

ARCHITECTURE ET TECHNOLOGIE DES ORDINATEURS

Sylvain TISSERANT

Université de la Méditerranée

Ecole Supérieure d'Ingénieurs de Luminy - Département d'informatique

2003

Introduction

I.1 Architecture des ordinateurs

I.1.a L'informatique aujourd'hui

Apparue au milieu du 20^{ème} siècle, l'informatique a connu et connaît une évolution extrêmement rapide. A sa motivation initiale qui était de faciliter et d'accélérer le calcul, se sont ajoutées de nombreuses fonctionnalités, comme l'automatisation, le contrôle et la commande de pratiquement tout processus, la communication ou le partage de l'information. Dans nos sociétés occidentales elle est omniprésente. Après avoir été un outil réservé aux centres de recherche, elle s'est implantée dans l'industrie et depuis les années 80 elle envahit nos foyers. Au départ nous n'avions que des systèmes centraux, puis sont apparus les postes de travail individuels, très rapidement reliés en réseaux locaux. Tous, ou presque, sont maintenant connectés à la "Toile". Nombreux sont les projets actuels d'utilisation et de gestion d'une puissance de calcul énorme distribuée sur un grand nombre de sites disséminés de par le monde. L'informatique est en grande partie responsable de la profonde transformation que connaît actuellement la civilisation des pays riches.

Les évolutions techniques sont telles que la durée de vie des matériels est relativement courte. Le marché est extrêmement instable et en permanente évolution. La position des constructeurs est presque remise en cause chaque jour. Face à cette situation, après une première période de systèmes propriétaires, dans lesquels un client ne pouvait en pratique s'approvisionner qu'auprès d'un seul constructeur, nous sommes passés aux systèmes ouverts. Cela sous-entend des architectures normalisées pour lesquelles les éléments peuvent provenir de plusieurs constructeurs. En parallèle, du côté logiciel un accent important est mis sur la réutilisation des produits, qui doit permettre de suivre ces évolutions au prix d'un effort moindre, sans avoir à repartir de zéro à chaque passage sur une nouvelle plate-forme par exemple. Cela explique l'émergence des technologies "Orientées Objets" et du "Génie Logiciel". Il est à remarquer que cela se fait souvent avec une perte d'efficacité, masquée par l'évolution des performances au niveau technique.

Face à cette situation très volatile, nous constatons tout de même que les concepts de base introduits par les pionniers perdurent, même s'ils peuvent se décliner de façons très différentes. Pour éviter d'avoir à réécrire ce cours d'Architecture des Ordinateurs tous les ans, j'ai donc pris le parti de principalement développer et expliquer ces concepts et les technologies de base. L'architecture du PC servira parfois d'exemple. Ce choix me semble être aujourd'hui (en l'an 2003) incontournable. Il y a quelques années j'aurais sans doute choisi les architectures RISC des stations de travail. Je ne prendrai pas le risque de prédire quel sera l'exemple à choisir dans cinq ou dix ans.

I.1.b Architecture générale et blocs fonctionnels

Initialement un ordinateur était un calculateur numérique. C'est aujourd'hui une machine de traitement de l'information. Il est capable d'acquérir et de stocker des informations, d'effectuer des traitements et de restituer des informations. L'informatique, contraction d'information et automatique, est devenue la science du traitement de l'information.

Un ordinateur peut être découpé en blocs fonctionnels. Le traitement de l'information est fait au niveau d'un processeur. Les actions que celui-ci doit effectuer sont définies par des instructions. Pour être accessible au processeur les données à traiter et les instructions doivent être stockées dans une mémoire. Le processeur et la mémoire sont reliés par un bus. Par ailleurs il faut que l'utilisateur puisse fournir à l'ordinateur des données et les instructions à suivre, tout comme il doit avoir connaissance des résultats. Il faut donc des dispositifs d'entrée et de sortie.

Chaque bloc fonctionnel peut lui-même être décrit par un ensemble d'unités fonctionnelles. On parle alors de l'architecture du processeur, de l'architecture de la mémoire, etc. Par ailleurs on peut étudier l'architecture d'un réseau rassemblant plusieurs ordinateurs.

On voit ainsi apparaître la notion de niveaux de représentation. A chaque niveau un objet considéré est décrit comme un ensemble de blocs interconnectés. Au niveau suivant chacun de ces blocs est à son tour décrit comme un ensemble de blocs interconnectés et ainsi de suite. Cette décomposition hiérarchique en niveaux dépend des besoins de la conception ou des outils utilisés : elle n'est ni figée, ni unique. Nous pouvons, par exemple, descendre au niveau des portes logiques et même au niveau des transistors qui composent ces portes.

L'architecture d'un ordinateur constitue la représentation de ses unités fonctionnelles et de leurs interconnexions. Le choix d'une architecture est toujours le résultat d'un compromis :

- entre performances et coûts;
- entre efficacité et facilité de construction;
- entre performances d'ensemble et facilité de programmation;
- etc ...

Chaque solution adoptée pour une machine donnée, à un instant donné et pour un marché donné, pourra très vite être remise en cause par un nouveau progrès technologique. De même, des architectures proposées à un moment donné, puis abandonnées pour des raisons techniques ou économiques, peuvent un jour être à nouveau employées.

