

**HELHa**

Haute École  
Louvain en Hainaut



Bachelier en Informatique de Gestion  
1ère année

# Structure des Ordinateurs

Mathieu Philippe

2009-2010 Édition n° 1

<b>INTRODUCTION HISTORIQUE : LES PRECURSEURS. ....</b>	<b>9</b>
<b>PRESENTATION GENERALE D'UN ORDINATEUR. ....</b>	<b>10</b>
L'UNITE CENTRALE ET LA MEMOIRE CENTRALE. ....	10
LES UNITES D'ECHANGE ET LES UNITES PERIPHERIQUES. ....	12
<b>PRINCIPE DU FONCTIONNEMENT D'UN ORDINATEUR PRESENTATION DE BOULIX. ....</b>	<b>12</b>
LES REGISTRES. ....	13
LA MEMOIRE CENTRALE. ....	13
LE PROGRAMME. ....	14
L'UNITE ARITHMETIQUE ET LOGIQUE. ....	14
L'UNITE DE CONTROLE. ....	15
DEROULEMENT D'UNE INSTRUCTION PORTANT SUR UN OPERANDE EN MEMOIRE. ....	16
INSTRUCTION DE RUPTURE DE SEQUENCE. ....	18
L'UNITE D'ECHANGE. ....	18
.....	19
LES UNITES PERIPHERIQUES. ....	19
LES INTERRUPTIONS. ....	19
CONFIGURATION D'UN SYSTEME INFORMATIQUE. ....	20
<b>QUELQUES ORDRES DE GRANDEUR. ....</b>	<b>20</b>
<b>NOTIONS DE HARDWARE ET DE SOFTWARE. ....</b>	<b>21</b>
<b>HISTORIQUE DES MACHINES LES GENERATIONS D'ORDINATEURS. ....</b>	<b>23</b>
L'EVOLUTION DE LA TECHNOLOGIE. ....	23
L'EVOLUTION DE L'EXPLOITATION DES ORDINATEURS. ....	25
<b>RETOUR SUR LE ROLE DU SYSTEME D'EXPLOITATION. ....</b>	<b>27</b>
<b>L'UTILISATION DES ORDINATEURS. ....</b>	<b>28</b>
LE CALCUL SCIENTIFIQUE. ....	29
LA GESTION. ....	29
LA CONDUITE DE PROCESSUS. ....	29
LES SYSTEMES CONVERSATIONNELS. ....	29
<b><u>CHAPITRE II - STRUCTURE DE L'INFORMATION DIGITALE. ....</u></b>	<b><u>30</u></b>
<b>L'INFORMATION DIGITALE. ....</b>	<b>31</b>
L'INFORMATION DIGITALE ELEMENTAIRE. ....	31
NOTION DE CODAGE. ....	31
NOTION DE CODE REDONDANT. ....	32
TAILLE DES INFORMATIONS DANS UN CALCULATEUR. ....	34
<b>LA REPRESENTATION DES NOMBRES. ....</b>	<b>35</b>
RAPPEL SUR LES SYSTEMES DE NUMERATION. ....	35
LE SYSTEME BINAIRE. ....	36
EN MARGE DU SYSTEME BINAIRE. ....	38
NUMERATION OCTALE ET HEXADECIMALE. ....	47
REPRESENTATION DES NOMBRES BINAIRES NEGATIFS. ....	48
LE FORMAT DES NOMBRES EN MACHINE. ....	51
<b>CODAGE DES INFORMATIONS NON NUMERIQUES. ....</b>	<b>54</b>
CODAGE DES CARACTERES. ....	55
CODAGE DES INSTRUCTIONS. ....	60
<b>L'ORGANISATION ET LA RECHERCHE DES INFORMATIONS EN MEMOIRE. ....</b>	<b>61</b>
NOTION DE TABLEAU, DE LISTE ET DE POINTEUR. ....	61
NOTION DE TABLE. ....	62
NOTION DE PILE ET DE FILE D'ATTENTE. ....	62
NOTION DE PREFIXE. ....	63

<b>ELEMENTS DE LANGAGE MACHINE.....</b>	<b>63</b>
FORME GENERALE DES INSTRUCTIONS ARITHMETIQUES. ....	63
LE JEU D'INSTRUCTIONS D'UN CALCULATEUR. ....	65
LES DIFFERENTS TYPES D'INSTRUCTIONS. ....	66
<b>NOTIONS ELEMENTAIRES SUR LA STRUCTURE DES PROGRAMMES.....</b>	<b>69</b>
CALCUL D'UNE EXPRESSION ARITHMETIQUE.....	69
NOTION DE RECURRENCE; BOUCLES, INDICES, REGISTRES D'INDEX. ....	71
NOTION DE SOUS-PROGRAMME.....	73
INVARIANCE, REENTRANCE ET RECURSIVITE. ....	76
<b><u>LA CARTE MERE.....</u></b>	<b><u>78</u></b>
<b>CARACTERISTIQUES.....</b>	<b>78</b>
<b>FACTEUR D'ENCOMBREMENT D'UNE CARTE MERE.....</b>	<b>79</b>
FORMAT XT.....	82
FORMAT AT – BABY AT.....	82
FORMAT ATX.....	82
FORMAT LPX.....	83
FORMAT NLX.....	83
<b>ALIMENTATION PC ET BOITERS ASSOCIES.....</b>	<b>83</b>
ALIMENTATION AT.....	84
ALIMENTATION ATX POUR PC.....	85
ALIMENTATION ATX POUR PENTIUM 4.....	85
AUTRES CARACTERISTIQUES DES ALIMENTATIONS.....	86
<b>COMPOSANTS INTEGRES.....</b>	<b>86</b>
<b>LE CHIPSET.....</b>	<b>86</b>
<b>ARCHITECTURE PENTIUM.....</b>	<b>87</b>
<b>ARCHITECTURE PENTIUM II 233 – 450 MHZ, PENTIUM III JUSQU'A 600 MHZ, PENTIUM PRO.....</b>	<b>89</b>
440 LX: PENTIUM II 233 -333 MHZ.....	89
ARCHITECTURE PENTIUM II DE 350 A 450– PREMIERS PENTIUM III 450 A 600 MHZ.....	89
<b>PENTIUM PRO.....</b>	<b>90</b>
<b>LE CAMINO INTEL I820.....</b>	<b>90</b>
<b>ALI ALADDIN TNT2.....</b>	<b>91</b>
<b>VIA APOLLO PRO 133 ET APOLLO VCM133. (PENIUM III).....</b>	<b>91</b>
<b>VIA APOLLO KX133 ET APOLLO KT133 (ATHLON ET DURON).....</b>	<b>91</b>
<b>VIA PROSAVAGE PM133. (CELERON / PENTIUM II / PENTIUM III).....</b>	<b>92</b>
<b>INTEL I840 (XEON).....</b>	<b>92</b>
<b>INTEL I820E.....</b>	<b>92</b>
<b>LE I850 (PENIUM IV).....</b>	<b>92</b>
<b>AMD-760 (ATHLON ET DURON).....</b>	<b>93</b>
<b>NVIDIA nFORCE POUR ATHLON ET DURON.....</b>	<b>93</b>
<b>INTEL I845.....</b>	<b>95</b>
<b>CHIPSET PROCESSEUR PENTIUM INTEL.....</b>	<b>97</b>
<b>CHIPSET PENTIUM II, CELERON ET PENTIUM III.....</b>	<b>99</b>
<b>CHIPSET PENTIUM PRO.....</b>	<b>101</b>
<b>CHIPSET INTEL XEON.....</b>	<b>101</b>
<b>CHIPSET INTEL PENTIUM IV.....</b>	<b>101</b>
<b>CHIPSET DURON ET ATHLON.....</b>	<b>103</b>
<b>CHIPSET ITANIUM.....</b>	<b>104</b>
<b>ATHLONS 64 BITS.....</b>	<b>104</b>
<b><u>L'HORLOGE ET LA PILE DU CMOS.....</u></b>	<b><u>104</u></b>

<b><u>LE BIOS .....</u></b>	<b><u>105</u></b>
<b><u>LE SUPPORT DE PROCESSEUR .....</u></b>	<b><u>105</u></b>
<b><u>TYPE DE SOCKET .....</u></b>	<b><u>106</u></b>
SOCKETS PRENANT EN CHARGE LES CPU D'INTEL.....	106
SOCKETS PRENANT EN CHARGE LES CPU D'AMD .....	107
<b><u>LES MICROPROCESSEURS POUR PC .....</u></b>	<b><u>107</u></b>
LE 8088 ET 8086. ....	108
LE 286.....	108
LE 386.....	109
LE 486.....	109
LES PENTIUM .....	110
LES PENTIUM MMX, K-6, K6-2, K6- CYRIX 6X86 .....	110
LES PENTIUM II – CELERON ET PENTIUM III.....	111
MICROPROCESSEUR INTEL PENTIUM III COPERMINE .....	112
L'ATHLON, DURON ET THUNDERBIRD D'AMD.....	113
EVOLUTION.....	113
MICROPROCESSEUR PENTIUM IV (FIN 2000).....	114
CARACTERISTIQUES. ....	114
EVOLUTION .....	115
NOTATIONS DES PENTIUM IV .....	116
PENTIUM IV EE.....	116
SEMPRON D'AMD .....	117
LES PROCESSEURS 64 BITS. ....	117
AMD 64 BITS.....	118
INTEL CORE, INTEL CODE 2 DUO.....	119
<b><u>LES MICROPROCESSEURS POUR PC PORTABLES.....</u></b>	<b><u>120</u></b>
LES PROCESSEURS CRUSOE. ....	120
PENTIUM III-M.....	120
PENTIUM IV MOBILE .....	121
PENTIUM M.....	121
INTEL CORE .....	121
A. INTEL CORE2 EXTREME.....	122
B. INTEL CORE 2 DUO .....	122
C. INTEL CORE 2 SOLO .....	122
ET AMD? .....	122
TURION.....	122
<b><u>MICROPROCESSEUR DUAL-CORE .....</u></b>	<b><u>123</u></b>
DIFFERENTES TECHNOLOGIES DU DUAL CORE.....	123
AVANTAGES. ....	124
PROCESSEURS DUAL CORE INTEL.....	124

DUAL CORE AMD .....	125
LES SYSTEMES D'EXPLOITATION. ....	125
<b><u>STRUCTURE INTERNE DES PROCESSEURS .....</u></b>	<b><u>126</u></b>
STRUCTURE INTERNE D'UN PROCESSEUR DE BASE .....	126
UTILISATION D'UNE MEMOIRE CACHE .....	128
PASSAGE DU CISC AU RISC .....	129
ARCHITECTURE PIPELINE, SUPERPIPELINE ET SUPERSCALAIRE .....	130
TAILLE DES TRANSISTORS, ALIMENTATIONS. ....	132
ARCHITECTURE INTERNE DES PROCESSEURS .....	132
CONCLUSION.....	134
<b><u>STRUCTURE INTERNE DES MICROPROCESSEURS INTEL ACTUELS.....</u></b>	<b><u>134</u></b>
L'ARCHITECTURE NETBURST.....	134
INTEL CORE ET CORE 2 DUO.....	136
<b><u>STRUCTURE DES MICROPROCESSEURS ATHLON 64 .....</u></b>	<b><u>136</u></b>
ARCHITECTURE STANDARD DES ATHLON 64, ATHLON 64FX ET OPTERON .....	137
ARCHITECTURE INTERNE D'UN ATHLON 64.....	140
<b><u>LES CONNECTEURS DE MEMOIRE VIVE.....</u></b>	<b><u>141</u></b>
TYPES DE MEMOIRES VIVES.....	142
FONCTIONNEMENT DE LA MEMOIRE VIVE.....	142
FORMATS DE BARRETTES DE MEMOIRE VIVE .....	143
DRAM PM.....	144
DRAM FPM.....	144
DRAM EDO.....	145
SDRAM .....	145
DR-SDRAM (RAMBUS DRAM).....	145
DDR-SDRAM .....	145
DDR2-SDRAM.....	146
TABLEAU RECAPITULATIF .....	146
SYNCHRONISATION (TIMINGS) .....	147
LA CORRECTION D'ERREURS.....	147
BIT DE PARITE.....	148
BARRETTES ECC.....	148
DUAL CHANNEL.....	148
RECAPITULATIF.....	148
MEMOIRE SPECIALE ORDINATEUR PORTABLE. ....	148
MEMOIRES AVEC CORRECTION D'ERREURS.....	149
<b><u>LES CONNECTEURS D'EXTENSION .....</u></b>	<b><u>149</u></b>
<b><u>LES CONNECTEURS D'ENTREE-SORTIE.....</u></b>	<b><u>150</u></b>

<b><u>LES BUS INTERNES DES PC .....</u></b>	<b><u>150</u></b>
LE BUS ISA.....	151
LE BUS MCA D'IBM.....	151
LE BUS VLB .....	151
LE BUS PCI .....	151
BUS PCI-X .....	152
LE BUS AGP .....	153
PCI-EXPRESS 1.1 ET 2.0.....	154
LE BUS PCMCIA .....	155
CALCULER LA BANDE PASSANTE D'UN BUS INFORMATIQUE. ....	157
<b><u>LE BUS PCI EXPRESS .....</u></b>	<b><u>157</u></b>
CARACTERISTIQUES DU BUS PCI EXPRESS.....	157
CONNECTEURS PCI EXPRESS .....	157
<b><u>LES CARTES GRAPHIQUES ACCELERATRICES 2D.....</u></b>	<b><u>158</u></b>
<b><u>FONCTIONNEMENT D'UNE CARTE GRAPHIQUE.....</u></b>	<b><u>160</u></b>
INTRODUCTION. ....	160
CARACTERISTIQUES D'UNE CARTE GRAPHIQUE (ECRAN).....	160
LE BUS UTILISE.....	161
LE CIRCUIT INTEGRE .....	161
LA MEMOIRE.....	161
LE CONVERTISSEUR DIGITAL – ANALOGIQUE.....	163
LES POSSIBILITES 3D .....	163
LE TAUX DE RAFRAICHISSEMENT. ....	163
LE BIOS VIDEO.....	164
LE CONNECTEUR DE SORTIE. ....	164
DOUBLE AFFICHAGE ET SORTIE TV .....	164
TECHNOLOGIES DIVERSES, SLI, SHADER... ..	164
<b><u>S-ATA - SERIAL ATA.....</u></b>	<b><u>166</u></b>
<b><u>CARACTERISTIQUES D'UN DISQUE DUR.....</u></b>	<b><u>167</u></b>
TYPES DE PARTITIONS WINDOWS.....	169
LA NORME IDE, E-IDE OU ULTRA-IDE. ....	169
MODES DE TRANSFERT.....	170
LES DISQUES UDMA-33 OU ATA-33 OU ATA 4.....	170
L'UDMA-66 OU ULTRA ATA 66.....	171
ATA / 100 ET ATA 133- ATA6 .....	171
RESUME DES MODES.....	171
LIMITATIONS DE CAPACITE.....	172
INSTALLER UN DISQUE DUR IDE (OU CD-ROM) .....	172
<b><u>DISQUE DUR SERIAL ATA (S-ATA).....</u></b>	<b><u>173</u></b>

LA COMMANDE DOS FDISK.....	174
LE BLOCK MODE.....	175
<u>QU'ENTEND-ON PAR FORMATAGE DE DISQUE ?.....</u>	<u>175</u>
SYSTEMES DE FICHIERS .....	176
<u>COMPREHENSION DES PARTITIONS.....</u>	<u>178</u>
COMPREHENSION DES LETTRES D'UNITE .....	180
<u>DISQUE DUR SCSI, TECHNOLOGIE RAID.....</u>	<u>183</u>
TECHNOLOGIE SCSI.....	183
CONNEXION DE PERIPHERIQUES SCSI.....	186
RAID (REDUNDANT ARRAY OF INDEPENDANT DISK) .....	187
RAID 0 (STRIPING).....	187
RAID 1 (MIRRORING) .....	188
RAID 2.....	189
RAID 3.....	189
LE RAID 4 .....	190
LE RAID 5 .....	191
ORTHOGONAL RAID 5.....	191
LES AUTRES RAID .....	191
RAID HARDWARE ET SOFTWARE. ....	192
SOLUTIONS RAID HARDWARE. ....	192
<u>LES IRQ ET LES DMA.....</u>	<u>192</u>
<u>LES INTERRUPTIONS .....</u>	<u>192</u>
LES DIFFERENTES INTERRUPTIONS.....	193
LES ADRESSES E/S (ENTREE/SORTIE).....	193
LES ATTRIBUTIONS HABITUELLES DES PORTS E/S.....	193
CANAUX DMA. (DIRECT MEMORY ACCESS).....	194
<u>LE PLUG AND PLAY .....</u>	<u>194</u>
1993 NAISSANCE DU PLUG AND PLAY. ....	194
<u>PRESENTATION DU DVD .....</u>	<u>196</u>
STRUCTURE PHYSIQUE .....	197
FORMATS STANDARD DE DVD .....	198
FORMATS STANDARD DE DVD ENREGISTRABLES .....	198
DVD-R/RW .....	198
DVD+R/RW .....	199
DIFFERENCES ENTRE DVD+ ET DVD- .....	199

DVD DL.....	200
STRUCTURE LOGIQUE.....	200
SYSTEME DE FICHIERS ET REPERTOIRES .....	201
STRUCTURE D'UN DVD VIDEO.....	201
ZONES.....	203
<b><u>LECTEUR DE BANDE.....</u></b>	<b>203</b>
INTRODUCTION. ....	203
LES TYPES.....	204
<b><u>LES DISQUETTES DE GRANDE CAPACITE.....</u></b>	<b>204</b>
INTRODUCTION .....	204
IOMEGA ZIP 100 MB, 250 MB ET 750 MB. ....	205
LS-120 MB DE SYQUEST.....	205
MEMOIRES SUR PORTS USB (STICK USB) .....	205
DISQUE DUR EXTERNE. ....	205
<b><u>ACQUISITION IMAGE: APPAREIL PHOTO NUMERIQUE ET SCANNER. ....</u></b>	<b>206</b>
APPAREIL PHOTO-NUMERIQUE .....	206
TYPES DE MEMOIRES FLASH .....	207
MEMORY STICK.....	207
SECURE DIGITAL (SD - SDHC), LA PLUS COURANTE. ....	208
COMPACT FLASH.....	208
XD PICTURE CARD .....	209
<b><u>LES SCANNERS.....</u></b>	<b>209</b>
LE PILOTE. ....	210
LA RESOLUTION ET LA PALETTE. ....	210
DETRAMAGE. ....	210
LE FONCTIONNEMENT D'UN SCANNER A PLAT. ....	211
L'OCR.....	212

## Chapitre 1 - Introduction

Ce chapitre dédié au lecteur débutant en informatique n'a d'autre but que d'introduire, sous une forme très intuitive et simplifiée, les notions d'informatique générale nécessaires à la bonne compréhension de cet ouvrage. Certaines d'entre elles qui font partie de notre propos seront reprises avec rigueur dans les chapitres suivants; d'autres sont simplement mentionnées pour permettre au lecteur de replacer les développements ultérieurs dans un contexte plus général. Pour tenter de donner une certaine vie à cette introduction, nous adoptons la voie historique, si tant est que l'on puisse parler d'histoire à propos d'une technique qui n'a guère que vingt ans et dont le nom même, l'informatique, est encore tout récent.

