD et DVD, cela peut continuer à croître, c'est un peu hypothétique, mais on peut imaginer que cela se développe de façon importante dans les années à venir.

Jusqu'à présent, je vous ai parlé de l'ordinateur de bureau que vous avez chez vous, de votre portable, mais il y a un autre aspect qui est la quantité d'informations stockées numériquement dans le monde :

## Historique

Quelle quantité d'information numérique  
mémorisée dans le monde :

- 1 mégaoctet : 1 livre
- Sur un ordinateur domestique : 500 000 mégas
- Un cerveau humain : 3 000 000
- Disponibles sur le Web : n\* 1000 000 000
- Produites chaque année : n\* 1000 000 000 000

Il y a une explosion de la quantité de la capacité individuelle des machines, mais il faut savoir qu'Internet repose sur une quantité de machines, qui sont un peu partout dans le monde, et qui sont interconnectées. On a une explosion donc de la capacité des machines, du nombre de machines, de leur connectivité : la plupart des machines dans le monde sont interconnectées entre elles de façon plus ou moins directe. Je reviendrai là-dessus lorsque l'on parlera plus directement d'Internet.

Je vous l'ai dit tout à l'heure, un mégaoctet c'est un livre, sur un ordinateur domestique imaginons que vous avez 500 Go, un cerveau humain, on considère que c'est 3 To, ce qui n'est pas mal, sauf qu'il est difficile de comparer les choses de cette façon, pour les raisons que j'évoquais tout à l'heure, c'est-à-dire que la façon de stocker l'information dans le cerveau est très différente de celle que l'on utilise, au niveau matériel en tous les cas, dans les ordinateurs. Il faut se méfier un peu de cette comparaison, mais en gros la capacité de mémorisation d'un cerveau est de 3 To.

Sur le Web, c'est plusieurs milliers de To, ce qui commence à faire beaucoup de livres ! Sauf que tout n'est pas accessible sur le Web : il y a des bases de données qui ne sont pas indexées sur les moteurs de recherche et il y a des entreprises qui ont de grandes quantités d'informations qui ne sont pas forcément ouvertes sur l'extérieur, je pense aux militaires qui ont des bases de données qui ne sont pas ouvertes à tous et là, cela devient difficile de voir à quoi correspond cet empilement de nombres de zéro correspondants.

Ce qu'il faut garder à l'esprit, c'est que la quantité d'informations mémorisées informatiquement dans le monde est énorme et en explosion ces dernières années, et va au-delà de l'augmentation de la capacité individuelle des machines. Il y a aussi une augmentation du nombre de machines, qui fait que la quantité d'informations stockées dans le monde est énorme. Ceci correspondait à la perspective historique.

Sur le codage et les représentations :

## Mémoire : codage et représentations

### Les mémoires d'un ordinateur :

- La mémoire stocke des ensembles de nombres (binaires)
- C'est l'organisation de ces nombres qui constitue la représentation interne d'un « objet » mémorisé
- Ces représentations sont très variables :



Didier en a parlé rapidement tout à l'heure : la mémoire ce sont des ensembles de nombres binaires, aussi bien sur les disques que sur les barrettes mémoire. Dans la mémoire on a une suite de nombres, et les nombres sont stockés comme des paquets d'octets. En réalité un octet est un nombre binaire, écrit en base 2. Pour comprendre plus simplement, un nombre entier que l'on utilise est écrit en base 10, une suite de chiffres qui vont de 0 à 9. Un nombre binaire est un nombre qui s'écrit en base 2, avec des chiffres qui vont de 0 à 1.

Fondamentalement, cela ne change rien, techniquement, cela change tout : parce que, dans un processeur et dans une mémoire, en fait, on a des interrupteurs électroniques microscopiques (on fait cela presque au niveau atomique). Mais l'idée, c'est quand même que l'on a des interrupteurs qui sont positionnés et c'est le positionnement de ces interrupteurs, qui vont être globalement par paquets de 8, interprétés comme des octets, et ce sont ensuite ces paquets d'octets que l'on va utiliser pour stocker des informations plus complètes.