Quelle que soit sa dimension (mini, micro, super, hyper, etc.) on retrouve toujours dans un ordinateur les mêmes blocs fonctionnels :

- une ou plusieurs unités de traitement;
- de la mémoire;
- un ou plusieurs disques durs, constituant une mémoire permanente;

- des dispositifs de communication avec l'opérateur : un écran, un clavier, très souvent une souris ;
- des dispositifs de communication avec l'extérieur comme une imprimante ;
- des dispositifs permettant l'archivage : disquette, bande magnétique, disque optique;
- etc.

reliés par des bus, des liaisons ou des réseaux. L'informatique se trouve également être embarquée dans un grand nombre de dispositifs de la vie courante, aussi variés que machines à laver, distributeurs de boissons, voitures ou cartes bancaires. Si, par rapport à un ordinateur classique, les tâches à réaliser ne sont pas aussi polyvalentes et si les contraintes sont différentes, on y retrouve cependant les mêmes briques. Partant des portes logiques, nous allons essayer d'étudier dans ce cours les principes de base de chacun de ces composants.

D'autre part, il est important de réaliser que matériel et logiciel sont indissociables, en particulier au niveau de la conception de tout système informatique. Nous rencontrerons dans ce cours un ou deux exemples de l'imbrication du logiciel et du matériel. Pour ce qui concerne le logiciel nous nous limiterons à quelques considérations simples sur le langage machine.

Mais avant d'aborder l'étude des portes logiques, puis des premières unités fonctionnelles remontons dans le temps.

I.2 Un peu d'histoire

I.2.a Le calcul mécanique

Pour résumer l'histoire de l'informatique ou des ordinateurs, il est souvent nécessaire de schématiser. Pour chaque invention, nous ne retenons en général qu'un nom. Mais souvent cette invention concrétise une démarche collective qui s'est étalée sur plus ou moins de temps. Très souvent aussi des prédécesseurs ont été oubliés. Je vais ici résumer l'histoire de l'informatique en m'appuyant sur quelques dates clefs les plus fréquemment citées.

De tout temps l'homme a eu besoin de moyens de calcul. Citons à titre d'exemple le mot même, dont l'étymologie *calculi* signifie cailloux en latin (utilisés sur les abaques des Romains), ou encore la numérotation décimale liée à la technologie de la première calculatrice de poche : la main. Calcul digital signifiait alors compter sur ses doigts.

Par ailleurs, toutes les grandes civilisations de l'Antiquité avaient leurs symboles pour coder les nombres et compter. Au début il s'agissait de signes simples (comme un trait) correspondant chacun au nombre 1. Un symbole représentant le nombre 10 est apparu en Egypte vers 3400 av. J.-C.. Il était ainsi possible d'écrire les nombres de manière un peu plus compacte. D'autres symboles spéciaux ont ensuite représenté les nombres 100, 1000 et 10000. La numération cunéiforme de Babylone utilisait un système sexagésimal (base 60). Le premier système binaire connu remonte à 3000 ans av. J.-C. environ. Il s'agit du Yin et du Yang chinois. On en trouve trace sur les figures magiques de l'empereur Fou-Hi, comme l'octogone à trigramme où les huit premiers nombres sont représentés sous forme binaire :



Figure 1

Les civilisations ont également cherché des outils pour faciliter et accélérer les calculs nécessaires à leur développement. Ce furent par exemple les bouliers en Chine et les abaques dans le Monde Méditerranéen.

Une innovation très importante fut l'emploi de la notation positionnelle, qui confère des valeurs différentes aux symboles numériques selon leur position dans le nombre écrit. Cette notation positionnelle n'est possible qu'avec un symbole pour le zéro. Grâce au symbole 0, il devint possible de différencier 11, 101 et 1 001 sans recourir à des symboles supplémentaires. Cette notation a été introduite en Europe par l'intermédiaire des Arabes, l'Islam s'étendant alors des frontières de la Chine à l'Espagne. Le système dit arabe avait été développé en Inde environ 300 ans av. J-C. Cette introduction s'est faite grâce notamment à la traduction, vers 820 apr. J-C, des ouvrages du mathématicien de Bagdad Al-Khuwarizmi, dont le titre d'un des livres (al jabr) est à l'origine du mot algèbre. Les premiers documents attestant de l'utilisation du système arabe en Europe datent de 976, mais il faut attendre le XIV^{ème} siècle pour qu'il remplace totalement la numérotation romaine. Non seulement l'écriture des nombres devint plus compacte, mais les calculs écrits ont été largement simplifiés.

Il faut ensuite sauter au XVII^{ème} siècle pour voir apparaître les premiers systèmes mécaniques de calcul, basés sur les roues dentées mises en œuvre depuis longtemps dans les horloges, automates, etc. On fait généralement commencer cette nouvelle période en 1614, date à laquelle John Napier (ou Neper), mathématicien écossais, inventa les premiers logarithmes. Il ramenait ainsi les opérations de multiplication et de division à de simples additions ou soustractions. Napier inventa également des machines à multiplier basées sur le déplacement de tiges : les Bâtons ou Os de Napier. Par ailleurs, il fut l'un des premiers (le premier ?) à utiliser le point décimal.