Entre la « préhistoire » et l' « histoire », le lecteur s'y trouve de but en blanc confronté avec la structure et le fonctionnement de Boulix. Boulix est l'ordinateur simplifié - est-il besoin de le spécifier que nous nous proposons de construire élément par élément - sur le papier cela s'entend en tant qu'illustration des chapitres consacrés à la logique de base des ordinateurs. Cette première présentation, très intuitive, permettra au lecteur de mieux comprendre les buts et les positions respectives des différents organes d'un ordinateur, lorsqu'ils lui seront décrits au long des pages qui suivront.

### Introduction historique : les précurseurs.

Les ordinateurs traitent de **l'information digitale**, c'est-à-dire de l'information fondamentalement discontinue comme des chiffres ou des caractères alphabétiques, par opposition aux calculateurs analogiques et simulateurs qui traitent de **l'information analogique** qui a un caractère essentiellement continu, comme des tensions ou des intensités électriques. La nature n'a pas attendu les fruits de l'intelligence humaine pour produire des machines digitales : la cellule vivante en est le meilleur exemple ; on a découvert ces dernières années qu'elle utilise un système de type digital pour conserver et transmettre la précieuse information héréditaire. Le boulier, qui semble dater du troisième millénaire avant Jésus Christ, apparaît comme la première machine typiquement digitale que l'homme ait utilisée pour l'aider à résoudre ses problèmes arithmétiques.

Les jaquemarts (Figure de métal ou de bois représentant un homme armé tenant un marteau qui frappe les heures sur la cloche ou le timbre de l'horloge d'une tour, d'un beffroi), qui fleurirent sur les clochers de nos églises, furent les ancêtres directs des automatismes à séquence, très répandus dans l'industrie. On appelle également ces derniers **calculateurs à programme intérieur** ou à programme figé, indiquant par là qu'ils sont condamnés à exécuter toujours la même suite d'opérations, c'est-à-dire le même **programme**.

Au milieu du dix-septième siècle, Pascal a fait un premier pas dans la mécanisation du calcul arithmétique en construisant une machine sur un concept nouveau : la roue dentée. Les roues de Pascal comportaient 10 positions (de 0 à 9) et chaque fois qu'une roue passait de la position 9 à la position 0, la roue immédiatement à gauche avançait automatiquement d'une position pour tenir compte de la retenue. Les machines électroniques de bureau et l'unité arithmétique de nos ordinateurs de gestion ne fonctionnent pas selon un autre principe, la roue dentée étant simplement remplacée par un circuit électronique. La machine de Pascal connaissait l'addition et la soustraction ; Leibniz lui a ajouté la multiplication et la division.

Le pas suivant dans la mécanisation du calcul a été réalisé par l'Anglais Babbage au dix-neuvième siècle. Il a pensé qu'il était inutile d'augmenter les performances des machines mécaniques tant que les opérations devaient être posées manuellement. Il fallait donc rendre automatique le passage d'une opération à la suivante. Mais, contrairement aux jaquemarts de nos églises, qui répètent inlassablement les mêmes gestes aux mêmes heures, les suites d'opérations à faire

exécuter n'étaient pas toujours les mêmes. Il a eu l'idée de reprendre un concept précédemment utilisé par Jacquard pour l'automatisation des métiers à tisser, celui de **programme extérieur**. La machine de Babbage devait en fait il n'a jamais pu la finir - lire une carte perforée, définissant une opération à exécuter ; exécuter l'opération; lire la carte suivante, etc. Sur le plan des principes, la machine de Babbage est l'ancêtre des premières machines électriques à relais construites dans certaines firmes et universités américaines au cours de la dernière guerre mondiale.

Nous désignons ce type de machines sous le nom de machines à programme extérieur, pour souligner que le programme n'est plus intérieur à la machine, donc figé, mais s'exécute pas à pas à partir d'un support extérieur, donc interchangeable.

**C'est Von Neumann** qui a fait faire vers 1945 le pas décisif à la mécanisation du traitement digital de l'information en introduisant deux concepts nouveaux :

(1) **Le programme enregistré**. Les machines à relais de l'époque possédaient des éléments de mémoire capables de conserver au cours du traitement des résultats partiels en vue de leur utilisation ultérieure. Von Neumann a eu l'idée d'utiliser les mémoires du calculateur pour emmagasiner également le programme : d'où le nom de **machine à programme enregistré** donné au nouveau type de calculateur. Au lieu d'exécuter les opérations au fur et à mesure de leur lecture sur le ruban perforé comme dans les machines à programme extérieur, la machine à programme enregistré suppose l'enregistrement du programme en mémoire avant de commencer l'exécution des opérations. Tous les ordinateurs fonctionnent sur ce principe.

(2) **La rupture de séquence**. La machine à programme extérieur supposait l'intervention de l'homme chaque fois qu'une décision devait être prise, c'est-à-dire que la suite du traitement dépendait des résultats déjà acquis.

Du fait des nouvelles possibilités offertes par la technique du programme enregistré, Von Neumann a eu l'idée de rendre automatiques les opérations de décision logique en munissant la machine d'une instruction appelée branchement conditionnel ou encore rupture de séquence conditionnelle. Suivant la valeur d'un résultat précédemment obtenu, positive ou négative par exemple, la machine devra exécuter telle ou telle partie du programme.

La plupart des ordinateurs fonctionnent selon le schéma originel de Von Neumann; nous les appellerons machines de type Von Neumann ou machines à registres.

### **Présentation générale d'un ordinateur.**

De multiples dénominations ont été utilisées pour désigner les machines électroniques à programme enregistré : calculateur, calculatrice, machine, ces termes étant suivis d'un adjectif tel que numérique, digital, ou tout simplement électronique. Le terme **ordinateur** est maintenant passé dans la langue française. Nous emploierons indifféremment l'une ou l'autre de ces dénominations.

Un ordinateur apparaît au premier abord comme composé d'une mémoire centrale qui contient programmes et données, d'une unité centrale de traitement qui exécute le programme, et d'unités d'entrées-sorties permettant des échanges avec l'extérieur. On désigne quelquefois l'ensemble de l'ordinateur et des éléments qui lui sont rattachés par le terme de **système informatique**.

### **L'unité centrale et la mémoire centrale.**

Le programme est enregistré en mémoire avant le début de son exécution. Cette mémoire, à partir de laquelle le programme peut être exécuté, est appelée **mémoire centrale** ou **mémoire principale**. Les différentes unités de la machine s'organisent autour de cette mémoire centrale.

La mémoire centrale contient deux types d'information : d'une part, les **instructions** du programme (ou informations traitantes) que la machine devra exécuter ; d'autre part, les données, souvent appelées **opérandes** (ou informations traitées), sur lesquelles la machine effectuera les traitements dictés par les instructions. A ces deux types d'information, traitante et traitée, correspondent deux unités particulières de la machine : l'unité **de contrôle**<sup>1</sup> également appelée **unité d'instructions** ou **unité de commande** pour les informations traitantes et l'**unité arithmétique et logique** ou unité **de traitement** pour les informations traitées.

(REMARQUE : en toute rigueur, le terme unité de commande conviendrait mieux que le terme unité de contrôle, eu égard à la définition du mot contrôle en français. Nous préférons cependant suivre la terminologie courante aux U. S. A. où l'on emploie « control unit ».)

On peut résumer schématiquement les fonctions principales de l'unité de contrôle<sup>1</sup> : elle extrait de la mémoire centrale la nouvelle instruction à exécuter; elle analyse cette instruction et établit les connexions électriques correspondantes dans l'unité arithmétique et logique ; elle extrait de la mémoire centrale les données sur lesquelles porte l'instruction; elle déclenche le traitement de ces données dans l'unité arithmétique et logique ; éventuellement, elle range le résultat dans la mémoire centrale.

L'unité arithmétique et logique effectue sur les données qu'elle reçoit les traitements commandés par l'unité de contrôle.

L'ensemble de l'unité de contrôle et de l'unité arithmétique et logique forme un tout dans la plupart des ordinateurs. On l'appelle **unité centrale** ou **unité centrale de traitement** ou encore **processeur central**. Pour certains auteurs, l'unité centrale comprend la mémoire centrale.

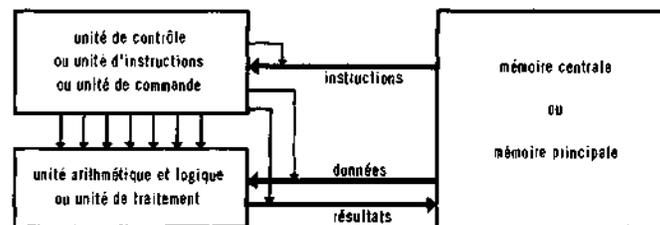


Fig.1 Organisation générale d'une unité centrale.

<sup>1</sup> REMARQUE : en toute rigueur, le terme unité de commande conviendrait mieux que le terme unité de contrôle, eu égard à la définition du mot contrôle en français. Nous préférons cependant suivre la terminologie courante aux U. S. A. où l'on emploie « control unit ».

## Les unités d'échange et les unités périphériques.

Telle qu'elle a été définie jusqu'à présent, la machine peut exécuter un programme initialement enregistré en mémoire centrale, portant sur des données également enregistrées en mémoire centrale, les résultats étant rangés en mémoire centrale au fur et à mesure de leur obtention. Il faut maintenant lui donner les moyens de communiquer avec l'extérieur : c'est le rôle des unités périphériques.

Il existe deux grandes classes **d'unités périphériques** :

Les **unités de communication** (lecteur de cartes, imprimante, unité de visualisation oscilloscopique, etc.) qui permettent le dialogue avec l'extérieur, et les **mémoires auxiliaires** (disques, bandes magnétiques, etc.) dont les capacités sont très nettement supérieures à celle nécessairement limitée de la mémoire centrale.

les unités périphériques sont reliées soit à l'unité centrale, soit directement à la mémoire par l'intermédiaire d'unités spécialisées dans la gestion des transferts d'informations appelées unités d'échange ou canaux. L'unité de contrôle commande les unités d'échange lorsqu'elle rencontre des instructions d'échange d'informations avec l'extérieur, appelées instructions d'entrée-sortie.

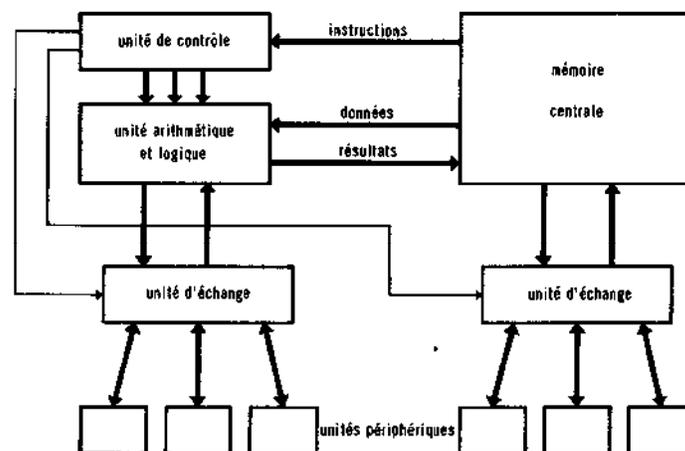


Fig. 2. Schéma général d'un ordinateur.

Ainsi un calculateur apparaît comme un assemblage d'unités distinctes dont le fonctionnement est dicté par le programme contenu en mémoire centrale. L'unité de contrôle commande l'exécution des opérations demandées par le programme. Celles-ci sont exécutées par l'unité arithmétique et logique s'il s'agit d'un traitement, par une unité d'échange s'il s'agit d'un transfert d'informations avec l'extérieur.

## Principe du fonctionnement d'un ordinateur présentation de Boulix.

Nous allons faire un pas de plus dans la description du fonctionnement d'un ordinateur de type Von Neumann, en cherchant à obtenir une compréhension intuitive des mécanismes plutôt qu'à fournir des définitions précises qui seront le fait des chapitres suivants. Ce sera l'occasion de procéder à une toute première présentation du calculateur, baptisé Boulix, qui nous servira d'exemple par la suite.

## Les registres.

Nous admettrons qu'au cours du traitement les différentes informations, instructions et données, peuvent être temporairement conservées dans des éléments de mémoire capables de contenir une information et appelés registres. Sur un ordre émanant généralement de l'unité de contrôle, une information peut être transférée d'un registre à un autre registre, ce transfert ne modifiant pas le contenu du premier registre.

## La mémoire centrale.

Elle peut être considérée comme formée d'un ensemble de **cellules**, chaque cellule pouvant contenir une information (nous dirons souvent un **mot**) donnée ou instruction. Les cellules sont numérotées et l'unité de contrôle connaît chaque cellule par son numéro, appelé **adresse**. Elle peut demander à **lire** le contenu d'une cellule d'adresse donnée ou à **écrire** une nouvelle information dans une cellule d'adresse donnée. Pour réaliser ces opérations, l'unité de contrôle fournit l'adresse de la cellule concernée dans un registre associé à la mémoire centrale et appelé **registre d'adresse** ou encore **registre de sélection mémoire**.

Le dispositif de sélection mémoire analyse l'adresse contenue dans le registre de sélection et sensibilise la cellule adressée soit pour une lecture, soit pour une écriture. Dans le cas d'une lecture, l'information contenue dans la cellule adressée est transférée dans un deuxième registre, associé à la mémoire et appelé **registre d'échange** ou **registre mot**. Dans le cas d'une écriture, il faut préalablement charger ce même registre de l'information à transférer dans la cellule adressée. L'opération de lecture ne détruit pas l'information contenue dans la cellule adressée. L'opération d'écriture la détruit en la remplaçant par une nouvelle information.

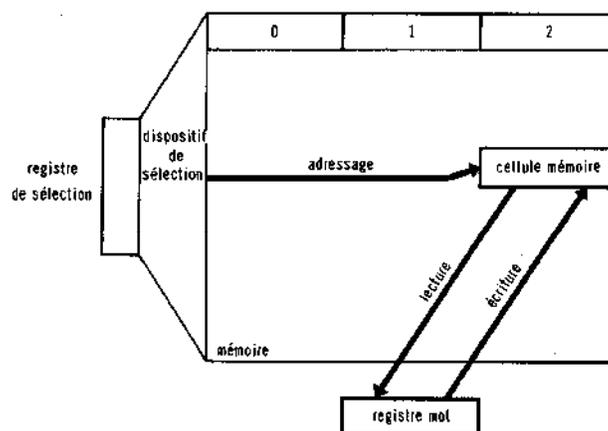


Fig. 3. Lecture et écriture en mémoire.

## Le programme.

Il se compose des instructions supposées rangées séquentiellement dans la mémoire. Ceci implique que normalement les instructions qui doivent s'exécuter à la suite les unes des autres soient rangées à des adresses successives de la mémoire : à la fin d'une instruction d'adresse A, le calculateur enchaîne automatiquement sur l'instruction d'adresse A+ 1, sauf dans le cas d'une rupture de séquence.

Schématiquement, nous distinguerons trois grands types d'instructions :

- (1) les instructions de traitement portant sur des opérandes en mémoire, comprenant essentiellement les opérations arithmétiques et logiques et l'opération de rangement en mémoire ;
- (2) les instructions de rupture de séquence permettant de rompre l'enchaînement séquentiel des instructions et de passer à une autre partie du programme si certaines conditions sont réalisées ;
- (3) les instructions d'échange permettant les échanges d'informations entre le calculateur et le milieu extérieur.

## L'unité arithmétique et logique.

Pour commander au calculateur une opération arithmétique, une addition par exemple, l'instruction doit lui fournir les informations suivantes :

- (1) le type d'opération à réaliser, ici une addition ; c'est le rôle du code opération ;
- (2) l'adresse de la cellule mémoire qui contient la première donnée, ou premier opérande ;
- (3) l'adresse de la cellule mémoire qui contient le deuxième opérande ;
- (4) l'adresse de la cellule mémoire où doit être rangé le résultat. On en déduit la forme de l'instruction qui contient un code opération et trois adresses :

code	adresse	adresse	adresse
opération	1 <sup>er</sup> opérande	2 <sup>ème</sup> opérande	résultat

La figure 4a représente l'unité arithmétique et logique capable d'exécuter cette opération et qui est entourée de trois registres mémorisant les deux opérandes et le résultat. L'instruction d'addition nécessite quatre accès à la mémoire centrale, permettant respectivement de rechercher l'instruction, puis le premier opérande, puis le second opérande, et enfin de ranger le résultat. Les machines qui utilisent ce type d'instruction sont appelées machines à trois adresses.

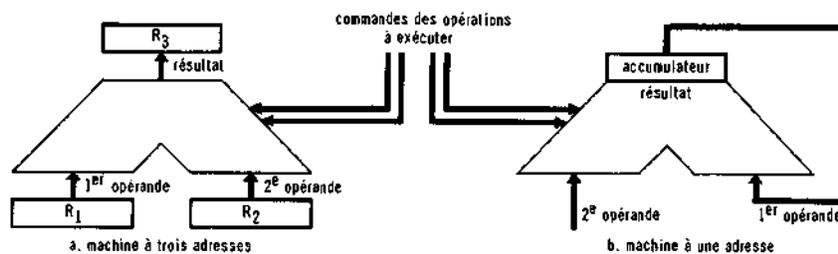


Fig. 4. Unité arithmétique et logique.

Boulx est une machine dite à une adresse. Son unité arithmétique et logique contient un registre particulier appelé accumulateur qui sert à contenir à la fois le premier opérande et le résultat, ce qui permet à l'instruction de ne contenir qu'une adresse : celle du deuxième opérande.

L'opération d'addition nécessite trois instructions :

- (1) chargement du premier opérande dans l'accumulateur;
- (2) addition du deuxième opérande au contenu de l'accumulateur ;
- (3) rangement du contenu de l'accumulateur en mémoire. Chacune de ces instructions comportera un code instruction et une adresse :

	Code opération	Adresse
--	----------------	---------

(1)	Chargement	Adresse 1 <sup>er</sup> opérande
(2)	Addition	Adresse 2 <sup>nd</sup> opérande
(3)	Rangement	Adresse résultat

L'unité arithmétique et logique est schématisée fig. 4b, les registres  $R_1$  et  $R_3$  de la figure 4a sont remplacés par l'accumulateur. Le deuxième opérande peut être mémorisé pendant l'opération par le registre mot associé à la mémoire. C'est le cas de Boulx.