On peut donc considérer que dans une mémoire on a une suite de nombres, mais cela ne dit pas tout sur la façon dont on mémorise les informations dans les ordinateurs. Au-dessus de ce point de vue très bas niveau, qui est : toutes les informations sont stockées comme des suites de nombres, on a une organisation de ces nombres qui va permettre de stocker des structures complexes. Ces informations complexes que l'on peut structurer et coder in fine comme une suite de nombres, cette façon de représenter l'information, peut-être, en fait, très diverse.

Je vais expliquer cela avec un exemple concret : je vais prendre une image numérique, que vous avez plus haut. Comment peut-on la stocker ? Si l'on considère juste la suite de nombres, ce n'est pas difficile en fait : vous savez qu'une image informatique est une suite de pixels, une suite de points. Chaque point a une couleur, donc si l'on considère que dans la mémoire on va mettre chaque point avec sa couleur, en gros on a stocké notre image.

Par contre, en vidéo, que fait-on ? On a une suite d'images, et une image ça peut être très lourd : je dis un peu au hasard, 500 pixels. Vous avez une image avec 500 lignes et 500 colonnes, soit 250 000 points. Avec le codage nécessaire de la couleur, cela fait au moins 1 Go (250 000 nombres qui vont occuper en mémoire 1 Go), avec une résolution qui n'est pas extraordinaire, vous pouvez avoir sur votre télévision une meilleure résolution que celle-là. Ensuite, une vidéo c'est une suite d'images, que l'on affiche les unes après les autres et cela donne l'illusion d'un mouvement continu. Typiquement, on va avoir 50 images à la seconde, ce qui nous donne 50 Go par seconde. Donc le film que vous voulez voir le soir sur votre ordinateur, à 50 Go par seconde, vous allez voir simplement les premières secondes, et en plus c'est un peu absurde, parce que, si cette image reste à l'écran une demi seconde, elle ne bouge pas du tout, c'est toujours la même image. Si cela bouge très peu, il n'est peut-être pas utile de tout sauvegarder. Vous voyez, en fait, que cette façon de coder (qui est très naturelle, c'est la première chose qui vient à l'esprit), arrive très vite à ses limites.

Comment peut-on représenter autrement cette image ? On pourrait, par exemple, considérer que l'on va stocker plutôt la différence d'un pixel à l'autre : il n'y a pas de rupture de couleur abrupte, en tout cas, on pourrait considérer que si l'on a une série de pixels qui sont de la même couleur, on pourrait ne pas avoir à répéter les informations pour chaque pixel.

Si l'on imagine le même type de procédé entre deux images d'une vidéo, on pourrait stocker la différence qu'il y a entre les deux images. Par exemple, si vous regardez la télé sur ADSL et qu'à un moment il y a un ralentissement du réseau, vous pouvez voir qu'à certains moments vous avez une sorte de pixellisation : certaines parties de l'image sont visibles et celles qui bougent ne le sont plus. Il en est ainsi parce que c'est codé comme cela. L'information n'arrive plus assez vite et du coup la mise à jour de l'image, cette information qui est liée au changement d'une image par rapport à sa précédente, n'arrive plus ; par contre, ce qui était là antérieurement et qui ne change pas, reste. Mais en fait, on peut utiliser beaucoup d'autres façons pour coder les images : on peut par exemple coder des textures, en créant des modèles numériques de textures, au sens graphique, par rapport au modèle de texture standard.

On peut aussi considérer que cette image est, en fait, un ensemble de formes élémentaires qui sont déformées et réassemblées. C'est un point de vue qui est très intéressant, et, pour être honnête, cette image a été conçue comme cela.