La règle à calcul, basée sur le principe des logarithmes, fut conçue par William Oughtred en 1622. Sa précision sera suffisante pour les calculs scientifiques jusqu'à la première moitié du 20^{ème} siècle. Par contre elle ne peut être utilisée en comptabilité.

En 1623, Wilhelm Schickard inventa pour Kepler une "horloge calculante" destinée à calculer les éphémérides. Il utilisait des roues dentées et il avait abordé le problème du report de retenue. Mais cette machine a disparue en 1624 dans les ravages de la Guerre de Trente Ans en Allemagne du Sud. Elle a été reconstruite en 1960 d'après les plans originaux.

En 1642, à 19 ans, Blaise Pascal construisit, pour aider son père commissaire pour l'impôt à Rouen, une "machine arithmétique" capable de traiter des additions et des soustractions à six chiffres. La Pascaline est considérée comme le premier additionneur à retenue automatique. A

chaque chiffre correspond une roue avec dix crans. Lorsqu'une roue a effectué un tour complet (passage de 9 à 0), elle décale d'un cran la roue supérieure. Chaque chiffre est lu dans une fenêtre au-dessus de la roue. Les multiplications étaient possibles par l'exécution d'additions répétitives, comme l'a montré en 1666 Moreland.

En 1673, Gottfried Leibniz améliora la Pascaline en lui ajoutant un chariot et un tambour à dents inégales. Cela permit d'automatiser, au moyen d'une manivelle, les additions et les soustractions répétitives nécessaires aux multiplications et divisions. Leibniz inventa également le système binaire et montra la simplicité de l'arithmétique binaire. Il l'appliqua à l'analyse des octogones de Fou-Hi.

En 1728, le mécanicien français Falcon construit le premier métier à tisser dont il commande le fonctionnement avec une planchette en bois percée de trous. C'est la première machine commandée par un programme. Ce système fut perfectionné par Joseph-Marie Jacquard en 1805. La planchette de bois fut remplacée par des cartes en carton perforées articulées, comme on peut encore en voir dans les orgues portatifs. Les mécanismes d'exécution et les mécanismes de contrôle étaient séparés. Ce système permettait de reproduire des motifs compliqués avec une qualité égale, mais il supprimait également cinq ouvriers par métier à tisser. Cela déclencha une réaction violente des canuts de Lyon. Il s'agit sans doute d'un des premiers exemples de l'ère industrielle du remplacement d'hommes par des machines automatiques. Cependant dès 1812 on comptait 10000 de ces métiers à tisser.

En 1820 apparaît le premier "Arithmomètre" de Charles Xavier-Thomas de Colmar, sur la base de la machine de Leibniz. Il pouvait exécuter les quatre opérations arithmétiques (addition, soustraction, multiplication, division). La première version ne disposait que de 3 chiffres pour les opérands et 6 pour le totaliseur. Mais cette capacité augmenta progressivement jusqu'à 30 chiffres. Sa facilité d'emploi assura son succès jusqu'à la Première Guerre Mondiale.

Jusqu'alors les calculateurs mécaniques étaient capables d'exécuter des opérations une à une. Dans les années 1820 et 1830, un mathématicien anglais, Charles Babbage, rapprocha les machines à calculer et les systèmes de commande de Jacquard dans le but d'effectuer des calculs complexes nécessitant l'exécution en séquence de plusieurs opérations arithmétiques. Il a d'abord tenté de réaliser une "machine à différences" destinée au calcul de tables de navigation ou de tir. Il parviendra à construire un prototype partiel. Mais les difficultés techniques ralentissent son projet, qu'il devra abandonner après une dizaine d'années de travail, faute de moyens financiers suffisants. Il eut alors l'idée d'une machine plus universelle : la "machine analytique" capable d'enchaîner l'exécution d'opérations arithmétiques en fonction d'instructions données par l'utilisateur. Cette machine contenait les concepts de base qui se retrouvent dans les ordinateurs modernes : une unité de traitement (le moulin) commandée par une unité de contrôle, une mémoire (le magasin), une unité d'entrée pour recevoir des cartes perforées (opérations et nombres) et une unité de sortie pour imprimer ou perforer le résultat. Charles Babbage a été aidé dans la conception de cette machine par sa collaboratrice Ada Lovelace (fille du poète Lord Byron). Elle a en particulier défini le principe de l'enchaînement d'itérations successives pour la réalisation d'une opération, baptisé algorithme en l'honneur du mathématicien arabe Al-Khuwarizmi. Cette machine était cependant trop ambitieuse pour la technologie de l'époque. Un exemplaire fut partiellement construit par son fils en 1888.

En 1854 George Boole propose sa formulation mathématique des propositions logiques qui appliquée au système binaire est à la base du fonctionnement des ordinateurs.