#### L'unité de contrôle.

C'est l'unité qui extrait les instructions de la mémoire et qui les analyse.

Pour cela, elle doit comporter deux registres :

- (1) un registre appelé **compteur d'instructions** ou **compteur ordinal** qui contient l'adresse de la nouvelle instruction à exécuter. Son nom provient de ce que, en dehors des ruptures de séquences, ce registre doit voir son contenu augmenté de un pour passer à l'instruction suivante ;
- (2) un registre, appelé **registre instruction**, qui contient l'instruction extraite de la mémoire. Le registre instruction de Boulx comprend deux parties : une partie contenant le code opération qui définit l'instruction à exécuter (addition, multiplication, branchement...) et une partie qui contient l'adresse de l'opérande.

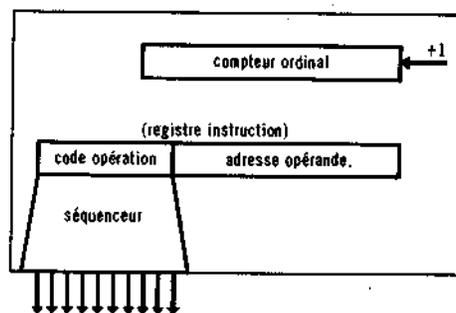


Fig. 5. Unité de contrôle.

De plus, l'unité de contrôle comporte un organe appelé **séquenceur** qui, après analyse du code opération, fournit les commandes à l'ensemble des unités de la machine (mémoire, unité arithmétique et logique) pour leur faire exécuter les différentes phases de l'instruction.

### Déroulement d'une instruction portant sur un opérande en mémoire.

Le déroulement d'une instruction de traitement dans un calculateur à une adresse tel que Boulix peut se décomposer en trois phases :

- (1) recherche et analyse de l'instruction;
- (2) recherche et traitement de l'opérande ou bien rangement de l'opérande ;
- (3) préparation de l'instruction suivante.

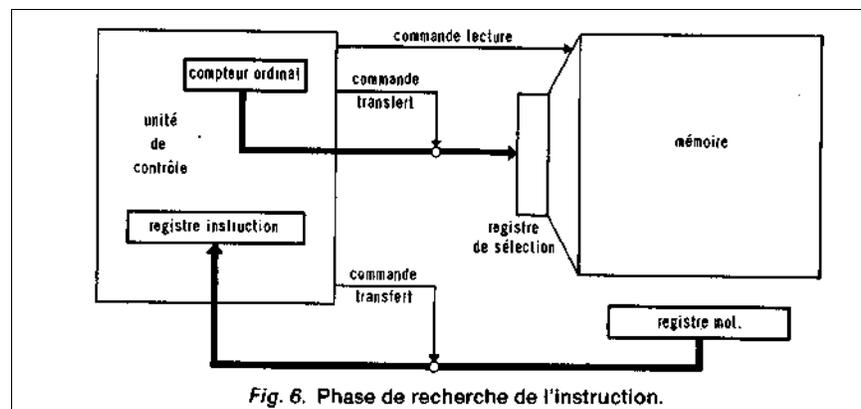
### Phase de recherche de l'instruction.

L'unité de contrôle commande le transfert du contenu du compteur ordinal (c'est-à-dire l'adresse de l'instruction à exécuter) dans le registre de sélection mémoire et envoie à la mémoire une commande de lecture.

Une fois cette opération de lecture terminée, l'instruction est disponible dans le registre mot.

L'unité de contrôle commande alors le transfert du contenu de ce registre dans le registre instruction.

La partie code opération de l'instruction peut être analysée par les circuits de l'unité de contrôle. Cette première phase est commune à tous les types d'instruction.



### Phase de recherche ou de rangement de l'opérande.

Cas de la recherche de l'opérande avec traitement.

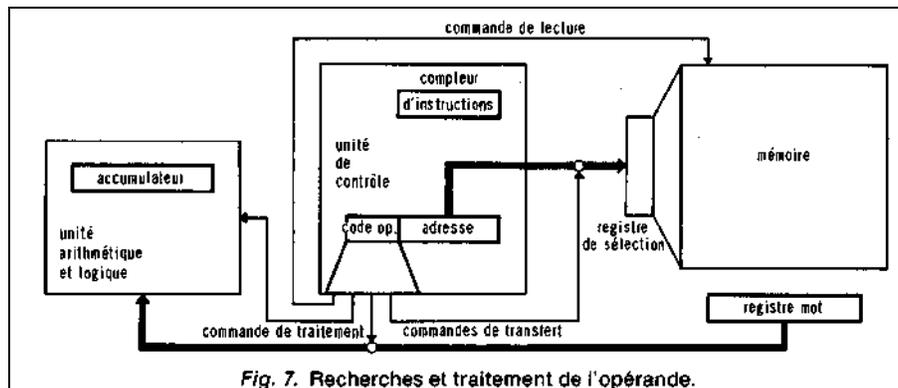
Une fois le code opération de l'instruction analysé, l'unité de contrôle sait qu'il s'agit d'une instruction de traitement avec recherche d'opérande préalable.

L'adresse de l'opérande se trouve dans la zone adresse de l'instruction.

L'unité de contrôle commande son transfert dans le registre de sélection de la mémoire puis commande à cette dernière une opération de lecture.

Une fois cette opération terminée, l'opérande cherché est disponible dans le registre mot.

L'unité de contrôle positionne les circuits de l'unité arithmétique et logique pour exécuter le traitement demandé par le code opération et commande le transfert de l'opérande vers l'unité arithmétique et logique. Le résultat du traitement de l'opérande est rangé dans l'accumulateur (cf. fig. 7). Notons que le traitement peut se réduire à un simple transfert de l'opérande dans l'accumulateur.

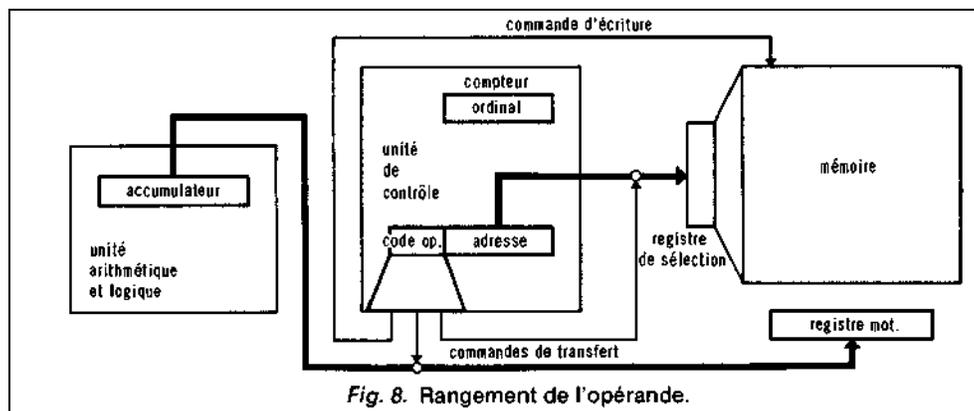


Cas du rangement de l'opérande.

L'adresse de rangement de l'opérande se trouve dans la partie adresse du registre instruction :

l'unité de contrôle commande son transfert dans le registre de sélection de la mémoire.

L'opérande à ranger est dans l'accumulateur : l'unité de contrôle commande son transfert dans le registre mot. Il ne lui reste plus qu'à commander à la mémoire une opération d'écriture. (Cf. fig. 8.)



### Phase de préparation de l'instruction suivante.

Elle consiste à augmenter de 1 le contenu du compteur ordinal, de façon qu'il contienne l'adresse de l'instruction suivante. Cette opération est appelée incrémentation du compteur ordinal.

### Instruction de rupture de séquence.

Ce type d'instruction, également appelé instruction de branchement ou de saut, permet de modifier le déroulement séquentiel du programme, en faisant suivre l'instruction de rupture de séquence non par l'instruction placée à l'adresse suivante, mais par une instruction dont l'adresse est fournie par l'instruction de rupture de séquence. Le branchement peut être conditionnel ; il ne sera alors effectif que si une condition, portant généralement sur le contenu de l'accumulateur, est réalisée; sinon le programme se continuera en séquence. Le code opération définit la condition, l'adresse, l'emplacement de l'instruction à exécuter si la condition est réalisée (fig. 9). Si la réponse de l'unité de traitement est que la condition est réalisée, l'unité de contrôle commande le transfert de l'adresse vers le compteur ordinal, et inhibe l'addition de 1 au compteur ordinal ; dans le cas contraire, elle commande uniquement l'incrémentation de 1 du compteur ordinal.

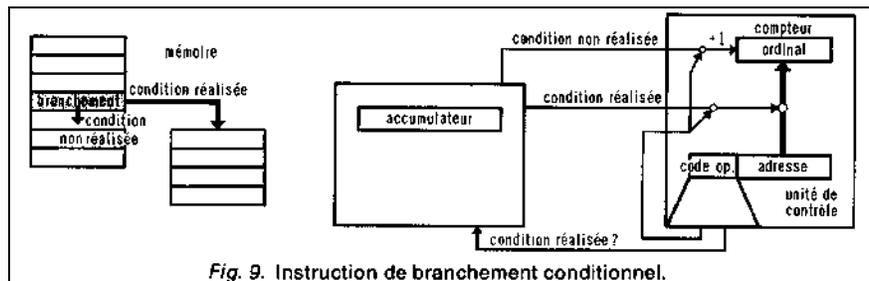


Fig. 9. Instruction de branchement conditionnel.

### L'unité d'échange.

Elle est généralement capable de gérer le transfert, aussi bien en entrée qu'en sortie, d'un ensemble d'informations entre les unités périphériques et la mémoire centrale ; dans les ordinateurs d'aujourd'hui, ce transfert peut être exécuté simultanément avec le déroulement d'un programme de calcul. Les informations transférées sont rangées en mémoire séquentiellement. Pour initialiser un tel transfert, des instructions spéciales d'entrée - sortie doivent fournir à l'unité d'échange, d'une part, l'adresse de l'unité périphérique concernée et, d'autre part, l'adresse de rangement de la première information et le nombre d'informations à transférer. Par la suite, l'unité d'échange se chargera intégralement de la gestion du transfert : à chaque information transférée, elle ajoutera 1 à l'adresse de rangement et retranchera 1 au nombre d'informations à transférer. Elle avertira l'unité de contrôle lorsque toutes les informations auront été transférées. (Fig. 10.)

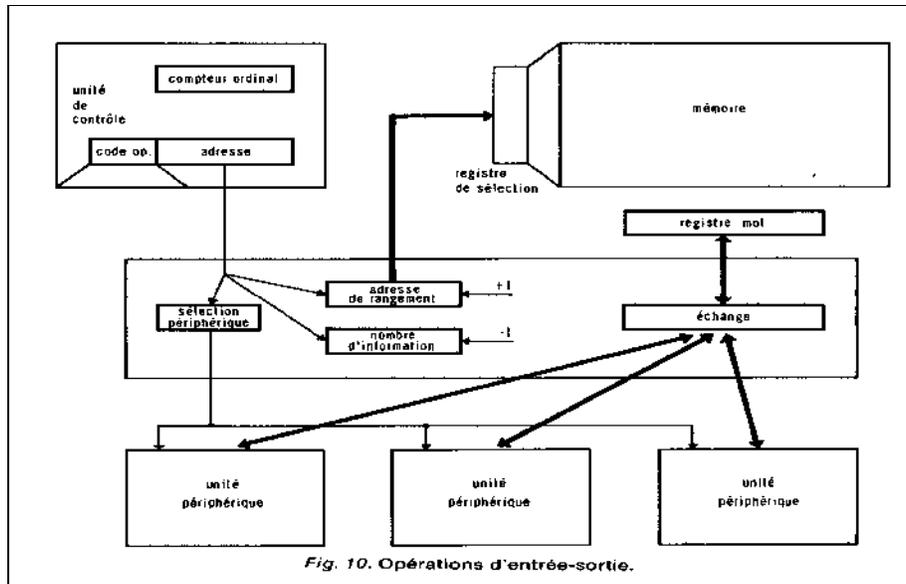


Fig. 10. Opérations d'entrée-sortie.

## Les unités périphériques.

On distingue

(1) les mémoires auxiliaires qui servent de moyen d'enregistrement de grande capacité et de moyen de communication à l'intérieur du système. Ce sont les tambours, les disques, les bandes magnétiques, les mémoires à feuillets magnétiques ;

(2) les unités permettant les communications entre le système et le milieu extérieur; notamment, les unités permettant les échanges d'informations sur supports générés ou lus par l'homme, telles que lecteurs et perforateurs de rubans ou cartes perforées, imprimantes., etc. ; les unités permettant à l'homme de dialoguer avec le calculateur telles que machines à écrire, unités de visualisation, etc. ; les unités permettant des échanges d'informations à distance avec des matériels fournissant ou acceptant des données digitales telles que lignes de transmission, terminaux, autres calculateurs, etc. ; les unités permettant des échanges d'informations avec les matériels fournissant ou acceptant des données analogiques (convertisseurs analogiques-numériques ou numériques-analogiques).

La plupart de ces unités se divisent en deux parties:

(1) une partie électronique appelée **unité de commande**, unité de contrôle de périphérique ou encore unité de liaison, et

(2), commandée par cette dernière, une unité électromécanique réalisant la lecture ou l'écriture des informations.

## Les interruptions.

Les **interruptions** sont des commandes qui émanent du milieu extérieur et qui demandent au calculateur d'exécuter un programme associé à la commande. Le programme en cours est alors interrompu pour permettre au programme demandé par l'interruption et considéré comme prioritaire de s'exécuter. Une fois ce dernier terminé, l'exécution du programme interrompu est

reprise. C'est par exemple grâce aux interruptions que les unités d'échange préviennent l'unité de contrôle que les opérations d'entrée-sortie sont terminées.

### Configuration d'un système informatique.

On appelle configuration d'un système de traitement de l'information la liste et éventuellement les caractéristiques des unités qui le composent ainsi que la manière dont elles sont interconnectées.

### Quelques ordres de grandeur.

Comme dans toute technique, il est nécessaire d'utiliser comme références les ordres de grandeur les plus habituels. C'est d'autant plus vrai dans le domaine des ordinateurs que les ordres de grandeur des temps d'exécution sont difficilement appréhendables dans notre échelle humaine des temps.

On exprime les durées d'opérations en sous-multiples de la seconde qui sont

la milliseconde (ms) =  $10^{-3}$  seconde

la microseconde ( $\mu$ s) =  $10^{-6}$  seconde

la nanoseconde (ns) =  $10^{-9}$  seconde

la picoseconde (ps) =  $10^{-12}$  seconde.

Les temps de propagation des signaux électriques et les temps de commutation (un signal passe ou ne passe pas) s'expriment en nanosecondes, exceptionnellement en centaines de picosecondes. Une impulsion électrique ne peut se propager plus rapidement que la vitesse de la lumière dans le vide, ce qui correspond à 30 cm par nanoseconde environ.

Voici quelques ordres de grandeur dans le domaine des performances : opérations arithmétiques : quelques dizaines de nanosecondes à quelques microsecondes ; opération de lecture ou d'écriture en mémoire centrale : quelques centaines de nanosecondes à quelques microsecondes ; nombre d'instructions exécutées par seconde : quelques dizaines de milliers à quelques millions ; un lecteur de cartes lit entre 50 et 1 500 cartes à la minute, une imprimante imprime entre 80 et 2 200 lignes à la minute ; une bande magnétique peut être lue ou écrite à la vitesse de 10 000 à 320 000 caractères à la seconde ; les transferts entre mémoire centrale et disque magnétique peuvent atteindre plusieurs millions de caractères par seconde.

On exprime les capacités de mémoire par le nombre total de positions mémoire ou de **caractères** (lettres, chiffres ou signes de ponctuation) qu'elles peuvent contenir.

La mémoire centrale varie généralement entre quelques milliers et quelques millions de caractères ; l'ensemble des unités de mémoire de masse (disques et feuillets magnétiques) d'un gros système peut atteindre plusieurs milliards de caractères. A titre de comparaison, ce livre comporte un peu moins d'un million de caractères.

1 K (Kilo)	$2^{10}$	= 1024
1 M (Méga)	$2^{20}$	= 1 048 576
1 G (Giga)	$2^{30}$	= 1 073 741 824
1 T (Téra)	$2^{40}$	= 1 099 511 627 776

SI Prefixes

Préfixes SI

In English Name	International Symbol	Value	En français Nom
Yotta	Y	$10^{24}$	Yotta
Zetta	Z	$10^{21}$	Zetta
Exa	E	$10^{18}$	Exa
Peta	P	$10^{15}$	Peta
Tera	T	$10^{12}$	Téra
Giga	G	$10^9$	Giga
Mega	M	$10^6$	Méga
Kilo	k	$10^3$	Kilo
Hecto	h	$10^2$	Hecto
Deka	da	$10^1$	Deca
Deci	d	$10^{-1}$	Déci
Centi	c	$10^{-2}$	Centi
Milli	m	$10^{-3}$	Milli
Micro	$\mu$	$10^{-6}$	Micro
Nano	n	$10^{-9}$	Nano
Pico	p	$10^{-12}$	Pico
femto	f	$10^{-15}$	Femto
Atto	a	$10^{-18}$	Atto
Zepto	z	$10^{-21}$	Zepto
Yocto	y	$10^{-24}$	Yocto

### Notions de hardware et de software.

Nous avons maintenant une compréhension suffisante de ce qu'est un ordinateur pour pouvoir aborder certains points de terminologie ayant un caractère très général.

Nous remarquerons d'abord qu'un ordinateur sans programme est parfaitement inutilisable. Ceci conduit à distinguer deux notions : le hardware et le software.

Le **hardware**, mot anglais signifiant quincaillerie, désigne tout ce qui a un caractère matériel dans une machine, tout ce qui est figé; c'est en gros l'ensemble des éléments décrits au paragraphe précédent.

Le **software**, mot fabriqué pour les besoins de la cause en remplaçant hard (dur) par soft (mou), désigne au contraire tout ce qui n'est pas matériel, tout ce qui peut être facilement modifié, tout ce qui par analogie peut être assimilé à la matière grise ; c'est en première approximation le domaine de la programmation.

On admet généralement que le hardware recouvre deux disciplines :

la technologie et la logique,

la **technologie** ayant pour objet l'étude des composants :

transistors, résistances, éléments de mémoire, etc., et

la **logique** l'étude de l'assemblage de ces composants pour en faire des circuits tels que des registres, des additionneurs, des circuits de sélection mémoire ou des unités telles que l'unité de contrôle, l'unité arithmétique et logique, etc.

A ceci vient s'adjoindre une troisième discipline dont l'importance va croissant avec l'augmentation de complexité de l'organisation des systèmes informatiques et à laquelle nous donnerions volontiers le nom d'**architecture** : il s'agit de l'assemblage des unités que la logique a permis de concevoir.