## Mémoire : codage et représentations

Exemple : les représentations d'une image

- Tableau de points colorés
- Ensemble d'objets graphiques superposés
- Assemblage de formes élémentaires déformées
- Mémoriser le procédé de construction de l'image :



C'est une image de synthèse qui a été générée par un logiciel qui part d'une description de la scène, une description de l'image, assez abstraite, qui va permettre à l'ordinateur de générer l'image telle qu'elle serait vue par l'utilisateur. C'est quelque chose de très intéressant, parce que au-delà du codage de l'image dans l'ordinateur, cela montre qu'il y a un moment où l'on dévie du stockage de l'information au procédé : finalement, si l'on arrive à avoir une description de cette scène (il y a de l'eau, une surface plane, avec un coefficient de réflexion,...., si l'on décrit tout de cette façon, on stocke finalement assez peu d'informations par rapport à la description de chacun des pixels, mais par contre, on met en œuvre des processus qui sont capables de construire l'image à partir de cette représentation qui est finalement assez éloignée de l'image que l'on voit.

On bascule d'une description qui est très exhaustive (de chaque pixel et de sa couleur), à une description qui est très générale de l'image, et on a le mécanisme qui nous permet de reconstruire l'image à partir de cette description. Et cette idée que l'on a d'un côté de l'information qui est stockée, est associée à cette information stockée, un processus qui permet de générer la vue finale, c'est un mécanisme que l'on va trouver à de nombreux niveaux dans la représentation de la mémoire, mais aussi dans les algorithmes de compression, ou dans les algorithmes de correction des erreurs.

On passe maintenant à optimiser l'espace : il y a une différence entre l'espace que l'on occupe, et la quantité d'informations que l'on stocke, ce qui correspond exactement à ce que je vous disais à l'instant : pour stocker une image il nous faut 1 Go, ou 50 Go la seconde de vidéo. Si l'on code simplement la différence d'une image à l'autre, en partant du fait que d'une image à l'autre, il y a très peu de différences, on réduit l'information qui est stockée, sans rien perdre. Autrement dit, on a la même quantité d'informations qui est supportée par moins d'espace.

## Optimiser la mémoire

- Optimiser l'espace :
  - Espace et quantité d'information
  - Optimiser l'espace : compresser les données
    - Objectif : plus d'information dans moins de place
  - Quelques possibilités :
    - Elimination des répétitions (exemple : la vidéo)
    - Factorisation de l'information
    - Avec erreur : supprimer l'information la moins importante (exemple : MP3)
  - Remarque : la compression fait gagner de l'espace mais :
    - Et la vitesse d'accès ? Et les temps de compression/décompression ?

Il faut faire attention à cela, l'espace physique de la mémoire et la quantité d'informations contenues sont deux choses totalement différentes. Il y a un domaine de recherche et de technologies qui porte sur la compression d'informations : comment trouver des représentations de l'information qui sont les plus compactes possibles pour occuper le moins d'espace possible en mémoire, en éliminant les répétitions. La factorisation de l'information est une sorte de généralisation de cette façon de procéder : plutôt que de tout écrire, il y a des parties que l'on va retrouver, pas seulement d'une image à l'autre, mais il y a aussi peut-être des motifs que l'on va

retrouver en de nombreux endroits différents. Si vous regardez par exemple du texte, vous avez évidemment des séquences de lettres qui reviennent tout le temps : si vous stockez des séquences de lettres au lieu de stocker chaque lettre, cela va prendre beaucoup moins de place. Il y a des algorithmes de compression qui sont basés sur ce procédé, avec un dictionnaire de ces formes fréquentes, et l'on va stocker l'adresse de la forme qui revient régulièrement, qui peut être assez volumineuse.

Ce qui vient d'être décrit correspond à des approches sans erreurs : on ne perd rien, on a exactement la même quantité d'informations. Mais, ensuite, a-t-on besoin de tout stocker ? Il est évident que, sur un texte, c'est embêtant si l'on perd des caractères, mais est-ce le cas aussi sur une image ou sur un enregistrement audio ? Un enregistrement classique d'un fichier audio est très volumineux, parce que l'on stocke tout, qui correspond à peu près à la représentation d'une image pixel par pixel : on aura exactement la forme de l'onde acoustique qui est stockée.