I.2.b Le calcul électromécanique

En 1890, Hermann Hollerith construisit un calculateur de statistiques qui fut utilisé pour le recensement américain. Il s'agissait d'une machine électromécanique plus performante que les machines mécaniques. Il développa à cette occasion la carte perforée et inventa le système de codage des informations qui porte son nom. La détection de la présence ou de l'absence d'un trou se faisait au moyen d'aiguilles qui traversaient les trous et fermaient chacune un circuit électrique en trempant dans un godet de mercure. Par ailleurs la plupart des questions du recensement étaient posées sous forme binaire : premier exemple de l'imbrication de l'architecture matérielle et de l'organisation des données. Hollerith fonda en 1896, la Tabulating Machine Company pour produire ses cartes et machines mécanographiques. En 1924, celle-ci devint l'International Business Machines Corporation : IBM.

En 1914, Leonardo Torres y Quevedo proposa de réaliser une version électromécanique de la machine de Babbage. Faute de moyens ce fut un échec.

En 1930, Vannemar Bush construisit au MIT un analyseur différentiel, capable de résoudre des équations différentielles rencontrées dans l'étude des circuits électriques. La même année, Georges Stibitz réalisa un additionneur binaire à relais, le "Modèle K", s'appuyant sur les travaux de Boole.

Les années 30 virent également de nombreux progrès sur le plan théorique. Par exemple, en 1936 Alan Turing énonce le principe d'une machine (virtuelle), qui pouvait en théorie réaliser tous les calculs mathématiques, avec en particulier des instructions conditionnelles. En 1938, dans sa thèse, Claude Shannon reprenant les travaux de Leibniz et Boole, mis en évidence une analogie entre l'algèbre binaire et les circuits électriques. Il montra plus tard que tout calcul logique ou arithmétique peut être réalisé avec uniquement les trois opérations logiques fondamentales : ET, OU et NON. Il est également l'inventeur du terme bit pour binary digit pour désigner le chiffre binaire.

En 1938, Konrad Zuse créa avec des moyens modestes un ordinateur binaire programmable mécanique, le Z1. En 1939, il perfectionna sa machine en remplaçant une partie des pièces mécaniques par des relais électromécaniques (Z2). Le Z3 et le Z4, virent le jour en 1941 et furent utilisés pour des calculs aéronautiques. En 1939, John Atanasoff et Clifford Berry réalisèrent un additionneur binaire à 16 bits. Ils furent les premiers à utiliser des tubes à vide.

Parmi les derniers calculateurs électromécaniques il nous faut citer le Mark 1 ou ASCC (Automatic Sequence Controlled Calculator), développé en 1941 conjointement par IBM et Harvard, sous la direction de Howard Aiken. Il s'agissait d'une énorme machine de 5 tonnes, couvrant 25 m² et consommant 25 kW. Il comptait 3000 relais et 760000 pièces mécaniques. Le programme était lu sur une bande de papier, les données à partir d'une seconde bande de papier ou d'un lecteur de cartes. A l'origine les sauts conditionnels étaient impossibles. Par la suite il fut équipé de plusieurs lecteurs de cartes pour permettre ces sauts conditionnels et l'exécution de

sous-programmes. En 1945, un insecte (bug) coince un relais, provoquant un dysfonctionnement de cette machine analytique.

I.2.c L'ère de l'électronique

Egalement en 1941, John Atanasoff et Clifford Berry construisirent le premier ordinateur binaire à lampes : l'ABC (Atanasoff-Berry Computer). Il comportait une mémoire de 60 mots de 50 bits et une unité arithmétique et logique. Bien que le programme ne fut pas stocké en mémoire, l'ABC est souvent considéré comme le premier vrai ordinateur.

L'année 1945 vit la mise en service de l'ENIAC (Electronic Numerical Integrator And Calculator) proposé en 1942 par Presper Eckert et John Mauchly. Il était composé de 19000 tubes, 1500 relais, consommait 170 kW, pesait 30 tonnes et couvrait une surface de 72 m². Il était environ 500 fois plus rapide que le Mark 1 (environ 330 multiplications par seconde). Cependant sa programmation s'effectuait à l'aide de fiches à brancher dans un tableau de connexions. Le travail de programmation pouvait prendre plusieurs jours. A la fin de cette même année, John von Neumann, consultant sur l'ENIAC, proposa de coder le programme sous forme numérique et de l'enregistrer en mémoire, avec un gain en souplesse et rapidité. Il jetait ainsi les bases de l'architecture de l'ordinateur moderne. Il ne fut cependant pas le premier à réaliser cette machine. Des querelles entre Eckert, Mauchly et von Neumann portant notamment sur la paternité de l'idée et des brevets retardèrent le projet EDVAC (Electronic Discret VARIable Computer). Celui-ci fut devancé en 1948 par un prototype, appelé Manchester Mark 1, puis en 1949 par l'EDSAC (Electronic Delay Storage Automatic Computer) première machine opérationnelle construite par Maurice Wilkes.

En 1948, William Shockley, John Bardeen et Walter Brattain inventèrent le transistor bipolaire. Celui-ci remplaça rapidement les lampes apportant fiabilité et rapidité aux ordinateurs de seconde génération.