On désigne plus couramment par **structure des ordinateurs** la réunion de deux concepts de logique et d'architecture.

Outre la programmation d'application dans les domaines aussi variés que le calcul scientifique, la gestion, la reconnaissance de formes, les problèmes d'intelligence artificielle, etc., le software se divise en deux branches :

la première assez théorique recouvre les problèmes liés aux **langages de programmation**,

la seconde, peu formalisée à ce jour, correspond aux **systèmes d'exploitation** qui regroupent les programmes chargés de la gestion du fonctionnement d'un système informatique.

L'ensemble des instructions telles que nous les avons abordées dans la description succincte de Boulix forme le **langage machine** de l'ordinateur.

Deux caractéristiques :

Il est spécifique de chaque ordinateur.

Il possède une forme interne à la machine et une forme externe lisible par le programmeur appelée **langage assembleur**.

Très vite s'est fait sentir le besoin de langages plus généraux et plus proches de l'écriture mathématique. D'où le développement de **langages évolués** tels que

Fortran pour le calcul scientifique,

Algol pour le calcul scientifique et la description formelle des algorithmes,

Cobol pour la gestion,

PL1 qui tente de regrouper ces trois langages,

ainsi que de multiples langages orientés vers des types d'applications plus particuliers.

L'utilisation de ces langages suppose des programmes capables de traduire le langage de sa forme évoluée connue du programmeur à sa forme interne connue de la machine.

Ces programmes de traduction sont appelés **compilateurs**, sauf si l'on part du langage assembleur auquel cas on utilise souvent le terme **d'assembleur**.

Il n'est pas possible à ce niveau de passer en revue toutes les fonctions d'un **système d'exploitation** :

Indiquons

- (1) qu'il gère les travaux des utilisateurs, c'est-à-dire qu'il enchaîne les différentes étapes du passage d'un programme (chargement en mémoire, assemblage ou compilation, exécution) et enchaîne les travaux entre eux'
- (2) qu'il gère les ressources de la machine en affectant les zones de la mémoire centrale et les unités périphériques aux différents travaux, etc.

La conception d'une machine suppose la mise en oeuvre de ces différentes disciplines, la technologie, la logique et l'architecture, d'une part, les langages et les systèmes d'exploitation d'autre part. Or, pendant longtemps, la conception de la machine a paru être l'oeuvre des hommes du hardware, les hommes du software devant par la suite se débrouiller avec le hardware qui leur était fourni. Aujourd'hui la conception de l'architecture d'un ordinateur suppose une collaboration très étroite entre les concepteurs de hardware et les concepteurs de software. C'est dans ce sens qu'ont été conçus les derniers chapitres de cet ouvrage.

### Historique des machines les générations d'ordinateurs.

Il n'a fallu que quelques années pour transposer la machine de Von Neumann en calculateur électronique commercialisable.

Dans les années 1952 à 1955, les diverses firmes présentèrent leurs premières machines. On se plaît à distinguer trois générations d'ordinateurs depuis leur commercialisation, et l'on entrera bientôt dans l'ère de la quatrième. Cette distinction est communément fondée sur des critères d'ordre technologique; mais les mutations d'une génération à l'autre, concernant l'architecture, sont pour nous encore plus intéressantes et significatives.

### L'évolution de la technologie.

Aux quatre, générations, on peut en première approximation faire correspondre successivement les tubes électroniques pour la première, les transistors pour la deuxième, les circuits intégrés pour la troisième, et très probablement les circuits intégrés à moyenne ou à grande échelle pour la quatrième.

Un transistor se réduit à un petit parallélépipède de silicium à base carrée, de quelques dixièmes de millimètre de côté et de 150 microns d'épaisseur. Dans la deuxième génération, les transistors, montés chacun dans un boîtier, étaient assemblés aux autres composants (diodes, résistances, capacités) sur des plaquettes de quelques centimètres ou quelques dizaines de centimètres de côté, les interconnexions entre ces composants étant réalisées par impression métallique:

Dans un **circuit intégré**, dont la taille est du même ordre de grandeur que le transistor, on trouve jusqu'à une dizaine de composants élémentaires interconnectés entre eux.

On emploie souvent le terme d'intégration à petite échelle (small scale integration, S. S. I.).

Les circuits intégrés à moyenne et grande échelle (middle scale and large scale integration : M. S. I. et L. S. I.) peuvent contenir respectivement quelques centaines ou quelques milliers de composants interconnectés.

Leur taille peut atteindre deux à trois centimètres de côté pour une épaisseur de 150 microns.

D'une génération à l'autre, on a pu noter un progrès quelquefois considérable sur les propriétés des circuits: **miniaturisation, fiabilité, complexité et vitesse.**

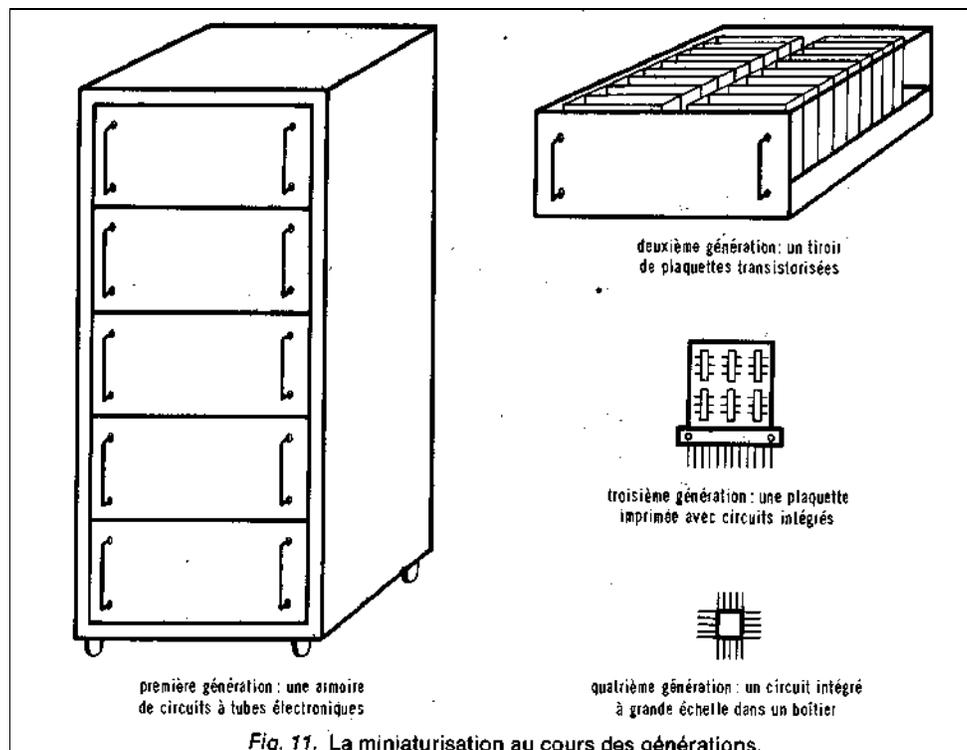
La **miniaturisation** est illustrée par la figure 11, où la même fonction logique nécessite une armoire dans la génération des tubes, un tiroir ou une partie de tiroir dans la génération des

transistors, une plaquette imprimée dans la génération des circuits intégrés à petite échelle, un boîtier de circuit intégré dans la génération des circuits intégrés à grande échelle.

La **fiabilité** introduit la notion de qualité de fonctionnement d'un composant ou d'un ensemble. Elle se mesure par le M. T. B. F. (Mean Time Between Failure) qui exprime la moyenne des temps entre pannes ou encore, pour conserver le sigle, la Moyenne des Temps de Bon Fonctionnement. Cette moyenne est passée de quelques dizaines de minutes pour une unité centrale moyenne de la première génération à quelques milliers d'heures pour une unité centrale équivalente de la troisième génération.

Ce progrès dans le domaine de la fiabilité a deux causes : le progrès dans la fiabilité du composant et la diminution du nombre d'interconnexions due à l'intégration. Le passage du tube au transistor a représenté un gain d'un facteur 10, tandis qu'un circuit intégré qui contient une dizaine de composants a la même fiabilité qu'un transistor, ce qui représente encore un gain d'un facteur 10.

Il ne faut pas confondre la notion de fiabilité qui représente une probabilité de fonctionnement sans aucune panne et celle de disponibilité qui correspond à la permanence du service rendu par un système : on peut concevoir des architectures de machines pour lesquelles des pannes partielles n'excluent pas un fonctionnement global correct ni même éventuellement le dépannage simultané des éléments défailants.



**La complexité.** La possibilité de concevoir des ensembles électroniques de plus en plus complexes est un corollaire direct du gain de fiabilité. Il serait parfaitement illusoire de concevoir une machine dont la moyenne des temps entre pannes serait de quelques secondes. C'est ainsi qu'à fiabilité équivalente on peut réaliser aujourd'hui des ensembles électroniques 1 000, à 10 000 fois plus complexes qu'en technologie à tubes.

**La vitesse.** Les temps de commutation des circuits logiques sont passés de quelques microsecondes pour la première génération à quelques nanosecondes pour la troisième génération. Ceci a permis de passer, à complexité sensiblement égale, de machines à un millier d'instructions par seconde à des machines à un million d'instructions par seconde.

### L'évolution de l'exploitation des ordinateurs.

En fait, si les trois générations recouvrent des technologies différentes tubes, transistors, circuits intégrés, elles recouvrent surtout des organisations et des techniques d'exploitation des ordinateurs différentes.

Le **calculateur de la première génération** exécutait ses travaux de manière purement séquentielle, chaque travail étant effectué en trois temps :

- (1) le programme perforé sur cartes ou sur ruban de papier était lu et enregistré en mémoire grâce à un programme appelé chargeur;
- (2) le programme était exécuté ;
- (3) les résultats étaient imprimés. (Cf. fig. 12.)

Des lectures de nouvelles données ou des impressions de résultats partiels pouvaient être incluses dans le programme, mais les opérations de traitement, d'entrée ou de sortie ne pouvaient que s'enchaîner dans le temps, leur durée s'ajoutant.

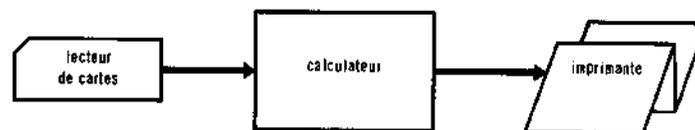


Fig. 12. Traitement séquentiel des travaux.

Le calculateur **de la deuxième génération** offrait des possibilités de simultanéité entre calcul et opérations d'entrée-sortie.

Cependant l'enchaînement des travaux restait séquentiel comme dans les machines de la première génération de telle sorte que les possibilités de simultanéité ne pouvaient s'exercer qu'à l'intérieur d'un même programme et, de ce fait, étaient généralement peu utilisées.

En particulier l'unité centrale restait en attente pendant les chargements des nouveaux programmes en mémoire.

La disproportion entre vitesse de calcul et vitesse de lecture des cartes ou d'impression est devenue telle que l'unité centrale n'était effectivement utilisée qu'un faible pourcentage du temps.

On a donc restreint les entrées et sorties de l'ordinateur aux seules bandes magnétiques, beaucoup plus rapides que les lecteurs de cartes et imprimantes. Les grosses installations possédaient alors une machine auxiliaire qui exécutait les conversions de supports carte à bande magnétique et bande magnétique à imprimante, l'ordinateur principal ne connaissant que les bandes magnétiques. (Cf. fig. 13.) Cette méthode d'exploitation est quelquefois appelée traitement par lot, pour indiquer qu'il fallait attendre que le lot de travaux chargés sur la bande magnétique soit entièrement traité pour obtenir les résultats de l'un d'entre eux, ou en charger un nouveau.

C'est au cours de la deuxième génération de calculateurs qu'est né un nouveau type d'application : le contrôle de processus qui nécessite que le calculateur soit directement connecté au processus contrôlé et travaille en synchronisation avec lui. Cette synchronisation est obtenue grâce aux interruptions de programme qui permettent au processus de prévenir le calculateur de tout événement, et de commander la prise en compte en priorité des programmes traitant ces événements.

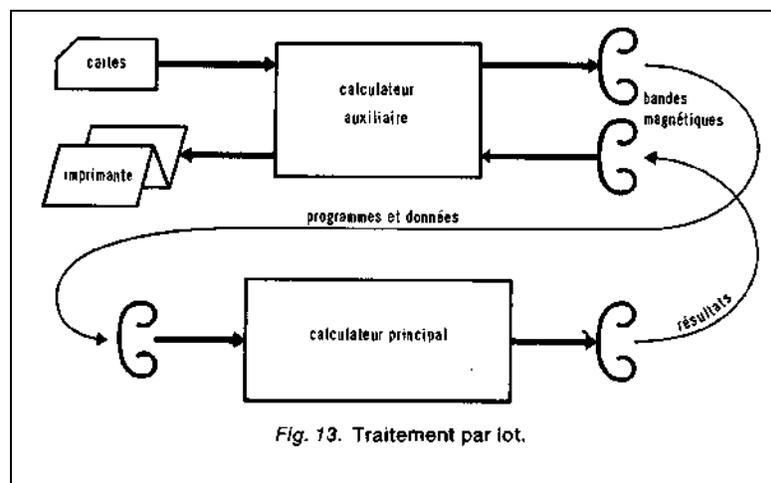
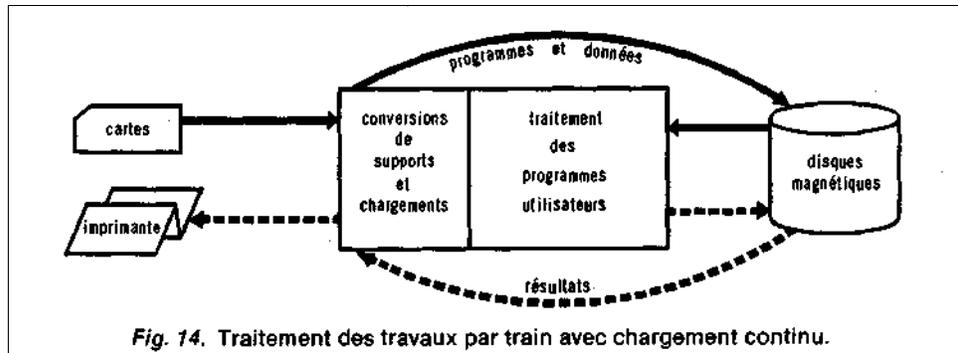


Fig. 13. Traitement par lot.

Le **calculateur de la troisième génération** permet d'exploiter efficacement les simultanités qui étaient plus potentielles que réelles dans le calculateur de la deuxième génération. Plusieurs programmes peuvent résider simultanément en mémoire; à un instant donné un seul d'entre eux utilise l'unité centrale, les autres pouvant simultanément effectuer des opérations d'entrée-sortie. Lorsque le programme qui utilise l'unité centrale s'arrête en attente d'une opération d'entrée-sortie, un autre programme prend sa place, ce qui évite les temps d'attente de l'unité centrale. Cette méthode d'exploitation est appelée **multiprogrammation**. Elle permet une meilleure utilisation de l'ensemble des ressources d'un système informatique.

L'exploitation normale d'un calculateur de la troisième génération consiste à partitionner, c'est-à-dire à diviser, la mémoire en deux zones, l'une réservée au train de travaux des utilisateurs, l'autre contenant les programmes de conversion de support et le système d'exploitation. (Cf. fig. 14.)



En première approximation, ces deux partitions correspondent respectivement au calculateur auxiliaire et au calculateur principal de la génération précédente.

Il y a cependant une différence importante : le chargement par lot est remplacé par le chargement continu des travaux au fur et à mesure de leur arrivée.

Les travaux sont mis en file d'attente sur le disque magnétique, puis chargés en mémoire pour être exécutés sous contrôle du système d'exploitation qui tient compte de leur priorité. Les résultats sont rangés sur disque puis sortis sur imprimante en tenant également compte de la priorité du travail correspondant. L'utilisateur prioritaire n'a donc pas à attendre le traitement de tout un lot de programmes pour obtenir ses résultats.

Le passage du traitement par lot au chargement continu représente un premier pas dans la recherche du « confort » de l'accès à l'ordinateur. La troisième génération a vu apparaître deux nouvelles étapes de cette recherche du confort : la possibilité de travailler à distance, ou télétraitement, et la possibilité de travailler de manière conversationnelle.

Le **télétraitement** n'est autre qu'une extension du système de chargement continu, dans lequel les travaux peuvent être soumis à l'ordinateur à partir de terminaux situés à distance, les résultats pouvant être renvoyés sur ces mêmes terminaux. Les travaux ainsi soumis sont inclus dans la file d'attente des travaux à exécuter au même titre que les travaux qui sont chargés localement, en modulant éventuellement les priorités respectives.

Les systèmes **conversationnels** permettent aux utilisateurs de suivre le déroulement des différentes étapes de leurs travaux ainsi que de réagir sur ce déroulement par l'intermédiaire de terminaux adaptés au dialogue (clavier et affichage sur oscilloscope, machine à écrire connectée, etc.). Afin de servir un grand nombre d'utilisateurs simultanément, le calculateur peut travailler en partage de temps. On entend par là qu'il alloue successivement à chaque utilisateur une tranche de son temps avec une périodicité telle que, compte tenu des temps de réponses respectifs de l'ordinateur et de l'homme, chaque utilisateur ait l'impression d'avoir la machine pour lui tout seul.

Outre de multiples applications particulières sur lesquelles nous reviendrons, le dialogue direct homme machine apporte une très grande souplesse lors de la mise au point ou de la modification des programmes.

Le concept de **machine virtuelle** relativement récent vise à simplifier la tâche du programmeur : ce dernier ne connaît qu'une machine fictive, appelée machine virtuelle, qui ne présente ni les limitations dues à la configuration de l'ordinateur utilisé (notamment en ce qui concerne la capacité de mémoire centrale), ni les limitations dues au partage de l'ordinateur avec d'autres utilisateurs (notamment en partage de temps).

**Retour sur le rôle du système d'exploitation.**

L'évolution des méthodes d'exploitation des ordinateurs permet de mieux comprendre le rôle et la complexité croissante des systèmes d'exploitation. Ces derniers tentent de réaliser, pour un type d'exploitation donné, le meilleur compromis entre les performances globales du système informatique, le « confort » d'accès des utilisateurs à l'ordinateur et la prise en compte du plus grand nombre de fonctions possible, déchargeant ainsi le travail des programmeurs, et de l'exploitant.

Réduits au chargeur et au traducteur de langage lors de la première génération, ils ont pris en charge les fonctions d'enchaînement des travaux et de gestion des entrées-sorties dès la deuxième génération, puis, successivement au cours de la troisième génération, la gestion de la multiprogrammation, du télétraitement, du temps partagé, et enfin ont éventuellement donné la possibilité de travailler en machine virtuelle.