Il y a une autre façon de procéder, qui est de considérer qu'un signal audio, un son, est la superposition de sons élémentaires qui seraient des sinusoïdes : si vous avez un son très pur, on change la fréquence, on change l'amplitude, et le mélange de toutes ces formes élémentaires, recompose n'importe quel signal acoustique. Sauf que dans ces fréquences (en gros, l'oreille est faite comme cela, vous avez une perception des fréquences, un son correspond à des fréquences superposées), votre oreille capte ces différentes fréquences, mais elle ne capte pas toutes les fréquences. Lorsque vous mémorisez de façon parfaite un son, en réalité vous stockez une partie du son qui n'est pas audible ; le principe du MP3 est de supprimer ces parties qui ne sont pas audibles et en fait, on va un peu plus loin : on supprime des parties non audibles, et on supprime les parties qui sont les moins perçues, et plus on veut compresser, plus on va enlever des composantes en fonction de l'impact perceptif qu'elles ont. Il y a une chose intéressante : au lieu de poser le problème en termes de quantité d'informations que l'on peut plus ou moins quantifier, on se pose la question de savoir à quoi cela sert, chose dont je parlais tout à l'heure.

Si l'on mémorise un fichier audio, la musique est là pour être entendue et finalement on peut comprimer, en enlevant de l'information, du moment que cette information n'est pas pertinente pour l'usage que l'on veut en faire. Lorsque l'on s'autorise à perdre de l'information, on peut comprimer beaucoup plus puisque l'on peut oublier les éléments qui sont complètement marginaux et qui arrivent rarement ; on peut enlever tous les éléments qui consomment de la place et ne garder que l'essentiel. Une petite remarque sur cette histoire de compression, qui fait gagner de l'espace, mais qui peut poser un problème au niveau de la vitesse d'accès : si l'on a un codage un peu compliqué, comment faire pour accéder à l'information que l'on avait stockée dans la barrette mémoire à telle position ? C'est un peu moins évident et le problème de la compression c'est aussi cela.

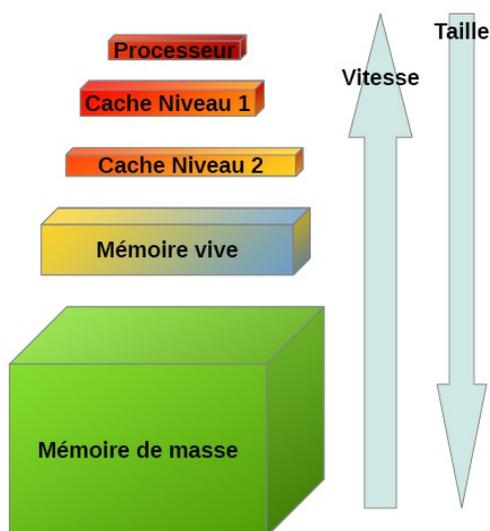
Il est évident que la compression fait gagner de la place, c'est fait pour cela, mais qu'est-ce que cela implique en termes d'accès, de mode d'accès, quels sont les processus mis en œuvre pour récupérer cette information ? et il peut y avoir aussi la robustesse. Lorsque vous stockez votre image comme suite de pixels, à un moment il y a une erreur qui vous fait perdre un pixel (par exemple le rayonnement cosmique perturbe le fonctionnement de l'électronique et fait perdre une information), au niveau de l'image, cela ne change rien, on ne voit pas la différence. Par contre si cela se passe sur une partie qui est compressée, vous pouvez perdre une information qui est indispensable pour la reconstruction de l'ensemble de l'image. Donc la question de la compression c'est une question d'espace, mais c'est aussi ce que cela implique en termes de temps d'accès, de robustesse de l'information qui est stockée et de capacité de récupération en cas de défaillance de la mémoire.