Figure 2

La taille et la consommation diminuèrent sensiblement. Le premier ordinateur utilisant des transistors fut le TRADIC en 1955. A la même époque, IBM commercialisa le premier disque dur (5 disques de 61 cm de diamètre pour 5 Mo). Les mémoires à tore de ferrite étaient également disponibles. Le PDP-8 de DEC fut le premier mini-ordinateur à être diffusé en grande série (50000 exemplaires).

La troisième génération, à partir de 1970, correspond à l'utilisation des circuits intégrés. L'Intel 4004, le premier microprocesseur 4 bits, fut en 1971 le premier circuit intégré incorporant unité de calcul, mémoire et gestion des entrées-sorties. Il comportait 2300 transistors. Le 8008 sortit un an plus tard. Le premier micro-ordinateur, le Micral N, fut construit en 1973 par une entreprise française R2E. Depuis le début des années 1980, le niveau d'intégration des puces ayant largement progressé en surface et en densité, on parle de quatrième génération.

I.3 Quelques généralités sur les architectures

I.3.a Architecture de Von Neumann

Cette architecture est caractérisée par un processeur et une mémoire reliés par un bus :

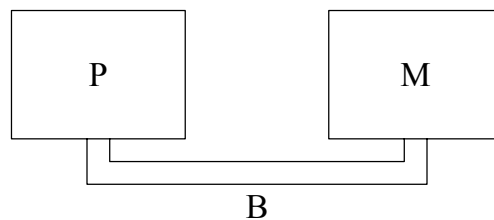


Figure 3

Les instructions et les données sont stockées dans la mémoire. Pour accéder à une information contenue en mémoire le processeur affiche sur le bus l'adresse de celle-ci. Après un certain temps le contenu demandé est affiché par la mémoire sur le bus.

On constate immédiatement que la rapidité d'une machine sera limitée par l'élément le plus lent. Il ne sert à rien de fabriquer un processeur très rapide si la mémoire n'est pas capable de lui fournir à un rythme suffisant les instructions à exécuter, ni les données à traiter. De manière générale lorsqu'on améliore les performances d'un élément de la machine il faut s'assurer que les autres éléments qui lui sont connectés seront capables de supporter des débits plus élevés. Dans le cas contraire il peut être favorable de revoir l'architecture de la machine. L'efficacité d'un ordinateur est directement liée au débit des informations (instructions et données) qui circulent entre les différents éléments. La conception d'une architecture a pour but d'optimiser ce débit, en prenant en compte des contraintes financières, commerciales et techniques.

La machine de base, dite de Von Neumann, que nous venons de décrire ne présenterait aucun intérêt sans au moins un dispositif d'entrées/sorties pour permettre une communication

avec l'extérieur. Pour une machine aussi simple il existe au moins trois méthodes pour installer un tel dispositif :

- Interface directement reliée au processeur :

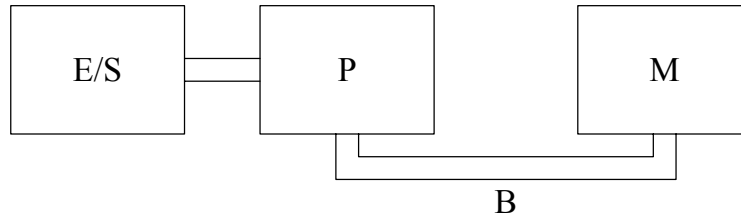


Figure 4

La communication est réalisée à l'aide d'instructions spécialisées ou par l'intermédiaire de registres spécialisés.

- Interface connectée au bus commun :

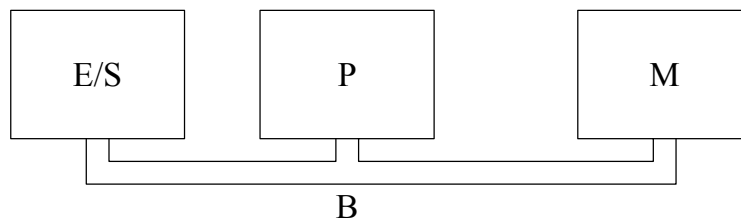


Figure 5

Un fil spécialisé du bus indique si le processeur s'adresse à la mémoire ou à l'interface d'E/S.

- Interface vue par le processeur comme une partie de la mémoire :

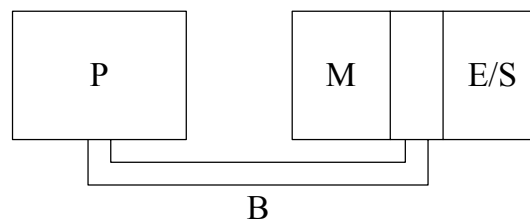


Figure 6

L'interface est sur le bus commun, le choix de l'adresse indique si le processeur s'adresse à la mémoire ou à l'interface. Si sur le plan du matériel ces deux solutions peuvent être très différentes, sur le plan du concept elles sont comparables : un fil supplémentaire correspond de fait à un bit. Dans les deux cas la mémoire et le dispositif d'entrées/sorties sont identifiés par des espaces d'adressage différents.