Les rôles principaux d'un système d'exploitation moderne sont les suivants :

- (1) Gestion de l'enchaînement des travaux des utilisateurs en tenant compte de leurs priorités respectives et gestion des différentes phases de chaque travail : chargement, compilation, exécution.
- (2) Gestion des entrées et des sorties d'informations, ce qui évite au programmeur d'avoir chaque fois à écrire les programmes correspondants.
- (3) Protection du système lui-même vis-à-vis des erreurs faites par les programmeurs et des programmes les uns vis-à-vis des autres (il s'agit par exemple d'éviter qu'un programme quelconque vienne par erreur écrire n'importe quoi dans les cellules mémoires réservées au système d'exploitation ou attribuées à un autre programme).
- (4) Gestion des mémoires centrales et auxiliaires.
- (5) Prise en charge des erreurs détectées lors de l'exécution des travaux.
- (6) Tenue à jour de la comptabilité permettant de facturer les temps d'utilisation aux différents utilisateurs, etc.

L'évolution de l'architecture du hardware et l'évolution de l'architecture du software sont très complémentaires : la possibilité de simultanéité entre traitements et entrées-sorties a conduit d'abord à la gestion des entrées-sorties par le système, puis à des formes simplifiées de multiprogrammation ; mais le développement de cette dernière a entraîné la réalisation de dispositifs hardware de protection et de gestion de la mémoire indispensables aux systèmes d'exploitation en multiprogrammation, etc. Nous verrons que cette complémentarité apparaît également sous une autre forme : si certaines fonctions sont systématiquement exécutées par hardware, et certaines par software, il en est qui peuvent être soit câblées, soit programmées, le software devant, lorsque c'est possible, prendre en compte ce qu'il n'a pas été prévu de faire exécuter par hardware.

### **L'utilisation des ordinateurs.**

Théoriquement un ordinateur a un certain **caractère d'universalité** : c'est le programme qui lui donne sa destination propre. Sous réserve d'une capacité et d'une rapidité suffisantes, il peut, suivant le programme, aussi bien calculer la paye des employés d'une entreprise que la trajectoire d'une fusée spatiale. Il n'en reste pas moins vrai que chaque type d'application a ses propres caractéristiques, et que pour des raisons tant techniques qu'économiques on est conduit dans une

certaine mesure à spécialiser les machines. On utilisera certainement des ordinateurs différents au siège d'une grosse entreprise pour assurer sa gestion et à bord d'un engin spatial pour contrôler sa mission.

### **Le calcul scientifique.**

Le calcul scientifique demande aux calculateurs des possibilités arithmétiques très développées. Ils travaillent sur une représentation des nombres appelée virgule flottante qui permet de conserver automatiquement, au cours des calculs, le maximum de chiffres significatifs. Ils évoluent depuis de petits ordinateurs adaptés à un laboratoire jusqu'à de très gros systèmes pour les centres de calculs des grands organismes scientifiques.

### **La gestion.**

La gestion d'une entreprise de moyenne importance suppose essentiellement la manipulation de fichiers : fichier des employés pour le calcul de la paye, fichier des clients pour la facturation, fichier des pièces détachées pour la tenue à jour des stocks, etc. Les machines correspondantes devront donc savoir traiter et structurer des chaînes de caractères (lettres, chiffres décimaux, ponctuations), et disposer de mémoires auxiliaires importantes. Par contre, leur possibilité arithmétique pourra être réduite, les comptables n'utilisant que les nombres entiers. On trouve toute une gamme de machines de gestion allant de la facturière-comptable à l'ordinateur moyen. Une grosse entreprise qui ajoute à ses travaux de gestion des études de recherche opérationnelle et de calculs scientifiques utilisera souvent une seule machine relativement puissante qui possède à la fois les caractéristiques d'un ordinateur scientifique et celles d'un ordinateur de gestion. Ce type de machine inconnu lors de la deuxième génération a tendance à se généraliser.

### **La conduite de processus.**

La conduite de processus regroupe des problèmes d'acquisition de données, de surveillance et de contrôle dans de multiples domaines : conduite d'unités industrielles, d'expériences, d'appareils d'analyse, d'engins spatiaux ; surveillance automatique des grands malades, etc. Elle demande de petits ordinateurs ayant des possibilités d'entrées-sorties spéciales très développées (convertisseurs analogiques-numériques et numériques-analogiques, grand nombre de lignes d'entrée et de sortie numériques) ainsi qu'un système d'interruptions permettant au calculateur de se synchroniser sur le processus à conduire ou contrôler et d'être immédiatement informé des événements extérieurs susceptibles de se produire. On dit que ces calculateurs travaillent **en ligne** pour exprimer qu'ils sont connectés avec le processus extérieur. Ils permettent un contrôle en **boucle ouverte** lorsque, à partir des mesures qu'ils acquièrent, ils élaborent des informations permettant à l'homme de réagir sur le processus ; le contrôle est dit en **boucle fermée** si l'ordinateur réagit directement sur le processus au moyen d'organes de sortie spécialisés. Enfin le concept de **travail en temps réel** correspond à la nécessité pour le calculateur d'élaborer ses résultats dans des temps cohérents avec l'évolution du processus contrôlé. Un système informatique qui assure des fonctions de sécurité comprend généralement deux calculateurs susceptibles de se relayer l'un l'autre en cas de panne. Il existe une classe spéciale de calculateurs, souvent appelés calculateurs embarquables, dont la fiabilité et la miniaturisation sont particulièrement étudiées en vue de leur intégration dans des systèmes d'armes, dans des engins spatiaux, dans des systèmes de pilotage automatique, etc.

### **Les systèmes conversationnels.**

Au lieu d'être en ligne sur un processus, la machine peut, dans un système conversationnel, être considérée comme étant en ligne sur l'homme. On retrouve ainsi une forme atténuée de

temps réel : l'homme devant son terminal attend une réponse relativement rapide, mais ici le délai de réponse n'est pas impératif comme par exemple pour la commande de l'allumage des rétrofusées lors d'un atterrissage sur la Lune.

Outre la souplesse d'accès à l'ordinateur notamment pour la mise au point ou la modification des programmes, les systèmes conversationnels fonctionnant en partage de temps répondent à des applications permettant une symbiose qui paraît aujourd'hui optimale entre l'ordinateur et l'homme par la mise en commun de l'intuition de celui-ci et les possibilités de traitement et de mémoire de celui-là. On distinguera deux grands types d'application de ces systèmes.

Les systèmes d'interrogation-réponse et l'accès aux banques de données.

Les utilisateurs ont accès en temps réel, grâce à un langage d'interrogation et, pour certains d'entre eux, de mise à jour, à un ensemble d'informations structurées. On en trouve des exemples dans les systèmes de réservations de places utilisés par les compagnies aériennes ou hôtelières, les systèmes de documentation automatique, et les banques de données qui se développent dans de multiples domaines gestion d'entreprise, gestion d'un grand projet, informations économiques, informations médicales, aménagement du territoire, etc.

La conception assistée et les techniques conversationnelles graphiques.

L'utilisation en temps réel de terminaux oscilloscopiques graphiques, permettant aussi bien l'affichage et la modification de figures par l'ordinateur que l'entrée de figures dessinées par l'utilisateur, ouvre de très larges possibilités dans les domaines de la recherche et de l'art de l'ingénieur. C'est en tout cas un outil fondamental lors de la conception de nouveaux ordinateurs, tant au niveau de l'étude logique qu'au niveau de la conception et de la fabrication des circuits intégrés ou qu'au niveau de l'implantation de ces derniers sur les plaquettes imprimées.

## **Chapitre II - Structure de l'information digitale**

Ce chapitre servira de référence pour toutes les notions de base concernant la structure de l'information traitée par les ordinateurs. Il suffit donc d'en connaître les grandes lignes pour aborder la suite de l'ouvrage, sachant que le lecteur non informaticien s'y référera à l'occasion de tel ou tel développement ultérieur.

Nous montrerons d'abord comment, à partir de l'information binaire élémentaire, les techniques de codage permettent de représenter des informations complexes, ce qui, au passage, nous mènera à une digression sur la redondance et son utilisation pour la détection et la correction des erreurs. Après avoir tenté de définir les moules dans lesquels elles doivent se loger, nous passerons en revue la représentation interne des différentes informations traitées par une machine : les nombres, les caractères de la machine à écrire, les instructions. La représentation des nombres nous conduira à rappeler la notion de système de numération et à présenter le système binaire, l'octal et l'hexadécimal n'étant effleurés que pour leur intérêt de représentation condensée du binaire. Nous réservons une place plus importante aux différents formats et conventions de représentation des nombres algébriques en système binaire et en système décimal : il est nécessaire de les connaître pour aborder l'étude de l'unité arithmétique à laquelle est consacré le chapitre V. Nous passerons très rapidement sur le codage des caractères puis sur celui des instructions qui sera repris au chapitre VI.

Un bref paragraphe introduit succinctement les structures d'ensembles d'informations que nous aurons à connaître au cours de développements ultérieurs, notamment ceux concernant

les problèmes de langages évolués ou de systèmes d'exploitation.

Nous terminons ce chapitre par des paragraphes d'introduction au langage machine des calculateurs de type Von Neumann et à leur programmation. Il sera bon de le relire au cours de l'étude du chapitre VI.

## L'information digitale.

### L'information digitale élémentaire.

L'information analogique est supportée par des grandeurs physiques, telles qu'une tension ou une intensité électriques, susceptibles de varier de façon continue, c'est-à-dire par une suite de modifications arbitrairement petites de leur valeur. A l'opposé, l'information digitale ou logique est fondamentalement discontinue. Son support élémentaire est un système à  $n$  états d'équilibre, chaque état d'équilibre correspondant à une valeur de l'information, nous dirons un digit. En fait, on n'utilise dans la pratique que des systèmes à deux états d'équilibre, d'où le nom de binaire communément donné à l'élément d'information. L'information digitale élémentaire sera donc l'alternative : OUI - NON, VRAI - FAUX... Conventionnellement, nous noterons ces deux états 1 et 0, et l'information contenue sera appelée digit binaire ou **bit**<sup>2</sup> (abréviation universellement reconnue de « binary digit »). Les informations plus complexes se ramèneront à un ensemble d'informations élémentaires, grâce aux techniques de codage.

### Notion de codage.

Un ensemble de deux digits binaires peut prendre les quatre états, ou configurations binaires, suivants : 0,0; 0,1; 1,0; 1,1. Par récurrence un ensemble de  $n$  digits binaires pourra prendre  $2^n$  états et donc représenter  $2^n$  informations différentes. Le codage consiste à établir une loi de correspondance appelée code entre les informations à représenter et les configurations binaires possibles, de telle sorte qu'à chaque information corresponde une et généralement une seule configuration binaire.

Pour coder un nombre  $N$  d'informations avec  $2^{n-1} < N \leq 2^n$  il faut donc au moins  $n$  digits binaires : l'opération de codage revient à passer d'une information sur  $N$  bits dont un seul a la valeur vraie à une information codée sur  $n$  bits ; l'opération inverse constitue le décodage.

Avec 6 digits binaires, on peut coder jusqu'à  $2^6 = 64$  informations différentes, par exemple les caractères d'une machine à écrire. Lorsque l'on appuiera sur une touche, le caractère correspondant sera codé sur 6 bits avant d'être envoyé au calculateur; inversement lorsque le calculateur enverra un caractère de 6 bits à la machine à écrire, il sera décodé afin de permettre la mise en mouvement du marteau correspondant. On utilise également les techniques de décodage dans l'analyse des instructions et l'adressage des mémoires.

L'application d'un changement de code à des informations codées est appelé transcodage.

### Exemple de codage : le code de l'hérédité

L'information héréditaire de tout être vivant est conservée au sein du noyau de chacune de ses cellules sous forme digitale. C'est à partir de cette information que la cellule synthétise le type de protéine approprié. On peut se représenter une protéine comme une grosse molécule constituée d'une longue chaîne d'acides aminés. Il existe une vingtaine d'acides aminés différents et donc une

---

<sup>2</sup> REMARQUE : nous avons fait la différence entre le support de l'information et la valeur de l'information qui est assimilable à l'un des états de ce support. En pratique les termes de « digit » et de « bit » serviront à désigner aussi bien l'élément d'information que la valeur de cet élément d'information.

variété quasi infinie de protéines possibles. L'information héréditaire doit donc préciser la succession des acides aminés dans la protéine qui doit être synthétisée.

Le support de cette information est une molécule d'acide désoxyribonucléique A.D.N. (ou d'acide ribonucléique A.R.N.). Un tel acide est formé par la répétition d'un grand nombre d'éléments appelés nucléotides mis bout à bout. Du point de vue informatique, on distingue quatre types de nucléotides possibles que, pour simplifier, nous noterons U, C, A et G.

Le problème du code de l'hérédité revient à déterminer la façon dont la succession des nucléotides dans une chaîne d'A.D.N. peut être l'image codée de la succession des acides aminés de la protéine. L'état actuel des recherches permet d'avancer l'hypothèse suivante. L'élément d'information significatif appelé codon est formé par la succession de trois nucléotides ; il peut coder soit un acide aminé, soit éventuellement une ponctuation dans la chaîne. Ainsi, par exemple, le codon UUU représenterait la phénylalanine, le codon GAG la glycine, le codon AUU la thyrosine, etc.

Le code de l'hérédité peut donc se représenter sous forme d'un tableau à deux colonnes, celle des codons et celle des acides aminés correspondants, et à 64 lignes puisqu'il y a  $4^3=64$  codons possibles.

Ce code n'est pas biunivoque : on a notamment constaté que plusieurs acides aminés pouvaient correspondre au même codon.

Après synthèse de la protéine, la chaîne des acides aminés qui la constitue est l'image, à travers cette table de codage, de la chaîne des nucléotides de l'A.D.N. Ce code paraît universel, tout au moins en ce qui concerne les formes de vie terrestre actuellement connues.

### Notion de code redondant

Lorsque l'on redoute une modification intempestive de l'information lors de sa transmission ou de son stockage en mémoire, on est conduit à utiliser des codes permettant soit de détecter les erreurs - codes **autovérificateurs** - soit de les corriger - codes **autocorrecteurs**. Ceci implique que les codes portent sur un nombre de digits binaires supérieur à celui strictement nécessaire pour coder l'information selon les principes précédemment exposés, d'où le qualificatif de redondant qui leur est souvent attribué. En voici quelques exemples.

#### Contrôle deux sur trois.

Concrétisons ces notions sur un exemple grossier mais significatif.

Au lieu de transmettre une information A codée sur n bits, on transmet l'information A répétée trois fois sur  $3^n$  bits.

Lors de l'analyse de l'information après transmission, trois éventualités peuvent se présenter

- (1) on a obtenu trois informations identiques : on en déduit que la transmission a été correcte (à moins qu'il n'y ait eu trois erreurs identiques, ce qui est hautement improbable);
- (2) on a obtenu deux informations identiques, la troisième était différente on en déduit qu'il y a eu une erreur sur la troisième, et que les deux premières correspondent à l'information correcte.

Le code est alors **autocorrecteur**;

(3) on a obtenu trois informations différentes. On sait qu'il y a deux erreurs au moins, et l'on n'est plus capable de reconstituer l'information correcte : le code n'est plus qu'**autovérificateur**.

## Contrôle de parité

Dans le domaine des calculateurs où les probabilités d'erreurs aléatoires sont très faibles, on emploie souvent un code autovérificateur dit **code de parité**, à  $n + 1$  bits, dans lequel les  $n$  premiers bits servent à coder de façon significative  $2^n$  informations. Le  $(n+1)^{\text{ème}}$  bit, dit « de parité », est positionné de telle sorte que le nombre total de bits 1 soit pair (code à parité paire) ou impair (code à parité impaire). Ce code n'est que partiellement autovérificateur car il ne permet pas de détecter une double erreur (deux bits erronés sur une même information).

## Codes autocorrecteurs de Hamming

Hamming a utilisé les tests de parité pour construire des codes **autocorrecteurs**. Voici le raisonnement de base dans le cas d'un code capable de corriger une seule erreur.

Etant donné une information de  $i$  digits binaires à transmettre, on ajoute  $p$  digits de parité, le calcul de la parité de chacun de ces digits portant sur un certain nombre de digits du message de  $i + p$  digits transmis. On associe  $p$  digits de test aux  $p$  digits de parité. Si, après transmission, la parité n'est pas vérifiée sur un digit de parité, on convient de donner la valeur 1 au digit de test correspondant ; on lui attribue la valeur 0 dans le cas contraire.

On choisit  $p$  suffisamment grand pour que les digits de test puissent coder soit l'absence d'erreur (les  $p$  bits sont à zéro) soit la position de l'erreur dans le message de  $i + p$  digits, en supposant qu'il n'y en a qu'une. Ayant la position de l'erreur, il suffit de changer la valeur du bit concerné pour rectifier le message.

Pour que les  $p$  digits de test puissent coder les  $i + p$  positions du message ainsi que l'absence d'erreur, il faut choisir  $p$  de telle sorte qu'il vérifie l'inégalité

$$2^p \geq i + p + 1 .$$

Le tableau suivant indique les valeurs maximales de  $i$  pour les premières valeurs de  $p$

$p$	2	3	4	5	6
$i$	1	4	11	26	57

On trouvera au paragraphe 2. 3 un exemple concret avec  $i = 4$  et  $p = 3$ .

## Codes autovérificateurs de type $n$ dont $p$

Ces codes, également appelés codes  $p$  parmi  $n$ , correspondent à un codage sur  $n$  bits sachant que l'on s'astreint à n'utiliser que les codes pour lesquels  $p$  bits parmi les  $n$  ont la valeur vraie. Ainsi le code 8 dont 4, qui utilise des messages de 8 digits dont 4 sont positionnés à 1 et 4 à 0, permet de représenter 70 informations différentes.

Ce type de code ne permet pas de détecter les doubles erreurs pour lesquelles un bit est passé intempestivement de 0 à 1 tandis qu'un autre est passé de 1 à 0.

## Utilisation des codes redondants

Dans les machines, les codes redondants sont surtout utilisés au niveau du stockage en mémoire centrale, du stockage en mémoire auxiliaire et de certains échanges d'information avec les périphériques électromécaniques.

Compte tenu de la mauvaise fiabilité des composants de l'époque, les premières machines ont souvent été munies de systèmes de codage permettant un contrôle poussé de la validité de l'information après un temps de stockage en mémoire ou après certains traitements. Le contrôle de parité a par la suite été jugé suffisant par la plupart des concepteurs d'ordinateurs de la deuxième et de la troisième génération, sauf dans le cas de machines très spécialisées pour lesquelles une très haute disponibilité était recherchée. Notons cependant que dans la série IBM 370, on trouve des codes auto correcteurs corrigeant à coup sûr une erreur de 1 bit au niveau des échanges avec la mémoire centrale et plusieurs erreurs portant sur des bits successifs au niveau des échanges avec les disques magnétiques (reconstitution d'informations perdues par défaut de surface).

