

CPU

Le CPU (*Central Processing Unit*) appelé aussi processeur ou microprocesseur est le cerveau du PC.

Sa puissance dépend

- du nombre de cœurs du processeur,
- de la taille de sa mémoire cache
- de l'architecture du processeur
- et de sa fréquence

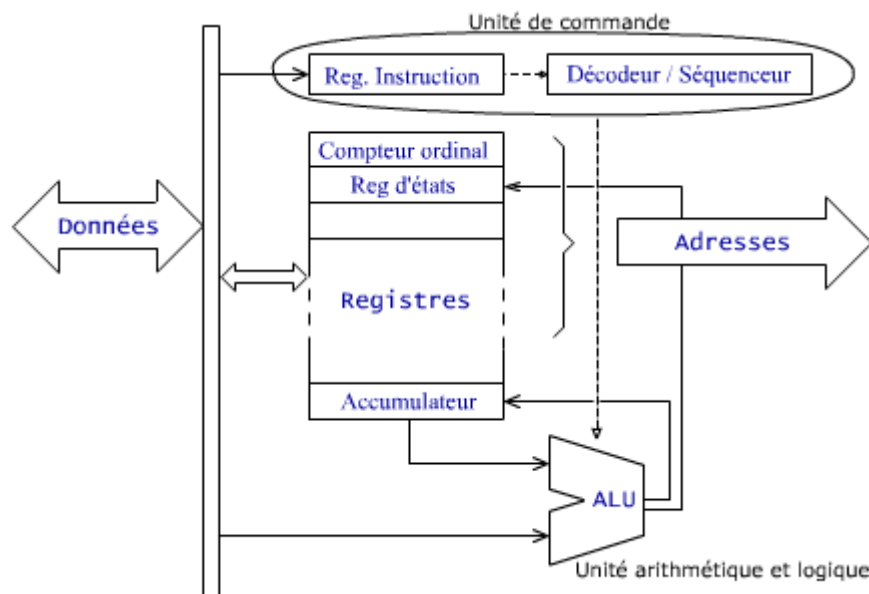
La **fréquence** du CPU s'exprime en MHz ou en GHz. C'est une valeur à laquelle les utilisateurs attachent le plus d'importance mais d'autres paramètres en déterminent les performances.

L'efficacité du CPU dépend aussi de son **architecture** interne. Alors qu'il fallait dans les premiers PC environ 12 cycles pour exécuter une instruction, les processeurs actuels traitent plusieurs instructions par cycle. Ces traitements se font en parallèle grâce notamment à la technique du double **pipeline** d'instructions, ce qui leur permet d'effectuer deux ou trois instructions par cycle. L'**hyperthreading** est une autre technique encore pour faire le traitement simultané de plusieurs flux d'instructions, les « *thread* ».

Le rôle de la **mémoire cache** est d'autant plus important que la vitesse du CPU est élevée par rapport à celle de la mémoire centrale. Cet écart entre la vitesse du processeur et celle des RAM ne cesse d'augmenter car l'évolution des CPU est bien plus rapide que celle des mémoires. La mémoire cache coûte cher. Sa taille explique souvent la différence de prix entre les processeurs.

L'augmentation des fréquences a atteint ses limites aux environs de 4 GHz. La seule façon d'augmenter encore les performances est maintenant de **multiplier les cœurs** sur la même puce.

Description



Nous avons vu que du CPU partent des ensembles de fils appelés «bus» auxquels sont connectés les autres composants du système.