Pour revenir à l'architecture matérielle, l'espace et la question du temps d'accès est une contrainte qui est très centrale, y compris sur les aspects technologiques, en construction matérielle d'un ordinateur : généralement plus les structures sont grosses, et plus l'accès est lent. Les supports matériels de la mémoire dans un ordinateur sont engagés de la façon suivante : il y a la mémoire de masse sur le disque dur, qui ne va pas très vite, mais qui stocke de très grandes quantités d'informations ; la mémoire vive va plus vite, est volatile, mais stocke un peu moins d'informations. En réalité, il y a d'autres niveaux entre la mémoire vive et la mémoire de masse sur un ordinateur : il y a des mémoires qui sont très rapides, qui coûtent très cher, et donc qui sont en quantités limitées. Au lieu d'avoir de la mémoire vive qui va très vite, en très grande quantité, mais qui est hors de prix, on met de la mémoire qui est très rapide et qui sert de tampon entre les deux. Si l'on accède tout le temps qu'à la même partie de la mémoire vive, cette partie de la mémoire vive à laquelle on accède tout le temps va être mise dans une mémoire « *cache* » qui est plus rapide ; donc le processeur va accéder très rapidement à cette mémoire, et lorsqu'il ne l'utilisera plus, il y a un mécanisme qui va l'enlever de la mémoire *cache*.

En fait, il y a différents niveaux de mémoires, à différentes vitesses, qui permettent d'accélérer les traitements, qui donnent au processeur l'illusion que la mémoire que l'on utilise est plus rapide que ce qu'elle est en réalité.

## Optimiser la mémoire

- Temps d'accès : les temps d'accès sont variables. Ils dépendent :
  - Des performances matérielles
  - Disque dur, mémoire vives, mémoires caches



Ce principe de mémoire *cache* est un principe que l'on retrouve à tous les niveaux, par exemple dans la mémoire vive, vous allez avoir une image du disque dur réduite qui va contenir les données les plus utilisées du disque. Et l'on a le même processus que l'on reproduit à des niveaux différents entre le niveau 2 et le niveau 1, pour avoir, en gros, à notre disposition une mémoire très importante, dont on voudrait qu'elle aille plus vite, et une mémoire réduite qui va très vite, mais en faible quantité. En fait on essaie de jongler entre les deux pour avoir tout le temps à notre disposition par une mémoire rapide, les informations dont on a besoin.

**Question d'un auditeur :** est-ce que la mémoire *cache* est la mémoire utilisée, par exemple, lorsque l'on redémarre l'ordinateur et que l'on va sur Internet, et qu'il y a des informations qui apparaissent ?

**Réponse :** c'est exactement le même principe : lorsque vous allez sur votre navigateur, vous avez des informations qui sont données : en fait, le *cache*, c'est une image locale sur votre ordinateur d'une page par exemple, qui a été récupérée d'un ordinateur distant sur Internet. Sauf que la récupération de cette page, sur l'ordinateur distant, est un processus qui est un peu lent, puisqu'il faut aller interroger un serveur distant par une requête, une information qui transite à travers Internet, pour aller jusqu'à ce serveur, et ensuite ce serveur renvoie ce que vous voulez (la page que vous attendez), à travers le réseau pour arriver sur votre ordinateur. Il est évident que si vous accédez toujours à la même page, vous n'allez pas faire les allers-retours en permanence ; donc l'ordinateur garde en mémoire les pages que vous avez vues récemment, ou que vous voyez fréquemment, de façon à vous en donner un accès immédiat. C'est exactement ce principe-là : la mémoire de l'ordinateur est utilisée comme une mémoire *cache* d'une mémoire plus générale et lente, qui est l'Internet.

**Question d'une auditrice :** combien de temps va rester stockée cette information sur la page que l'on va voir ?

**Réponse :** c'est plus tôt le navigateur qui gère cela, pour être honnête, je ne sais pas exactement ce qu'il en est, c'est peut-être paramétrable, il n'y a pas de règle universelle et systématique sur la gestion des *caches*. Vous pouvez décider d'utiliser plus ou moins de mémoire vive de votre ordinateur comme *cache* ; cela dépend de nombreux paramètres, en fonction de l'utilisation que vous faites de votre ordinateur, des fournisseurs, et il y a un problème : c'est en fait que la page peut changer : si vous revenez un an après sur la même page, il faut être sûr que l'on est bien sur des informations qui sont à jour. C'est pour cette raison que vous pouvez avoir des modifications quelque part, que vous ne voyez pas, parce que, en fait, vous n'avez pas rafraîchi votre page et vous avez une image obsolète qui est fournie par le *cache*. Ce mécanisme de *cache* se retrouve à tous les niveaux.