Les codes de type  $n$  dont  $p$ , quelquefois utilisés dans les machines de la première génération, sont actuellement réservés au domaine des transmissions.

### **Taille des informations dans un calculateur.**

Un des choix importants dans un projet de calculateur réside dans la définition de la taille, exprimée en nombre de bits, des informations qui devront être manipulées. On distingue deux concepts de base : le concept de caractère et le concept de mot.

#### **Le caractère**

Ce premier concept correspond au besoin de coder en machine les différents caractères de l'alphabet (lettres, chiffres, signes). Au départ, une longueur de 6 bits, permettant de coder 64 caractères différents, a été presque universellement retenue, à tel point que le mot caractère désigne souvent un ensemble de 6 bits. Aujourd'hui le caractère de 8 bits ou octet est le plus généralement adopté dans les machines orientées vers la gestion. Il permet de coder des alphabets dépassant 64 signes, ou encore de superposer au codage des différents caractères une marque de fin de chaîne de caractères.

#### **Le mot**

Le concept de mot est beaucoup plus difficile à définir que celui de caractère. Ce terme a, en effet, plusieurs acceptions.

(1) Le mot apparaît d'abord comme l'information de rang supérieur au caractère. Il contient généralement un nombre entier de caractères et varie entre 10 et 64 bits.

(2) Le mot apparaît ensuite comme l'unité d'information traitée par la machine. Dans Boulix par exemple, une instruction et un opérande ont la même taille : on parlera des mots machine.

(3) Généralement, et c'est le cas de Boulix, le mot machine est effectivement traité d'un seul bloc, aussi bien au niveau de l'accès mémoire et des transferts qu'au niveau du traitement : on l'appellera alors mot mémoire ou mot technologique.

(4) Dans des machines telles que celles de la série IBM 360, on traite des informations de 8, 16, 32 et 64 digits. Le mot machine est alors choisi arbitrairement à 32 digits. On appelle alors demi-mot l'ensemble de 16 digits et double-mot l'ensemble de 64 digits. La taille du mot mémoire varie de 8 à 64 bits selon le modèle dans la série.

#### **Machines caractères et machinés mots**

Certaines machines orientées vers la gestion ou le traitement de l'information non numérique travaillent exclusivement sur des chaînes de caractères. On les appelle souvent machines caractères. A l'opposé, des machines strictement orientées vers le calcul scientifique ne travaillent que sur les mots. On les appelle machines mots. Des machines telles que celles de la série IBM 360 travaillent aussi bien sur des mots que sur des chaînes de caractères.

### Notion de contenant et de contenu.

Dans une machine, l'information digitale doit être conservée sur un certain support physique qui, comme nous l'avons vu au début de ce chapitre, est un système à plusieurs états d'équilibre. Par exemple pour contenir un caractère de 6 bits, il faut un support de 6 éléments à deux positions d'équilibre (ce qui correspond à un support de 64 états d'équilibre). Ce support pourra être un registre, une cellule mémoire, un emplacement sur un ruban magnétique ou sur un disque, etc.

Afin de ne pas confondre le contenant de l'information et le contenu, à savoir l'information elle-même, nous désignerons le contenant par un symbole et le contenu par ce même symbole placé entre parenthèses.

Exemple : soit un registre R, le contenu de R est (R) ; soit AD une adresse de cellule mémoire, le contenu de cette cellule est (AD) ; si (R)=(AD), alors le contenu de la cellule d'adresse AD sera (AD)=(R). On notera le transfert du contenu du registre R<sub>1</sub> dans le registre R<sub>2</sub> par (R<sub>1</sub>) → R<sub>2</sub>, ce qui se lit : le contenu de R<sub>1</sub> est transféré dans R<sub>2</sub>.

### La représentation des nombres.

#### Rappel sur les systèmes de numération.

On peut représenter le nombre N par N tirets verticaux alignés côte à côte. Historiquement, plusieurs systèmes de représentation plus condensés ont vu le jour. Nous citerons, par exemple, le système romain qui ne permet pas l'utilisation d'algorithmes simples pour la réalisation des opérations arithmétiques courantes (il suffit, pour s'en convaincre, d'essayer de diviser MMXLVIII par LXIV).

Le système décimal qui nous vient des Arabes, qui le tenaient probablement de l'Inde, est un cas particulier des systèmes de numération qui sont d'un emploi très commode pour le calcul arithmétique.

Rappelons leur définition.

Pour représenter un nombre N d'objets, on les groupe par paquets de n, ces paquets de n étant eux-mêmes groupés par paquets de n et ainsi de suite. La formulation mathématique correspondante est la suivante

$$N = a_p \cdot n^p + a_{p-1} \cdot n^{p-1} + \dots + a_1 \cdot n + a_0 \quad (1)$$

Où

$a_0$  = le nombre d'objets (< n) qui n'ont pu être classés dans un paquet de n objets ;  
 $a_1$  = le nombre de paquets (< n) de n objets qui n'ont pu être classés dans un paquet de n<sup>2</sup> objets, etc.

n est appelé base du système de numération.

On remarque que  $n-1$  symboles (de 1 à  $n-1$ ) sont nécessaires pour représenter les  $a_p$ .

On leur adjoint un  $n^{\text{ième}}$  symbole 0 dont le rôle est fondamental : s'il n'y a aucun groupe de  $n^i$  objets, au lieu d'omettre le terme correspondant dans la formule (1), on le fera figurer avec  $a_i = 0$ .

Tous les termes  $a_i$  ( $i < p$ ) étant alors représentés, on pourra transcrire le nombre  $N$  sous la forme  $N = a_p a_{p-1} \dots a_1 a_0$  (2)

Compte tenu de sa construction, cette représentation de  $N$  est unique.

La position  $i$  occupée par le symbole  $a_i$  indique le type de paquets de  $n^i$  objets auquel il se rapporte : nous dirons encore le poids attaché à ce symbole. Le 0 placé à droite d'un symbole apparaît ainsi comme un multiplicateur par la base :  $a_0$  représente  $a$  paquets de  $n$  objets =  $a \cdot n$  ;

$a00$  représente  $a$  paquets de  $n^2$  objets =  $a \cdot n^2$ , etc.

Le système décimal étant le système de numération à base 10,

$a_0$  représente le chiffre des unités (poids 1),

$a_1$  celui des dizaines (poids 10),

$a_2$  celui des centaines (poids 100), etc.

Le choix du système décimal semble arbitraire, et d'aucuns prétendent qu'il serait préférable de l'abandonner en faveur du système duodécimal, à base 12, ce chiffre étant divisible par 1, 2, 3, 4, 6 et 12, tandis que 10 n'est divisible que par 1, 2, 5 et 10. (Notons qu'en duodécimal, il faut ajouter deux symboles pour représenter le 10 et le 11.)

Quand le système de numération dans lequel on travaille n'est pas implicitement évident, on le rappelle en indice

Exemple :  $100101_2$  pour du binaire  $128_{10}$  pour du décimal  $128_{12}$  pour du duodécimal.

### Le système binaire.

Introduit par Leibniz au XVII<sup>ème</sup> siècle, le système binaire ou système de numération à base deux est la numération qui apparaît comme la plus évidente dans les machines électroniques puisque, comme nous l'avons vu, elles utilisent essentiellement des systèmes à deux états d'équilibre. Les 10 premiers nombres binaires s'écrivent :

Décimal :		Binaire :	
	1		1
	2		10
	3		11
	4		100
	5		101
	6		110
	7		111
	8		1000

9

1001

10

1010

Les tables d'addition et de multiplication sont

addition :

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 10$$

Multiplication :

$$0 \times 0 = 0$$

$$0 \times 1 = 0$$

$$1 \times 0 = 0$$

$$1 \times 1 = 1$$

Exemples d'opérations :

**Exemples d'opérations :**

$$\begin{array}{r}
 101101 \\
 \times \quad 101 \\
 \hline
 101101 \\
 101101 \\
 \hline
 11100001
 \end{array}$$

$$\begin{array}{r}
 101101 \quad 11 \\
 \hline
 101 \quad 1111 \\
 100 \\
 011 \\
 00
 \end{array}$$

### Conversion binaire-décimal.

L'algorithme utilisé, qui est une application directe de la formule (1), est explicité sur l'exemple suivant

$$\begin{array}{r}
 1101,11 = 1 \times 2^3 = 8 \\
 + 1 \times 2^2 = 4 \\
 + 0 \times 2^1 = 0 \\
 + 1 \times 2^0 = 1 \\
 + 1 \times 2^{-1} = 1/2 \\
 + 1 \times 2^{-2} = 1/4 \\
 \hline
 13,75
 \end{array}$$

Il est utile de connaître la table des puissances de 2, dont nous donnons à titre indicatif, les premiers éléments dans le tableau 1.

$2^{-n}$	$n$	$2^n$
1	0	1
0.5	1	2
0.25	2	4
0.125	3	8
0.0625	4	16
0.03125	5	32
0.015625	6	64
0.0078125	7	128
0.00390625	8	256
0.001953125	9	512
0.0009765625	10	1024
0.00048828125	11	2048
0.000244140625	12	4096

### Conversion décimal-binaire.

On sépare la partie entière de la partie fractionnaire. Les algorithmes utilisés pour la conversion de chacune de ces parties sont une application directe de la méthode utilisée pour construire la formule (1); ils sont explicités sur l'exemple suivant.  
Soit à convertir 125,625

**Partie entière.** On divise le nombre par 2 autant qu'il est possible, les restes successifs étant les poids binaires obtenus dans l'ordre des puissances croissantes

$$\begin{aligned}
 125 : 2 &= 62 + 1 \\
 62 : 2 &= 31 + 0 \\
 31 : 2 &= 15 + 1 \\
 15 : 2 &= 7 + 1 \\
 7 : 2 &= 3 + 1 \\
 3 : 2 &= 1 + 1 \\
 1 : 2 &= 0 + 1
 \end{aligned}$$

Le résultat est : 1111101.

**Partie fractionnaire.** On multiplie la partie fractionnaire par 2, la partie entière obtenue représentant le poids binaire et la partie fractionnaire étant à nouveau multipliée par 2, etc.

$$\begin{array}{ll}
 0.625 \times 2 = 1,250 & \text{Poids binaire : } 1 \times 2^{-1} \\
 0.250 \times 2 = 0,500 & \text{Poids binaire : } 0 \times 2^{-2} \\
 0.500 \times 2 = 1,000 & \text{Poids binaire : } 1 \times 2^{-3}
 \end{array}$$

Le résultat est : 0,101;  
on a donc :  $125,625_{10} = 1111101,101_2$ .

En marge du système binaire.

### Application aux codes de Hamming

Ce paragraphe peut être considéré comme une application du système binaire à la conception des codes autovérificateurs sur une seule erreur.

Reprenons le développement du paragraphe 1. 1. 3. dans l'hypothèse où l'information à transmettre est de  $i = 4$  digits,

$p$	2	3	4	5	6
$i$	1	4	11	26	57

ce qui implique l'adjonction de  $p = 3$  digits de parité.

Organisons le message de  $i + p$  digits en numérotant les digits de 1 à 7 de droite vers la gauche et en plaçant les digits  $p$  dans les positions 1, 2 et 4

7	6	5	4	3	2	1
$i$	$i$	$i$	$p$	$i$	$p$	$p$

On veut que les 3 digits de test indiquant, comme décrit au paragraphe 1. 3. 3., le résultat du test de parité sur les 3 digits de parité donnent en binaire la position du digit erroné.

Ainsi la configuration 000 indiquera l'absence d'erreurs, la configuration 001 une erreur sur le bit de position 1, la configuration 010 une erreur sur le bit de position 2, etc.

Ces considérations permettent de déterminer sur quelles positions, parmi les 7, doit porter le contrôle de parité de chacun des 3 digits de parité.

On remarque en effet qu'une erreur détectée sur le bit de parité de position 1 ne doit pouvoir provenir que d'erreurs concernant les digits de positions 1, 3, 5 ou 7 correspondant respectivement aux configurations 001, 011, 101 et 111 des digits de test. On en déduit immédiatement que la parité du digit  $p$  de position 1 doit porter sur les digits de position 1, 3, 5 et 7 du message à transmettre.

On vérifiera de même que la parité du digit  $p$  de position 2 doit porter sur les digits de position 2, 3, 6 et 7, et que celle du digit  $p$  de position 4 doit porter sur les digits de position 4, 5, 6 et 7.

Exemple : soit à transmettre le message 1010.

7	6	5	4	3	2	1
$i$	$i$	$i$	$p$	$i$	$p$	$p$
1	0	1	-	0	-	-

Le digit  $p$  de position 1 doit rendre paire la somme de digits de position 1, 3, 5 et 7 ; il est donc égal à 0

7	6	5	4	3	2	1
$i$	$i$	$i$	$p$	$i$	$p$	$p$
1	0	1	-	0	-	<b>0</b>

La parité du digit p de position 2 porte sur les digits de position 2, 3, 6 et 7.

7	6	5	4	3	2	1
i	i	i	p	i	p	p
1	0	1	-	0	1	0

Enfin la parité du digit p de position 4 porte sur les digits de position 4, 5, 6 et 7 :

7	6	5	4	3	2	1
i	i	i	p	i	p	p
1	0	1	0	0	1	0

Supposons que pendant la transmission du message une erreur s'introduise sur le bit de position 5

7	6	5	4	3	2	1
i	i	i	p	i	p	p
1	0	0	0	0	1	0

après transmission le test de parité sera vérifié pour le digit p de position 2, et non vérifié pour les digits p de position 1 et 4. On en déduit que les digits de test prennent la configuration 101 qui indique en binaire la position du bit erroné.

On rectifiera le message en changeant la valeur du bit de position 5.

### Le contrôle d'erreurs

Le codage binaire est très pratique pour une utilisation dans des appareils électroniques tels qu'un ordinateur, dans lesquels l'information peut être codée grâce à la présence ou non d'un signal électrique.

Pendant le signal électrique peut subir des perturbations (distorsion, présence de bruit), notamment lors du transport des données sur un long trajet. Ainsi, le contrôle de la validité des données est nécessaire pour certaines applications (professionnelles, bancaires, industrielles, confidentielles, relatives à la sécurité, ...).

C'est pourquoi il existe des mécanismes permettant de garantir un certain niveau d'**intégrité** des données, c'est-à-dire de fournir au destinataire une assurance que les données reçues sont bien similaires aux données émises. La protection contre les erreurs peut se faire de deux façons : soit en fiabilisant le support de transmission, c'est-à-dire en se basant sur une protection **physique**. Une liaison conventionnelle a généralement un taux d'erreur compris entre  $10^{-5}$  et  $10^{-7}$ .

soit en mettant en place des mécanismes **logiques** de détection et de correction des erreurs. La plupart des systèmes de contrôle d'erreur au niveau logique sont basés sur un ajout d'information (on parle de « redondance ») permettant de vérifier la validité des données. On appelle **somme de contrôle** cette information supplémentaire.

### La correction d'erreurs

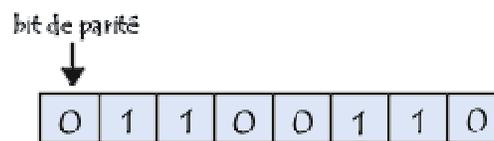
C'est ainsi que des systèmes de détection d'erreur plus perfectionnés ont été mis au point, ces codes sont appelés :

- codes autocorrecteurs
- codes autovérificateurs

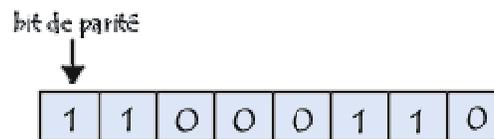
### Le contrôle de parité

Le contrôle de parité (appelé parfois VRC, pour Vertical Redundancy Check ou Vertical Redundancy Checking) est un des systèmes de contrôle les plus simples. Il consiste à ajouter un bit supplémentaire (appelé **bit de parité**) à un certain nombre de bits de données appelé mot de code (généralement 7 bits, pour former un octet avec le bit de parité) dont la valeur (0 ou 1) est telle que le nombre total de bits à 1 soit pair. Pour être plus explicite il consiste à ajouter un 1 si le nombre de bits du mot de code est impair, 0 dans le cas contraire.

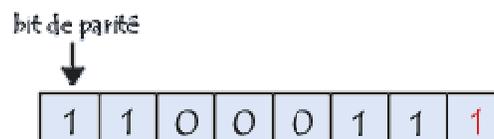
Prenons l'exemple suivant :



Dans cet exemple, le nombre de bits de données à 1 est pair, le bit de parité est donc positionné à 0. Dans l'exemple suivant, par contre, les bits de données étant en nombre impair, le bit de parité est à 1 :

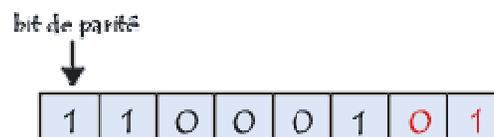


Imaginons désormais qu'après transmission le bit de poids faible (le bit situé à droite) de l'octet précédent soit victime d'une interférence :



Le bit de parité ne correspond alors plus à la parité de l'octet : **une erreur est détectée.**

Toutefois, si deux bits (ou un nombre pair de bits) venaient à se modifier simultanément lors du transport de données, aucune erreur ne serait alors détectée...



Le système de contrôle de parité ne détectant que les erreurs en nombre impair, il ne permet donc de détecter que 50% des erreurs. Ce système de détection d'erreurs possède également l'inconvénient majeur de ne pas permettre de corriger les erreurs détectées (le seul moyen est d'exiger la retransmission de l'octet erroné...).

## Le contrôle de parité croisé

Le contrôle de parité croisé (aussi appelé contrôle de redondance longitudinale ou Longitudinal Redundancy Check, noté **LRC**) consiste non pas à contrôler l'intégrité des données d'un caractère, mais à contrôler l'intégrité des bits de parité d'un bloc de caractères. Soit « HELLO » le message à transmettre, en utilisant le code ASCII standard. Voici les données telles qu'elles seront transmises avec les codes de contrôle de parité croisé :

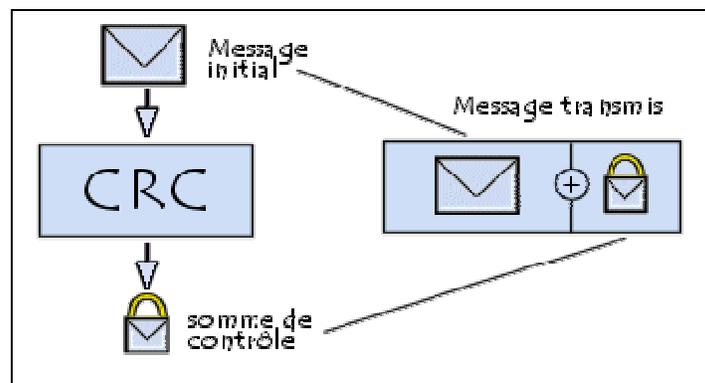
Lettre	Code ASCII (sur 7 bits)	Bit de parité (LRC)
H	1001000	0
E	1000101	1
L	1001100	1
L	1001100	1
O	1001111	1
<b>VRC</b>	1000010	0

## Le contrôle de redondance cyclique

Le **contrôle de redondance cyclique** (noté **CRC**, ou en anglais Cyclic Redundancy Check) est un moyen de contrôle d'intégrité des données puissant et facile à mettre en oeuvre. Il représente la principale méthode de détection d'erreurs utilisée dans les télécommunications.