Le CPU (mono cœur) est constitué essentiellement de trois parties :

- ❑ L'unité de commande qui cherche les instructions en mémoire, les décode et coordonne le reste du processeur pour les exécuter. Une unité de commande élémentaire se compose essentiellement d'un registre d'instruction et d'une unité «décodeur / séquenceur»
- ❑ L'unité Arithmétique et Logique (ALU) exécute les instructions arithmétiques et logiques demandées par l'unité de commande. Les instructions peuvent porter sur un ou plusieurs opérandes. La vitesse d'exécution est optimale quand les opérandes se situent dans les registres plutôt que dans la mémoire externe au processeur.
- ❑ Les registres sont des cellules mémoire interne au CPU. Ils sont peu nombreux mais d'accès très rapide. Ils servent à stocker des variables, les résultats intermédiaires d'opérations (arithmétiques ou logiques) ou encore des informations de contrôle du processeur.

La structure des registres varie d'un processeur à l'autre. C'est ce qui fait que chaque type de CPU a un jeu d'instruction qui lui est propre. Leurs fonctions de base sont néanmoins semblables et tous les processeurs possèdent en gros les mêmes catégories de registres.

L'**accumulateur** est principalement destiné à contenir les données qui doivent être traitées par l'ALU.

Les **registres généraux** servent au stockage de résultats intermédiaires

Les **registres d'adresses** servent à confectionner des adresses de données particulières, Ce sont, par exemples, les registres de base et d'index qui permettent entre autre d'organiser les données en mémoire comme des tables indicées.

Le **registre d'instruction** contient le code de l'instruction qui est traitée par le décodeur / séquenceur.

Le **compteur ordinal** ou *program counter* contient l'adresse de la prochaine instruction à exécuter. En principe, ce registre ne cesse de compter. Il génère les adresses des instructions à exécuter les unes à la suite des autres. Certaines instructions demandent quelquefois de changer le contenu du compteur ordinal pour faire une "rupture de séquence" c'est à dire un saut ailleurs dans le programme.

Le **registre d'état** appelé parfois **registre de condition** contient des indicateurs appelés flags (drapeaux) et dont les valeurs (0 ou 1) varient en fonction des résultats des opérations arithmétiques et logiques. Ces états sont utilisés par les instructions des sauts conditionnels.

Le **pointeur de pile** ou *stack pointer* gère certaines données en mémoire en les organisant sous forme de piles.

Principe de fonctionnement du CPU

Le contenu du compteur de programme est déposé sur le bus d'adressage pour y rechercher une instruction en code machine. Le bus de contrôle produit un signal de lecture et la mémoire, sélectionnée par l'adresse, renvoie le code de l'instruction au processeur via le bus des données.

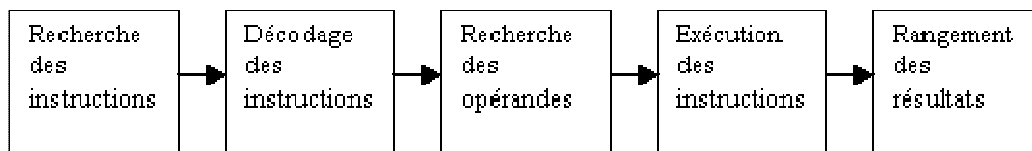
Une fois que l'instruction aboutit dans le registre d'instruction, l'unité de commande du processeur la décode et produit la séquence appropriée de signaux internes et externes qui coordonnent son exécution. Une instruction comporte une série de tâches élémentaires. Elles sont cadencées par les cycles d'horloge.

Toutes les tâches qui constituent une instruction s'exécutent les unes à la suite des autres. L'exécution d'une instruction dure donc plusieurs cycles. Comme il n'est pas toujours possible d'augmenter la fréquence, la seule manière d'accroître le nombre d'instructions traitées en un temps donné est de chercher à en exécuter plusieurs simultanément. On y arrive en fractionnant les ressources du processeur, les données et/ou les processus. C'est ce qu'on appelle la **parallélisation**.

Pipelines d'instructions et architecture superscalaire

L'unité de commande élémentaire qui, dans les premiers microprocesseurs, contenait un registre d'instruction et un décodeur / séquenceur, est remplacé depuis les processeurs de la troisième génération (386) par un ou des pipelines. Dans le 486, le pipeline est unique et se compose de cinq unités fonctionnelles.