I.3.b Architecture de Harvard

Dans cette architecture on sépare physiquement la mémoire des instructions et la mémoire de données. Chacune de ces mémoires est accessible par un bus différent. Il est alors possible d'accéder simultanément aux instructions et aux données, ce qui permet d'augmenter le flux des informations. En première approximation il est donc possible de multiplier la vitesse de la machine par deux, en multipliant par deux la vitesse du processeur tout en conservant le même type de mémoire. En pratique il peut être moins onéreux de mettre deux bus dans une machine que de doubler la vitesse de la mémoire.

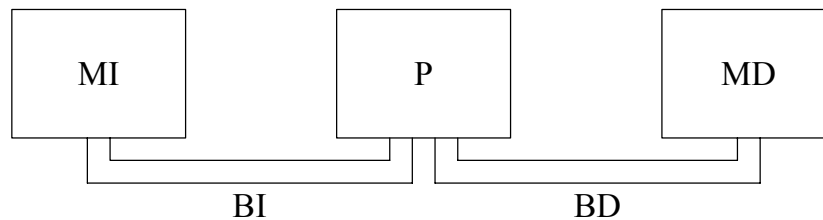


Figure 7

Cette architecture a été expérimentée très tôt, puis abandonnée pendant une vingtaine d'années. Elle est à nouveau employée dans certains micro-ordinateurs. Elle est également reprise par l'intermédiaire de deux caches mémoires dans l'architecture de microprocesseurs. Le 68030 fut l'un des premiers à disposer de deux caches consacrés l'un aux instructions et l'autre aux données.

Chercher à pouvoir effectuer plusieurs opérations (ici transferts des instructions et des données) en parallèle est un principe très souvent mis en œuvre en architecture des ordinateurs. Par exemple, pour permettre aux interfaces d'E/S de lire ou d'écrire des données en mémoire pendant que le processeur effectue des calculs sur les données précédentes, il est intéressant d'utiliser une mémoire de données accessible depuis deux ports différents.



Figure 8

De cette façon le processeur et l'interface peuvent chacun accéder alternativement à une moitié différente de la mémoire :

Phase 1 : le processeur lit des données provenant de la première moitié de la mémoire pendant que l'interface d'Entrées/Sorties range de nouvelles données dans la seconde moitié.

Phase 2 : le processeur accède aux données transférées par l'interface lors de la phase 1 et l'interface écrit dans la première moitié de la mémoire.

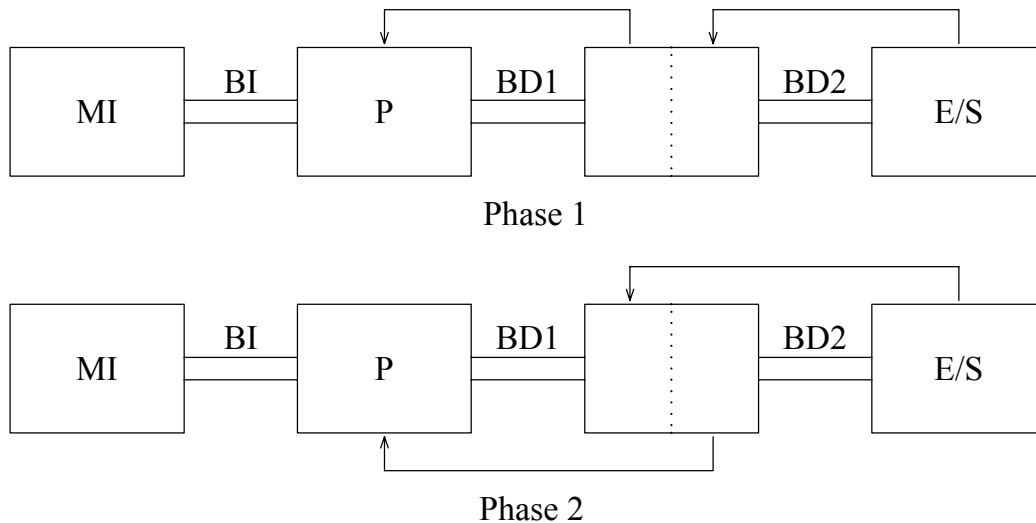


Figure 9

I.3.c Processeur

Sans développer ici les principes de fonctionnement d'un processeur, nous pouvons déjà dire que l'exécution des instructions peut se découper en grandes étapes :

- chargement de l'instruction à exécuter;
- décodage de l'instruction;
- localisation dans la mémoire des données utilisées par l'instruction;
- chargement des données si nécessaire;
- exécution de l'instruction;
- sauvegarde des résultats à leurs destinations respectives;
- passage à l'instruction suivante.