**Remarque d'un auditeur :** il y a une touche magique, F5, qui rafraîchit tout.

**Question d'un auditeur :** du coup, cette notion de mémoire *cache*, désigne une fonctionnalité de mémoire, pas un type de mémoire physique ?

**Réponse :** cela se confond un petit peu. Conceptuellement, c'est vrai, cela n'est pas la même chose : on utilise de l'espace mémoire comme *cache*. Lorsque l'on regarde la fiche technique d'un ordinateur, d'un processeur, vous avez pour tel processeur, tant de mémoire *cache*. Ce sont des éléments techniques que l'on retrouve aussi dans les offres commerciales, et là, pour le coup, c'est de la mémoire qui est intégrée au processeur. Cela peut se confondre un peu : la mémoire *cache* est une mémoire vive qui est très rapide, technologiquement c'est à peu près de la mémoire vive, avec des performances qui sont accélérées.

**Question d'un auditeur :** le support matériel de cette mémoire *cache*, c'est la barrette ?

**Réponse :** en fait, cela dépend de quelle mémoire. Par exemple, sur Internet, cela peut-être le disque dur qui fait office de *cache* pour la mémoire plus lente qu'est Internet. Et la mémoire vive va faire office de *cache* pour la mémoire de masse, le disque dur. La mémoire qui est sur la barrette va être utilisée comme *cache* pour le disque dur. Par contre le *cache* de niveau 2, qui est un *cache* sur la carte mère, cette mémoire-là va faire office de *cache* pour la mémoire vive. En fait, il y a toujours l'une qui est le *cache* de l'autre, tout s'imbrique, c'est un mécanisme que l'on trouve partout, indépendamment du support.

**Question d'un auditeur :** au point de vue vocabulaire, le terme anglais quel était-il, parce que *cache*, c'est bizarre.

**Réponse** : c'est une mauvaise traduction, c'est un néologisme du terme anglais « *cache memory* ». L'idée est de cacher la mémoire lente qui est derrière.

Nous allons passer à la robustesse, qui correspond à ce qui se passe en cas de défaillance. Les raisons de la défaillance peuvent être variables, cela peut-être le rayonnement cosmique, l'information stockée n'est plus la bonne, ou il peut y avoir des problèmes matériels, le temps aussi fait son œuvre sur les circuits imprimés, et donc il peut y avoir des problèmes qui apparaissent soit sur les barrettes, soit sur les disques durs, c'est vrai pour tous les supports.

Le problème est de savoir ce que l'on fait lorsqu'il y a des erreurs, car les ordinateurs supportent très mal les erreurs. Si vous perdez un pixel sur votre image, cela n'est pas très grave, si vous avez une erreur dans le programme (qui est stocké), qui reconstitue l'image d'une vidéo, qui la décompresse, potentiellement, cela va donner n'importe quoi : vous perdez tout. Ce problème de robustesse peut être très variable, mais c'est un point qui est très critique en informatique. Il y a une quantité de stratégies différentes pour rendre le système robuste aux erreurs, avec des bases théoriques qui sont plus ou moins compliquées, nous n'irons pas très loin sur ce sujet, mais le point le plus naturel est de faire de la redondance : en gros on duplique les informations. Au lieu d'avoir un disque dur, on en a deux, avec un qui est le miroir de l'autre.