Elles "travaillent à la chaîne" pour effectuer simultanément les étapes successives de plusieurs instructions.



La première unité fonctionnelle du pipeline va chercher les instructions en mémoire et les range dans un tampon, la seconde unité décode l'instruction et ainsi de suite.

Globalement, il faut toujours autant de temps pour exécuter une instruction mais la vitesse d'exécution est multipliée par cinq puisque avec ce système les instructions suivantes sont entamées dès que possible, les instructions se font donc en parallèle.

Les premiers Pentium (cinquième génération) ont deux pipelines en parallèle ce qui leur permet d'accomplir deux instructions par cycles.

Dans les Pentium actuels, l'architecture «superscalaire» multiplie les chemins des instructions au niveau de certains étages du pipeline, les unités fonctionnelles d'exécutions ont alors recours à plusieurs Unité Arithmétiques et Logiques (ALU). Le processeur est dit **superscalaire de rang "n"** s'il possède n unités arithmétiques et logiques.

Le pipeline du 486 perdait de son efficacité lors des ruptures de séquences. Lors des sauts conditionnels, les instructions qui suivent le saut et dont le traitement a été entamé doivent parfois être abandonnées après l'évaluation du test. En cas d'erreur, la perte de temps est d'autant plus importante que le pipeline est plus long.

A partir du Pentium, une **unité de prédiction de branchement**, s'attache à évaluer la suite d'instructions qui sera la plus probable. Lors des tests conditionnels, elle mémorise le résultat du test pour en tenir compte au passage suivant selon des algorithmes qui au fil de l'évolution des processeurs devient de plus en plus élaborés. Cette méthode est efficace notamment pour les tâches répétitives exécutées par des boucles de programme.

Hyper-threading

L'hyperthreading est une technique qui consiste à permettre l'exécution simultanée de plusieurs *threads*. On traduit généralement le terme *thread* par « fils d'exécution » ou « sous processus ». Il s'agit d'une partie d'un processus dont l'exécution peut être indépendante du reste de l'application. Le correcteur orthographique d'un traitement de texte en est l'exemple le plus facile à imaginer.

L'HTT (*Hyper-Threading Technology*) est apparu chez Intel avec le Pentium Xeon, il est disponible sur les Pentium 4 qui tournent à au moins 3,6 GHz. L'hyperthreading était selon la rumeur, présent dans les premiers Pentium 4 mais Intel l'aurait par prudence mi hors d'usage avant d'être absolument sûr de son bon fonctionnement.

On ne peut tirer avantage de l'hyperthreading que si les applications, le système d'exploitation, le processeur le BIOS et le Chipset sont prévus pour.

Certains éléments du processeur qui caractérisent son état sont dédoublés de sorte à permettre l'exécution concurrente de deux jeux d'instructions et d'éviter ainsi les temps morts. Ces temps morts sont par exemple dus à l'absence d'une donnée utile dans la cache (*cache miss*) ou à une erreur de prédiction de branchement qui annihile la séquence d'instructions entamée dans le pipeline.

Le processeur physique agit alors comme deux processeurs virtuels aussi dits "processeurs logiques" qui apparaissent dans le gestionnaire de périphérique comme deux processeurs distincts.

Le gain de performance annoncé par Intel va actuellement jusqu'à 30%. Il est optimum quand les logiciels sont développés dans l'optique de l'Hyper-Threading.

Processeurs multi-cœurs

Une autre manière de favoriser l'exécution simultanée de plusieurs processus ou de plusieurs threads est de placer plusieurs processeurs, on dira plusieurs cœurs, dans le même boîtier. Cette technique convient particulièrement au multitâche puisqu'avec elle, il y a réellement autant de tâches qui peuvent s'exécuter en parallèle que de cœurs dans le processeur. Les applications pour bénéficier pleinement du multi-cœur doivent avoir été repensées pour pouvoir se subdiviser en tâches parallèles.