A chacune de ces étapes peut correspondre une unité fonctionnelle. Cette séquence simple peut se compliquer, en particulier si la machine possède des instructions de longueur variable : certaines instructions étant codées sur un mot de la mémoire et d'autres sur plusieurs. Dans ce cas l'unité chargée de la recherche et du chargement de l'instruction doit être capable de reconnaître (ou de pré-décoder) le type de l'instruction pour déterminer s'il lui faut aller chercher la suite. Dans la plupart des processeurs, l'unité centrale de traitement se décompose en :

- l'unité de commande qui contient un dispositif de décodage des instructions (décodeur) et un séquenceur qui contrôle les circuits nécessaires à l'exécution de l'instruction en cours;

- l'unité arithmétique et logique (UAL ou ALU : Arithmetic and Logical Unit) qui exécute des opérations arithmétiques comme l'addition, la soustraction, des décalages, des opérations booléennes (ET, OU, etc...), des comparaisons, etc...
- les registres : mémoire locale très rapide qui permet de stocker des résultats temporaires ou des informations de commande. Parmi ceux-ci on distingue les registres arithmétiques qui servent aux opérations arithmétiques. D'autres registres ont des fonctions particulières. On trouve par exemple :
 - le compteur ordinal (CO) qui pointe sur la prochaine instruction à exécuter;
 - le registre d'instruction (RI) qui contient l'instruction en cours d'exécution;
 - le registre d'état (PSW : Processor Status Word) qui indique l'état du système : dépassement, retenue, etc...
 - une pile et un pointeur de pile (SP : Stack Pointer);
- des chemins de données.

Toutes ces unités fonctionnent au même rythme, à une cadence imposée par une horloge, généralement externe à l'unité centrale. A chaque cycle d'horloge, chaque unité va ouvrir ou fermer certaines portes pour déplacer, lire, écrire, comparer, additionner des bits. Ceci se fait en fonction d'ordres donnés par l'unité de contrôle. Ces ordres dépendent évidemment de l'instruction à exécuter.

Le tableau suivant rappelle quelques unités utilisées pour exprimer la fréquence de fonctionnement d'un processeur et la durée correspondante des cycles. Ainsi un processeur fonctionnant à 250 MHz a un cycle de 4 ns.

Fréquence	Préfixe	Hz	Cycle	Préfixe	s
1 kHz	kilo	10^3	1 ms	mili	10^{-3}
1 MHz	méga	10^6	1 μ s	micro	10^{-6}
1 GHz	giga	10^9	1 ns	nano	10^{-9}

Table 1

Nous verrons qu'une manière d'améliorer les performances d'une unité centrale est de faire fonctionner ses diverses unités fonctionnelles en parallèle. Cela signifie aller chercher l'instruction suivante alors que le décodeur travaille sur l'instruction en cours. Il y a anticipation. Il faut pour pouvoir profiter de cette technique il faut que les instructions soient exécutées en séquence et donc éviter les ruptures. Il faut donc que la programmation soit adaptée à l'architecture des machines. Il peut également être possible d'exécuter la même instruction simultanément sur plusieurs données. Il suffit d'une seule unité de commandes associée à plusieurs unités de traitement.

I.3.d Mémoires

Nous appelons mémoire tout dispositif capable de stocker des informations (instructions et données) de telle sorte que l'organe qui les utilise puisse à n'importe quel moment accéder à l'information qu'il demande.

Les informations peuvent être écrites ou lues. Il y a écriture lorsqu'on enregistre des données en mémoire, lecture lorsqu'on sort des informations précédemment enregistrées. La lecture peut être destructive (l'information lue n'est plus en mémoire) ou non. Le temps d'accès est le temps qui s'écoule entre l'instant où a été lancée une opération de lecture en mémoire et l'instant où la première information est disponible. Le temps de cycle représente l'intervalle minimum qui doit séparer deux demandes successives de lecture ou d'écriture. Le temps de cycle est égal ou supérieur au temps d'accès. On appelle cadence de transfert ou débit d'une mémoire, le nombre maximum d'informations lues ou écrites par unité de temps.

Une mémoire est formée d'un certain nombre de cellules, ou cases, contenant chacune une information. Chaque cellule a un numéro qui permet de la référencer et de la localiser. Ce numéro est son adresse. Avec une adresse de n bits il est possible de référencer directement au plus 2^n cellules. La capacité d'une mémoire est le nombre total de cellules qu'elle contient. Elle s'exprime en nombre de bits, d'octets (bytes) ou de mots (words). Compte tenu de l'adressage binaire, les capacités des mémoires s'expriment en puissances de deux ou en multiples de $2^{10} = 1024$. La table 2 explicite les unités utilisées.

Symbole	Préfixe	Décimal	Binaire	
1 k	kilo	10^3	$2^{10} =$	1024
1 M	méga	10^6	$2^{20} =$	1048576
1 G	giga	10^9	$2^{30} =$	1073741824
1 T	téra	10^{12}	$2^{40} =$	1099511627776

Table 2

Dans une mémoire à semi-conducteur, on accède directement à n'importe quelle information dont on connaît l'adresse, le temps pour obtenir l'information ne dépend pas de l'adresse. On dira que l'accès à une telle mémoire est aléatoire, direct ou encore sélectif. A l'inverse pour accéder à une information sur bande magnétique il faut dérouler la bande en repérant tous les enregistrements jusqu'à ce que l'on trouve celui que l'on adresse. On dit alors que l'accès à l'information est séquentiel. Le temps d'accès est variable selon la position de l'information recherchée. L'accès peut encore être semi-séquentiel : combinaison des accès direct et séquentiel. Pour un disque magnétique par exemple l'accès à la piste est direct, puis l'accès au secteur est séquentiel. Les mémoires associatives permettent l'accès par le contenu.