## Optimiser la mémoire

- Optimiser la robustesse :
  - Que se passe-t-il en cas de défaillance ?
    - Problème de robustesse
  - Correction des erreurs :
    - Redondance
      - Duplication, redondance partielle
      - Contrôle : somme de contrôle
    - codes correcteurs d'erreurs

Lorsque l'on a deux disques durs en miroir, que se passe-t-il s'il y a une erreur ? S'il y a une erreur, on ne sait pas d'où vient cette erreur : ce n'est pas le point de savoir qu'il y a un problème, le tout est d'arriver à le résoudre. Entre les deux disques, si l'un dit : «8» et l'autre : «18», lequel des deux a raison ? D'un côté on a un pixel bleu, de l'autre, un pixel vert ; quelle est la couleur du pixel ? Bleu ou vert ? Le fait d'avoir dupliqué les informations ne suffit pas pour reconstruire la bonne information. Alors, on peut avoir des mécanismes qui n'ont pas deux, mais trois disques, et l'on fait un vote majoritaire, et cela fonctionne vraiment comme ça !

Il y a ensuite des mécanismes très compliqués, je n'entrerai pas dans les détails, qui permettent une redondance partielle et du contrôle : il faut dissocier ce qui est la détection de l'erreur de sa réparation. Vous avez, par exemple, un ensemble de données qui transitent à travers Internet, vous les recevez, vous savez que la somme de toutes les valeurs qui ont été transportées doit être 500. Si vous n'avez pas 500, c'est qu'il y a eu un problème : la somme de contrôle n'est pas correcte, vous savez qu'il y a une défaillance. Soit vous n'utilisez pas vos données, soit vous redemandez à l'émetteur la bonne information.

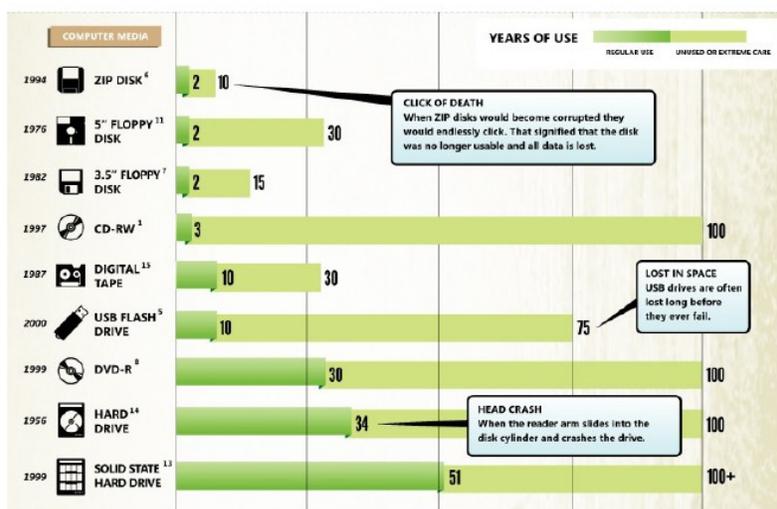
Il y a aussi une autre façon de procéder, compliquée, avec des codes correcteurs d'erreurs, qui est d'utiliser des bases mathématiques. C'est très utilisé dans l'informatique, c'est basé sur de l'algèbre, pour arriver à optimiser, à consommer le moins d'espace possible pour garantir l'intégrité des données.

Je vais laisser la parole à Didier.

**Didier.Josselin.** : Durée de vie des supports en informatique

Georges vous a parlé des différents éléments de la mémoire, comment ils étaient stockés ; (j'ai essayé, disons en terme d'utilisateur lambda), en géographie on a beaucoup de données de cartes, cela prend beaucoup de gigaoctets, et, du coup, la conservation de l'information est fondamentale. Il faut voir que l'on est passé d'un monde analogique à un monde numérique.

## Durée de vie des media informatiques



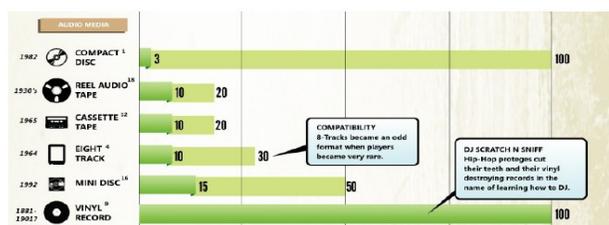
Par exemple, vous avez une représentation analogique, une sorte de courbe lissée, qui représente un son analogique, celui que vous écoutiez avec un disque. Et tout est transformé avec une perte d'informations que vous retrouvez finalement, avec ces écarts qu'il y a à la réalité, puisque vous passez d'une représentation qui est continue à une représentation qui est discrète, qui fait que finalement, en passant à la représentation numérique, vous allez perdre de la qualité. C'est là-dessus que l'on peut gagner, justement Georges en parlait à propos des MP3, c'est sur ce genre de simplification de l'information que l'on perd de la qualité. Par exemple, il y a des gens, avec de bonnes oreilles, qui vont arriver à détecter des MP3 avec de bonnes qualités.