Principe

Le **contrôle de redondance cyclique** consiste à protéger des blocs de données, appelés trames (frames en anglais). A chaque trame est associé un bloc de données, appelé code de contrôle (parfois CRC par abus de langage ou FCS pour Frame Check Sequence dans le cas d'un code de 32 bits). Le code CRC contient des éléments redondants vis-à-vis de la trame, permettant de détecter les erreurs, mais aussi de les réparer.



Le principe du CRC consiste à traiter les séquences binaires comme des polynômes binaires, c'est-à-dire des polynômes dont les coefficients correspondent à la séquence binaire. Ainsi la séquence binaire 0110101001 peut être représentée sous la forme polynomiale suivante :

$$0 \cdot X^9 + 1 \cdot X^8 + 1 \cdot X^7 + 0 \cdot X^6 + 1 \cdot X^5 + 0 \cdot X^4 + 1 \cdot X^3 + 0 \cdot X^2 + 0 \cdot X^1 + 1 \cdot X^0$$

soit

$$X^8 + X^7 + X^5 + X^3 + X^0$$

ou encore

$$X^8 + X^7 + X^5 + X^3 + 1$$

De cette façon, le bit de poids faible de la séquence (le bit le plus à droite) représente le degré 0 du polynôme ( $X^0 = 1$ ), le 4<sup>ème</sup> bit en partant de la droite représente le degré 3 du polynôme ( $X^3$ )...

Une séquence de n bits constitue donc un polynôme de degré maximal n-1. Toutes les expressions polynomiales sont manipulées par la suite avec une arithmétique modulo 2.

Dans ce mécanisme de détection d'erreur, un polynôme prédéfini (appelé polynôme générateur et noté G(X)) est connu de l'émetteur et du récepteur.

La détection d'erreur consiste pour l'émetteur à effectuer un algorithme sur les bits de la trame afin de générer un CRC, et de transmettre ces deux éléments au récepteur.

Il suffit alors au récepteur d'effectuer le même calcul afin de vérifier que le CRC est valide.

### Application pratique

Soit M le message correspondant aux bits de la trame à envoyer et M(X) le polynôme associé. Appelons M' le message transmis, c'est-à-dire le message initial auquel aura été concaténé le CRC de n bits. Le CRC est tel que  $M'(X)/G(X)=0$ . Le code CRC est ainsi égal au reste de la division polynomiale de M(X) (auquel on a préalablement concaténé n bits nuls correspondant à la longueur du CRC) par G(X).

Le plus simple est encore de prendre un exemple : prenons le message M de 16 bits suivant : 1011 0001 0010 1010 (noté B12A en hexadécimal).

Prenons  $G(X) = X^3 + 1$  (représenté en binaire par 1001).

Etant donné que G(X) est de degré 3, il s'agit d'ajouter 4 bits nuls à M : 10110001001010100000.

Le CRC est égal au reste de la division de M par G :

10110001001010100000

1001.....

-----

0100.....

0000.....

-----

1000.....

0000.....

-----

1000.....

1001.....

-----

1111.....

1001.....

-----

1100.....

```

1001,,,,.....
----,,,,.....
1101,,,,.....
1001,,,,.....
----,,,,.....
1000,,,,.....
0000,,,,.....
----,,,,.....
10001,.....
1001,,.....
----,,.....
10000,,.....
1001,,.....
----
1111,.....
1001,.....
----,.....
1100.....
1001.....
----.....
1100....
1001....
----....
1010...
1001...
----...
0110..
0000..
----.
1100.
1001.
----.
1010
1001
----
0011

```

Pour créer M' il suffit de concaténer le CRC ainsi obtenu aux bits de la trame à transmettre :

$$M' = 1011000100101010 + 0011$$

$$M' = 10110001001010100011$$

Ainsi, si le destinataire du message effectue la division de M' par G, il obtiendra un reste nul si la transmission s'est effectuée sans erreur :

```

10110001001010100011
1001,,,,,,
----,,,,,,
0100,,,,,,
0000,,,,,,
----,,,,,,
1000,,,,,,

```

```

1001,,,,
----,,
0010,,
0000,,
----,,
0101,,
0000,,
----,,
1010,,
1001,,
----,,
0110,,
0000,,
----,,
1101,,
1001,,
----,,
1010,,
1001,,
----,,
0111,,
0000,,
----
1110,,
1001,,
----,,
1111,,
1001,,
----
1100,,
1001,,
----
1010,,
1001,,
----
0110,,
0000,,
----,
1101,
1001,
----,
1001
1001
----
0

```

Polynômes générateurs :

Les polynômes générateurs les plus couramment employés sont :

- **CRC-12** :  $X^{12} + X^{11} + X^3 + X^2 + X + 1$
- **CRC-16** :  $X^{16} + X^{15} + X^2 + 1$

- **CRC CCITT V41** :  $X^{16} + X^{12} + X^5 + 1$   
(Ce code est notamment utilisé dans la procédure HDLC.)
- **CRC-32 (Ethernet)** :  $X^{32} + X^{26} + X^{23} + X^{22} + X^{16} + X^{12} + X^{11} + X^{10} + X^8 + X^7 + X^5 + X^4 + X^2 + X + 1$
- **CRC ARPA** :  $X^{24} + X^{23} + X^{17} + X^{16} + X^{15} + X^{13} + X^{11} + X^{10} + X^9 + X^8 + X^5 + X^3 + 1$

### Les codes de Gray et le binaire réfléchi

Si la représentation binaire des nombres est très commode au niveau des calculs, il n'en est pas de même lorsqu'il s'agit de lire au vol des grandeurs digitales en cours d'évolution (en provenance de compteurs ou de codeurs par exemple).

Lorsqu'on lit un compteur binaire au moment précis où il passe de 01111 à 10000, on peut obtenir n'importe quelle valeur comprise entre 0 et 10000, dès lors que les différents digits ne sont pas changés avec une simultanéité parfaite.

Les codes de Gray, du nom de leur inventeur, évitent cet inconvénient.

Ils se caractérisent par le fait que les configurations binaires représentant deux nombres successifs ne se différencient que par la modification d'un seul digit.

La figure 2 indique comment on peut construire un code de ce type. On construit d'abord le tableau matriciel des quatre configurations binaires possibles d'un nombre de 2 digits  $a_1 a_0$ .

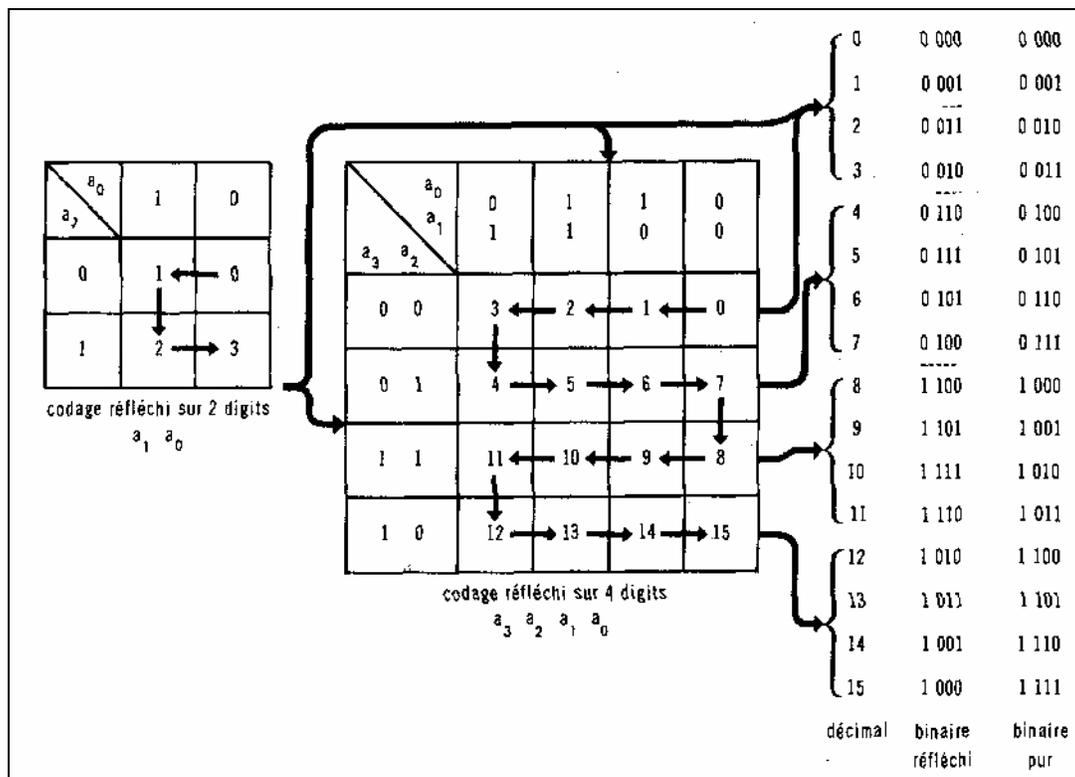


Fig. 2 – Construction du code binaire réfléchi

Dire que deux configurations binaires représentant des nombres successifs ne doivent différer que d'un digit, c'est dire qu'elles doivent se représenter dans deux cases adjacentes du tableau. On en déduit un cheminement possible correspondant au codage suivant

Décimal	$a_1$	$a_0$
0	0	0
1	0	1
2	1	1
3	1	0

A partir de ce premier code sur deux digits, on construit un code sur 4 digits  $a_3 a_2 a_1 a_0$  en traçant le tableau matriciel des 16 configurations binaires possibles et en choisissant dans ce tableau un cheminement qui ne permettra que le passage d'une case à une case adjacente, etc.

Le balayage horizontal dessiné sur la figure conduit au code de Gray le plus utilisé qui est appelé binaire réfléchi. (Son nom provient de ce que, si l'on établit une coupure à un code de type  $2^n$ , il y a symétrie, à l'exception du bit de poids fort, entre les  $2^n$  premiers codes et les  $2^n$  codes suivants.) On note que le binaire réfléchi est une représentation des nombres qui ne correspond pas à un système de numération.

Il n'est donc pas utilisé pour les calculs, mais seulement dans les organes d'acquisition que le calculateur peut être amené à lire au moment où leur valeur évolue.

### Numération octale et hexadécimale.

Si la représentation binaire est très commode pour le calculateur, il n'en est pas de même pour l'utilisateur qui préfère une écriture plus condensée.

BINAIRE	OCTAL	HEXAD.	DÉCIMAL
0	0	0	0
1	1	1	1
10	2	2	2
11	3	3	3
100	4	4	4
101	5	5	5
110	6	6	6
111	7	7	7
1000	10	8	8
1001	11	9	9
1010	12	A	10
1011	13	B	11
1100	14	C	12
1101	15	D	13

1110	16	E	14
1111	17	F	15
10000	20	10	16

Table. 3. Représentations octale et hexadécimale.

Suivant que la machine travaille sur des caractères de 6 bits ou sur des octets, on convient de regrouper les bits par 3 ou 4, ce qui revient à travailler en octal ou en hexadécimal qui sont les systèmes de numération de base 8 et 16.

Nous donnons les tables de correspondance entre binaire, décimal, octal et hexadécimal. On note que, pour l'hexadécimal, il faut ajouter aux dix symboles utilisés en numération décimale six symboles A, B, C, D, E et F représentant les nombres décimaux de 10 à 15 (tabl. 3).

Les représentations octale et hexadécimale d'un nombre se déduisent de sa représentation binaire par décomposition en groupe de 3 ou 4 bits à partir de la virgule et par remplacement de ces groupes par le chiffre correspondant du tableau 3.

Par exemple, le nombre  $10110110111,11101_2$  s'écrira également  $2667,72_8$  et  $5B7,E8_{16}$ , selon les décompositions suivantes

$$\begin{array}{cccccc} 10 & 110 & 110 & 111, & 111 & 010 \\ 2 & 6 & 6 & 7, & 7 & 2 \end{array}$$

$$\begin{array}{ccccc} 101 & 1011 & 0111, & 1110 & 1000 \\ 5 & B & 7, & E & 8 \end{array}$$

L'arithmétique octale et hexadécimale ainsi que les conversions avec le système décimal se déduisent directement de ce que nous avons présenté dans le cadre de la numération binaire.

### Représentation des nombres binaires négatifs.

La solution la plus immédiate est de réserver un digit binaire pour le signe, les autres digits représentant la valeur absolue.

La convention généralement adoptée consiste à représenter le signe + par le digit 0 et le signe - par le digit 1.

Ainsi : 0.11011 représentera  $27_{10}$  et 1.11011 représentera  $-27_{10}$ .

Cette représentation impose un traitement spécial du signe et, de ce fait, des circuits différents pour l'addition et la soustraction.

Cet inconvénient disparaît si l'on utilise la représentation des nombres négatifs sous forme complémentée. On distingue le complément vrai ou complément à 2 en binaire (complément à 10 en décimal) et le complément restreint ou complément à 1 en binaire (complément à 9 en décimal).

On obtient le complément restreint en soustrayant chaque digit de (n-1), n étant la base du système de numération, et le complément vrai en ajoutant 1 au complément restreint ainsi obtenu.

En système binaire, le complément restreint est obtenu par substitution des 0 par des 1 et vice versa.

Le tableau 4 compare, dans le cas du binaire, ces trois systèmes de représentation qui sont les plus utilisés, en supposant que l'on dispose de nombres de 4 digits, signe compris.

ÉQUIVALENT DÉCIMAL	SIGNE ET VALEUR ABSOLUE	COMPLÉMENT RESTREINT	COMPLÉMENT VRAI (à 2)
7	0111	0111	0111
6	0110	0110	0110
2	0010	0010	0010
1	0001	0001	0001
+ 0	0000	0000	
- 0	1000	1111	0000
- 1	1001	1110	1111
- 2	1010	1101	1110
- 6	1110	1001	1010
- 7	1111	1000	1001
- 8	-	-	1000

Tabl. 4. Les trois représentations des nombres algébriques binaires.

On remarque sur ce tableau qu'en complément à 2 on a une seule représentation de zéro tandis qu'il en existe deux dans les deux autres systèmes correspondant au + 0 et au - 0. Dans une machine travaillant en complément vrai, la soustraction sera obtenue par simple addition du complément, en ne tenant pas compte de la retenue. L'exemple qui suit est exprimé en numération décimale pour plus de clarté :

<b>SOUSTRACTION NORMALE</b> $\begin{array}{r} 63 \\ -28 \\ \hline 35 \end{array}$	complémentation →	<b>SOUSTRACTION COMPLÉMENT À 10</b> $\begin{array}{r} 63 \\ +72 \\ \hline 135 \end{array}$
--	-------------------	---

Dans une machine travaillant en complément restreint, la soustraction sera obtenue par addition du complément et report de la retenue. Le même exemple donne

<b>SOUSTRACTION NORMALE</b> $\begin{array}{r} 63 \\ -28 \\ \hline 35 \end{array}$	complémentation →	<b>SOUSTRACTION EN COMPLÉMENT À 9</b> $\begin{array}{r} 63 \\ +71 \\ \hline 134 \\ + \quad \underline{1} \\ \hline 35 \end{array}$
--	-------------------	---

On notera la complémentation par une barre au-dessus de la grandeur concernée :  $\overline{V}$  désigne le complément restreint de V. Représentation des nombres décimaux.

Certaines machines, notamment pour les applications de gestion, calculent en système décimal. On les appelle souvent machines décimales par opposition aux machines purement binaires. Comme on ne dispose pas de système élémentaire à 10 états d'équilibre, les digits décimaux doivent être codés à partir d'éléments binaires.

Quatre bits sont nécessaires puisque 10 est compris entre  $2^3$  et  $2^4$ . Mais comme quatre bits permettent de coder 16 informations différentes, nous avons  $A_{16}^{10}$  (de l'ordre de  $10^{10}$ ) façons de répartir les 16 codes possibles entre les 10 caractères à représenter.

Bien évidemment on choisit parmi ce nombre quasi illimité les systèmes de codage les plus commodes pour l'utilisation cherchée.

Le tableau 5 en donne trois parmi les plus usités, le code BCD (qui est une abréviation de Binary Coded Decimal, ce qui veut dire décimal codé binaire), également appelé code 8-4-2-1, le code 2-4-2-1 et le code excédent 3, auquel on a, pour mémoire, ajouté un code redondant, du type n dont p.

ÉQUIVALENT DÉCIMAL	CODE 8-4-2-1	CODE 2-4-2-1	CODE EXCÉDENT 3	CODE 5 DONT 2
0	0000	0000	0011	11000
1	0001	0001	0100	00011
2	0010	0010	0101	00101

3	0011	0011	0110	00110
4	0100	0100	0111	01001
5	0101	0101	1000	01010
6	0110	0110	1001	01100
7	0111	0111	1010	10001
8	1000	1110	1011	10010
9	1001	1111	1100	10100

Tabl. 5. Différents codes décimaux.

Le **code 8-4-2-1** représente les digits décimaux entre 0 et 9 au moyen de leurs équivalents binaires directs.

Par exemple : 0101 se lit

$$0 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1 = 5,$$

les poids 8, 4, 2 et 1 étant représentés par les bits successifs.

Deux problèmes se posent : l'addition binaire directe de deux chiffres décimaux codés en 8-4-2-1 ne donne pas un chiffre décimal codé en 8-4-2-1 dès lors que le résultat dépasse 9, ce qui n'est même pas toujours détecté par une retenue ; la complémentation binaire directe d'un digit décimal ainsi codé fournit un complément à 15 éventuellement sous forme d'un code illégal en BCD, alors que l'on préférerait obtenir le complément à 9.

Le code 2-4-2-1 associe aux digits successifs les poids 2, 4, 2 et 1 de telle sorte que 1110 se lit :  $1 \times 2 + 1 \times 4 + 1 \times 2 + 0 \times 1 = 8$ . Il a l'avantage de permettre la complémentation à 9 par complémentation binaire directe puisque la somme des poids est égale à 9, encore que l'on n'obtienne pas toujours une représentation orthodoxe. Par exemple, le complément binaire de 0111 qui représente 7 s'écrit 1000, ce qui n'est pas le code normal de 2, mais qui - eu égard aux poids des différents bits - donne le résultat 2.