L'information stockée est volatile si elle risque d'être altérée par un défaut d'alimentation électrique et non volatile dans le cas contraire.

En ce qui concerne les mémoires le problème essentiel du constructeur est celui du prix. En effet à un instant donné ce sont toujours les mémoires d'accès le plus rapide qui sont les plus onéreuses et en général le prix augmente plus vite que la vitesse. Or, comme nous l'avons vu, la vitesse d'accès à la mémoire conditionne, dans une large mesure les performances globales d'une machine. C'est pourquoi les ordinateurs actuels sont dotés de plusieurs types de mémoires de tailles et de vitesses différentes.

On trouve d'abord des mémoires très rapides (les plus chères), en petit nombre, accessibles immédiatement par le processeur. Ce sont les registres. On trouve ensuite le cache ou anté-mémoire. Accessible en quelques cycles d'horloge (1 ou 2 en général) il contient quelques dizaines de milliers de mots. On a ensuite la mémoire principale ou mémoire centrale. On dispose parfois d'une mémoire d'appui ou d'arrière-plan, intermédiaire entre la mémoire centrale et les mémoires auxiliaires. Elle est constituée de circuits semi-conducteurs de faible prix, dont la vitesse est faible pour des circuits de mémoire mais cent fois plus élevée que celle des disques magnétiques. Toutes ces mémoires sont volatiles.

Les disques magnétiques forment le niveau suivant (mémoires auxiliaires ou mémoires de masse) et constituent une mémoire permanente dont les données restent valides quand on coupe l'alimentation électrique. Le dernier niveau correspond à l'archivage des données. Jusqu'à présent ce domaine était l'exclusivité des bandes magnétiques. Ces dernières années ont vu apparaître un nouveau produit qui peut dans certains cas fournir une alternative pour l'archivage. Il s'agit du disque optique numérique. Celui-ci a des densités de stockage de 10^2 à 10^4 fois supérieures à celles des supports magnétiques, pour des temps d'accès cinq à dix fois plus grand.

La figure suivante schématise cette hiérarchie des mémoires :

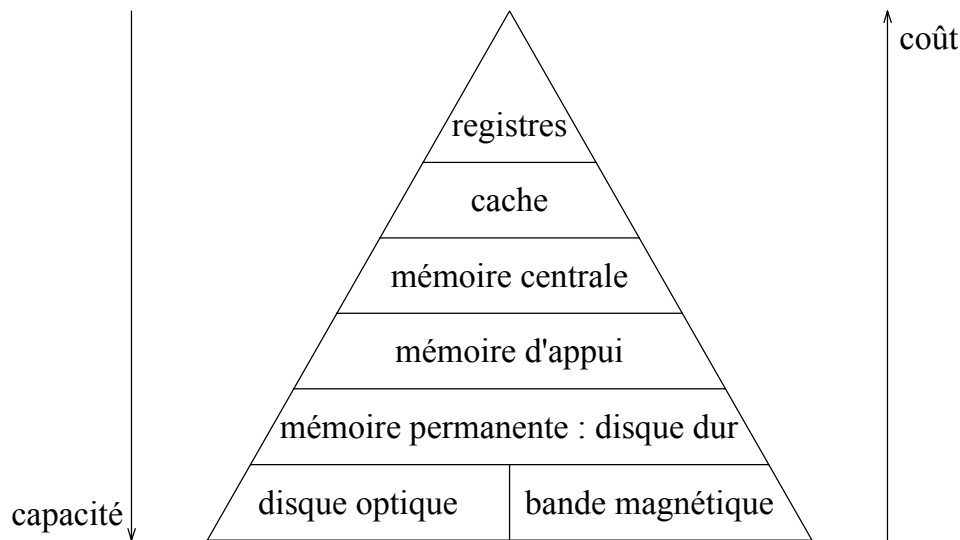


Figure 10

I.3.e Bus : le squelette des ordinateurs

Un bus est un moyen de communication entre les différents éléments constituant une machine. Il s'agit en général d'un ensemble de fils électriques. Les différents éléments sont reliés au bus par des connecteurs, dont l'ensemble constitue un fond de panier. Chaque emplacement (slot) du fond de panier (backplane) peut recevoir une carte électronique (board). Sur un bus circulent différents types de signaux : adresses, données, contrôle, alimentations, etc...

I.3.f Entrées/Sorties

Pour pouvoir exécuter une tâche, l'ordinateur doit disposer du programme et des données (s'il y a lieu) qui s'y réfèrent. Dès que le travail est accompli, les résultats du traitement sont communiqués à l'utilisateur. Les techniques d'échange d'informations entre l'ordinateur et son environnement externe sont appelées techniques d'Entrées/Sorties (E/S ou I/O pour Input/Output). L'ordinateur échange des informations non seulement avec des opérateurs humains, par l'intermédiaire de terminaux de communication (consoles avec clavier, souris, imprimantes, lecteurs, etc...), mais également avec des dispositifs variés : appareils de mesures, capteurs, etc...