Pour vous montrer les deux mondes, j'ai essayé de voir ce que l'on pouvait avoir sur les durées de vie des matériels.

Il y a plusieurs catégories : voici les différents matériels informatiques dont on a parlé. Vous vous souvenez sans doute du disque zip, mais on ne l'a pas gardé longtemps, la preuve est là. À l'opposé vous avez les nouveaux disques durs qui ont une capacité vraiment très importante, puisque 100, vous avez une estimation qui est au-delà de 100 ans ; à gauche vous avez un usage régulier, et de l'autre côté, une utilisation où vous utilisez peu. Vous pouvez voir, en fonction des supports, que vous avez des différences entre les deux modes d'utilisations qui sont très importants.

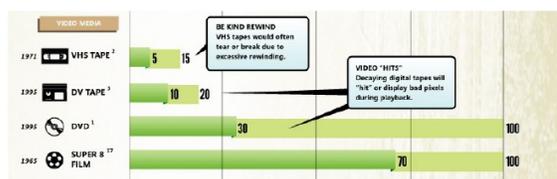
Typiquement, si vous regardez le CD, si vous l'utilisez beaucoup vous pouvez le jeter, par contre si vous stockez, vous pouvez le garder très longtemps. Je ne vais pas entrer dans le détail, mais moi, ce qui m'a étonné, c'est de voir la bonne capacité de l'USB flash qui est finalement, une technologie assez pérenne ; et lorsque l'on va vers un mode de stockage qui est de plus en plus élaboré, on arrive à des systèmes qui sont plus performants.

## Media audio



Sur l'audio, c'est un petit peu pareil, et regardez ce qu'est la pérennité du vinyle, et vous ne perdez jamais la totalité de l'information ; visiblement il faut investir dans le vinyle ! A l'inverse, le compact-disque, si vous l'utilisez souvent, c'est exactement comme le cd-rom, il a le même profit, par contre, si vous l'utilisez comme méthode de stockage, il a une bonne durée de vie. Et entre les deux, vous avez tout ce qui n'était pas très bon, comme les cassettes,....

## Media video

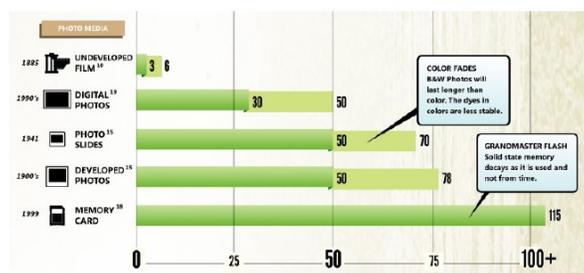


Pour les médias c'est un peu la même chose, pour les DVD, mais ce qui est étonnant, c'est de voir que les super huit ont une durée de vie extraordinaire.

Ce que l'on voit globalement apparaître de ces schémas, c'est qu'il y a un gain technologique et une perte de cette notion de pérennité, c'est-à-dire que l'on est en plein dans la civilisation de consommation, d'utiliser des moyens de stockage que l'on remplace très souvent, ce qui se traduit par des nuisances indirectes que l'on paie !

Passons aux photos :

## Media photo



Pour les photos, ce qui est intéressant, ce sont les « *memory card* », qui sont, plus ou moins, l'équivalent des clés USB, qui ont une grande pérennité, sauf si on les efface et les réécrit. La photo est un domaine dans lequel on voit qu'avec le temps il y a des améliorations ; il faut toujours noter que ce sont des estimations.

Fin de la première partie.