Le **code excédent 3**, contrairement aux deux précédents, n'est pas fondé sur les poids associés aux différents bits. Les digits décimaux sont représentés par leur équivalent binaire direct augmenté de 3. Ce code permet une complémentation à 9 directe et simplifie l'addition : on la traite en binaire et on retranche 3 au résultat obtenu s'il n'apparaît pas de retenue du cinquième ordre binaire, tandis qu'on lui ajoute 3 dans le cas contraire.

### Le format des nombres en machine.

Un opérateur arithmétique ne peut traiter que des nombres qui lui sont présentés selon un format ou un nombre restreint de formats bien définis.

La première composante d'un format est sa taille. Les machines mots, qui travaillent généralement en système binaire, admettent des formats de **longueur fixe**; suivant que l'information occupe un mot, un demi-mot, deux mots, etc., et on parlera de simple longueur, demi-longueur, double longueur, etc. Les machines caractères qui opèrent en système décimal travaillent généralement sur des formats de **longueur variable**, un nombre étant codé sur un nombre variable de caractères.

La deuxième composante d'un format est la convention dans laquelle le nombre est représenté ou codé. Nous nous proposons d'analyser trois types de formats qui sont : d'une part le format

fixe ou format en virgule fixe et le format flottant ou format en virgule flottante pour les nombres binaires ; d'autre part les formats de longueur variable pour les nombres décimaux.

## La virgule fixe

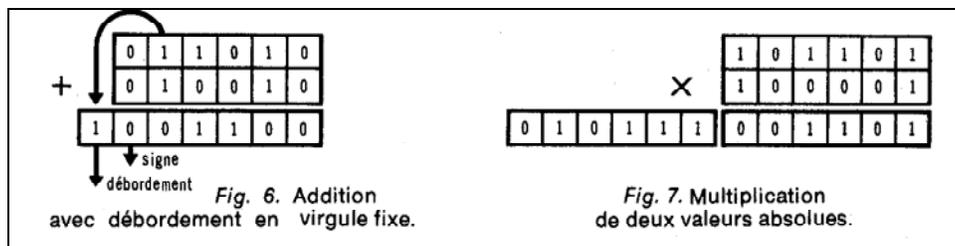
C'est la façon la plus naturelle d'écrire un nombre binaire dans un mot mémoire : les nombres sont considérés comme des entiers, le soin de placer la virgule étant laissé au programmeur. Les nombres négatifs peuvent être représentés dans l'une des trois conventions :  
 signe et valeur absolue, complément à 1 ou complément à 2.

Le premier bit du mot apparaît ainsi comme un bit de signe.

Un mot de  $n$  bits peut ainsi contenir les entiers allant de  $-2^{n-1}$  à  $2^{n-1}$

(auxquels il faut ajouter l'entier  $2^{n-1}$  dans la représentation en complément à 2).

On notera quelques particularités dans les opérations en virgule fixe. L'addition de deux nombres fixes peut donner lieu à ce que l'on appelle un débordement ou un dépassement de capacité (cf. fig. 6).



La multiplication de deux nombres de  $n$  chiffres donne un résultat de  $2n$  chiffres au plus qui sera donc représenté en virgule fixe double longueur. La figure 7 en donne un exemple sans tenir compte des signes pour simplifier.

## La virgule flottante

Lorsqu'il ne dispose que de la représentation en virgule fixe, le programmeur doit connaître et faire évoluer au cours des opérations la place de la virgule. S'il veut calculer avec le maximum de précision, il doit donc apprécier les ordres de grandeur de tous les résultats intermédiaires, afin de les cadrer de façon à garder toujours le maximum de digits significatifs.

La virgule flottante évite ces inconvénients. Elle consiste à représenter les nombres sous la forme :

$$S M \times \alpha^E$$

Où

S : est le signe du nombre

M : est la mantisse du nombre

E : est l'exposant du nombre  
 $\alpha$  : est généralement choisi égal à la base du système de numération ;

dans ce cas  $\alpha = 2$  en virgule flottante binaire.

Nous avons l'habitude de cette notation en écriture courante, où nous représentons

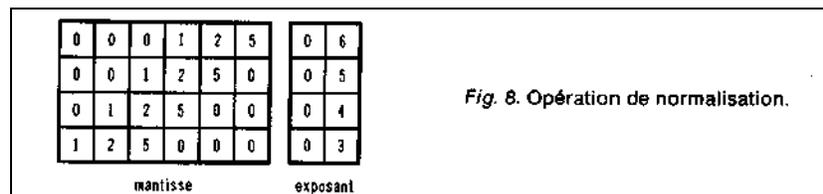
125 000 000 par  $125 \times 10^6$

Cette représentation n'est pas unique, puisque l'on peut aussi bien écrire :

$125 \times 10^6$  que  $12,5 \times 10^7$  ou  $1250 \times 10^5$ , etc. Parmi ces représentations, on en retient une,

dite normalisée, qui permet de conserver le plus de chiffres significatifs possible.

Pour normaliser un nombre, on décale sa mantisse vers la gauche jusqu'à ce que le premier digit soit significatif, et on diminue l'exposant du nombre de décalages qui ont été nécessaires (dans l'hypothèse où l'exposant exprime une puissance de deux) [cf. fig. 8].



En raison de la normalisation qui suit toute opération, la mantisse est toujours recadrée, de telle sorte que les dépassements de capacité ne puissent porter que sur l'exposant. Outre le dépassement de capacité pour exposant trop grand, il peut y avoir un sous-dépassement de capacité, lorsque l'exposant devient trop petit pour pouvoir être représenté.

La plupart des machines remplacent alors le résultat correspondant par zéro, ce qui fait exception aux règles de la normalisation.

On note que le nombre de digits de la mantisse est directement lié à la précision des calculs alors que le nombre de digits de l'exposant détermine les nombres extrêmes que la machine peut représenter.

Il y a plusieurs conventions de représentation des nombres flottants binaires en machines. De façon générale, on place les informations les plus significatives en tête : d'abord le signe, ensuite l'exposant, ensuite la mantisse.



On peut ainsi comparer deux nombres flottants avec les mêmes circuits que des nombres fixes, sous réserve que la convention adoptée<sup>o</sup> pour la représentation des exposants négatifs les fasse apparaître directement comme plus petits que les exposants positifs.

**Le signe.** Le plus (+) est généralement représenté par 0, le moins (-) par 1.

**L'exposant.** Il exprime généralement une puissance de 2 (quelquefois une puissance de 16 comme dans la série IBM 360).

S'il occupe  $e$  bits, il est généralement représenté par excès de  $2^{e-1}$  bits, c'est-à-dire que les exposants négatifs variant de  $-2^{e-1}$  à 0 sont représentés par les valeurs binaires allant de 0 à  $2^{e-1}$  tandis que les exposants positifs de 0 à  $2^{e-1} - 1$  sont représentés par les valeurs binaires allant de  $2^{e-1}$  à  $2^e - 1$ .

La **mantisse**. Elle peut être considérée comme entière, la virgule étant supposée immédiatement à droite, ou comme fractionnaire, la virgule étant supposée immédiatement à gauche. Ainsi  $125 \times 10^6$  s'écrira comme indiqué sur la fig. 9a dans la première convention ( $125\ 000 \times 10^3$ ) et comme indiqué sur la fig. 9b dans la deuxième convention ( $0.125 \times 10^9$ ).

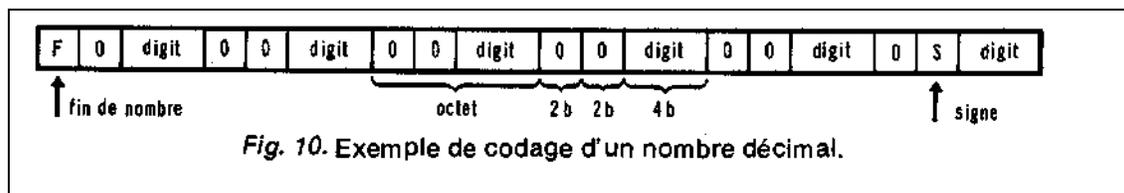


On trouve les trois représentations possibles pour les nombres négatifs : valeur absolue, complémentation à 1 ou à 2. Dans ces deux derniers cas, le signe est considéré comme faisant partie intégrante de la mantisse.

Il est important de savoir reconnaître si un nombre est normalisé. Par exemple, si l'exposant exprime une puissance de 2, et si les nombres négatifs sont représentés en complémentation, le nombre est normalisé lorsque le bit de signe et le bit de poids fort de la mantisse ont des valeurs différentes.

### Les chaînes décimales de longueur variable

Les digits décimaux sont généralement codés sur les quatre bits de poids faible des caractères de 6 bits ou des octets de 8 bits. On trouve également des représentations condensées avec deux digits par octet. Les nombres sont formés par une chaîne de tels caractères. Le signe d'un nombre peut être codé sur un caractère et placé en tête ou plus généralement en queue de la chaîne de caractères qui le constitue. Il peut également être inclus dans les bits de poids fort du premier ou du dernier caractère significatif.



Ces chaînes de caractères sont généralement de longueurs variables, et adressées par leur dernier caractère qui sera traité en premier à cause des retenues (cf. VI, 5.1.2.). Le premier caractère peut contenir, par exemple dans ses poids forts, une marque de fin de nombre ; en l'absence de marque de fin de nombre, l'instruction doit préciser au calculateur la longueur de la chaîne sur laquelle elle porte.

### Codage des informations non numériques.

On distingue deux types d'informations non numériques susceptibles d'être traitées par une machine

(1) des informations traitées par la machine au même titre que des nombres. On les représentera sous forme de **chaîne de caractères** (ces derniers comprenant les lettres, les chiffres et les signes de ponctuation ou d'opération). C'est sous forme de chaînes de caractères que se présentent aussi bien le fichier des comptes clients dans une banque que la suite des instructions d'un programme écrit en langage assembleur ou en langage évolué ;

(2) des informations reconnues par la machine comme des commandes de ses différents organes. Ce sont les **instructions** une fois transcrites en langage machine.

### **Codage des caractères.**

Nous avons vu au paragraphe 2.6. comment on pouvait coder les caractères numériques. Il s'agit maintenant d'étendre ce système de codage à l'ensemble des signes de la machine à écrire (alphabet, signes de ponctuation et d'opérations, caractères spéciaux : \$, £, etc.). On impose généralement à ces représentations certaines conditions parmi les suivantes

(1) la représentation doit englober la représentation des chiffres décimaux sous une forme simple à utiliser (8-4-2-1, 2-4-2-1, excédent 3 ...) et permettre de reconnaître rapidement les chiffres des autres caractères ;

(2) la représentation doit permettre l'adjonction de nouveaux caractères spécifiques à une application particulière ;

(3) dans le cas des transmissions, la représentation doit inclure un système de redondance permettant la détection des erreurs.

De nombreux codes sont utilisés pour la représentation interne des caractères en machine (sur 6, 7, 8 bits), pour le dialogue avec les unités périphériques (7 et 9 bits dont un de parité pour les bandes magnétiques, 12 bits pour les cartes perforées, etc.) ainsi que pour les télétransmissions. Les codes utilisés sont spécifiés dans les manuels d'utilisation des machines et des organes périphériques.

## **Le codage des informations**

Le morse a été le premier codage à permettre une communication longue distance. C'est Samuel F.B.Morse qui l'a mis au point en 1844. Ce code est composé de points et de tirets (un codage binaire en quelque sorte...). Il permet d'effectuer des communications beaucoup plus rapides que ne le permettait le système de courrier de l'époque aux Etats-Unis : le Pony Express.

L'interpréteur était l'homme à l'époque, il fallait donc une bonne connaissance du code...

De nombreux codes furent inventés dont le code d'Émile Baudot (portant d'ailleurs le nom de code Baudot, les anglais l'appelaient en revanche Murray Code).

Le 10 mars 1876, le Dr Graham Bell met au point le téléphone, une invention révolutionnaire qui permet de faire circuler de l'information vocale dans des lignes métalliques. Pour l'anecdote, la Chambre des représentants a décidé que l'invention du téléphone revenait à Antonio Meucci. Ce dernier avait en effet déposé une demande de brevet en 1871, mais n'avait pas pu financer celle-ci au-delà de 1874.

Ces lignes permirent l'essor des télécriteurs, des machines permettant de coder et décoder des caractères grâce au code Baudot (les caractères étaient alors codés sur 5 bits, il y avait donc 32 caractères uniquement...).

Dans les années 60, le code ASCII (American Standard Code for Information Interchange) est adopté comme standard. Il permet le codage de caractères sur 8 bits, soit 256 caractères possibles.

## Qu'est-ce que le code ASCII ?

La mémoire de l'ordinateur conserve toutes les données sous forme [numérique](#). Il n'existe pas de méthode pour stocker directement les caractères. Chaque caractère possède donc son équivalent en code numérique : c'est le code ASCII (American Standard Code for Information Interchange - traduisez « Code Américain Standard pour l'Echange d'Informations »).

Le code ASCII de base représentait les caractères sur 7 bits (c'est-à-dire 128 caractères possibles, de 0 à 127).

Les codes 0 à 31 ne sont pas des caractères. On les appelle caractères de contrôle car ils permettent de faire des actions telles que :

retour à la ligne (CR)

Bip sonore (BEL)

Les codes 65 à 90 représentent les majuscules

Les codes 97 à 122 représentent les minuscules

(Il suffit de modifier le 6ème bit pour passer de majuscules à minuscules, c'est-à-dire ajouter 32 au code ASCII en base décimale.)

caractère	code ASCII	code hexadécimal
NUL (Null)	0	00
SOH (Start of heading)	1	01
STX (Start of text)	2	02
ETX (End of text)	3	03
EOT (End of transmission)	4	04
ENQ (Enquiry)	5	05
ACK (Acknowledge)	6	06
BEL (Bell)	7	07
BS (Backspace)	8	08
TAB (Tabulation horizontale)	9	09
LF (Line Feed, saut de ligne)	10	0A
VT (Vertical tabulation, tabulation verticale)	11	0B
FF (Form feed)	12	0C
CR (Carriage return, retour à la ligne)	13	0D
SO (Shift out)	14	0E
SI (Shift in)	15	0F
DLE (Data link escape)	16	10
DC1 (Device control 1)	17	11
DC2 (Device control 2)	18	12
DC3 (Device control 3)	19	13
DC4 (Device control 4)	20	14
NAK (Negative acknowledgement)	21	15
SYN (Synchronous idle)	22	16
ETB (End of transmission block, fin de bloc de transmission)	23	17
CAN (Cancel, annulation)	24	18
EM (End of medium, fin du médium)	25	19
SUB (Substitute, substitut)	26	1A

ESC (Escape, caractère d'échappement)	27	1B
FS (File separator, séparateur de fichier)	28	1C
GS (Group separator, séparateur de groupe)	29	1D
RS (Record separator, séparateur d'enregistrement)	30	1E
US (Unit separator, séparateur d'enregistrement)	31	1F
SP (Space, espace)	32	20
!	33	21
"	34	22
#	35	23
\$	36	24
%	37	25
&	38	26
'	39	27
(	40	28
)	41	29
*	42	2A
+	43	2B
,	44	2C
-	45	2D
.	46	2E
/	47	2F
0	48	30
1	49	31
2	50	32
3	51	33
4	52	34
5	53	35
6	54	36
7	55	37
8	56	38
9	57	39
:	58	3A
;	59	3B
<	60	3C
=	61	3D
>	62	3E
?	63	3F
@	64	40
A	65	41
B	66	42
C	67	43
D	68	44
E	69	45
F	70	46
G	71	47
H	72	48

I	73	49
J	74	4A
K	75	4B
L	76	4C
M	77	4D
N	78	4E
O	79	4F
P	80	50
Q	81	51
R	82	52
S	83	53
T	84	54
U	85	55
V	86	56
W	87	57
X	88	58
Y	89	59
Z	90	5A
[	91	5B
\	92	5C
]	93	5D
^	94	5E
_	95	5F
`	96	60
a	97	61
b	98	62
c	99	63
d	100	64
e	101	65
f	102	66
g	103	67
h	104	68
i	105	69
j	106	6A
k	107	6B
l	108	6C
m	109	6D
n	110	6E
o	111	6F
p	112	70
q	113	71
r	114	72
s	115	73
t	116	74
u	117	75
v	118	76
w	119	77
x	120	78
y	121	79

z	122	7A
{	123	7B
	124	7C
}	125	7D
~	126	7E
Touche de suppression	127	7F

### Table des caractères ASCII Etendue

Le code ASCII a été mis au point pour la langue anglaise, il ne contient donc pas de caractères accentués, ni de caractères spécifiques à une langue. Pour coder ce type de caractère il faut recourir à un autre code. Le code ASCII a donc été étendu à 8 bits (un octet) pour pouvoir coder plus de caractères (on parle d'ailleurs de code ASCII étendu...).

Ce code attribue les valeurs 0 à 255 ([donc codées sur 8 bits, soit 1 octet](#)) aux lettres majuscules et minuscules, aux chiffres, aux marques de ponctuation et aux autres symboles (caractères accentués dans le cas du code iso-latin1).

Rem. : Le code ASCII étendu n'est pas unique et dépend fortement de la plateforme utilisée. Les deux jeux de caractères ASCII étendus les plus couramment utilisés sont :

Le code ASCII étendu OEM, c'est-à-dire celui qui équipait les premières machines de type IBM PC

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	ç	ü	é	â	ä	à	å	ç	ê	ë	è	ï	î	ì	ñ	
9	é	æ		ô	ö	ò	û	ù	ÿ	ÿ	ÿ		£	¥		f
A	á	í	ó	ú	ñ	ñ										
B																
C																
D																
E																
F																

Le code ASCII étendu ANSI, utilisé par les systèmes d'exploitation récents

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	□	□	,	f	//	...	†	#	^	⌘	Š	<	œ	□	□	□
9	□	\	'	“	”	•	-	-	~	™	š	>	œ	□	□	ÿ
A		ı	◊	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
D	ð	ñ	ò	ó	ô	õ	ö	×	ø	ù	ú	û	ü	ý	þ	ÿ
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ø	ñ	ò	ó	ô	õ	÷	ø	ù	ú	û	ü	ý	þ	ÿ	

## Le code EBCDIC

Le code EBCDIC (Extended Binary-Coded Decimal Interchange Code), développé par IBM, permet de coder des caractères sur 8 bits. Bien que largement répandu sur les machines IBM, il n'a pas eu le succès qu'a connu le code ASCII.

## Unicode

Le code Unicode est un système de codage des caractères sur 16 bits mis au point en 1991. Le système Unicode permet de représenter n'importe quel caractère par un code sur 16 bits, indépendamment de tout système d'exploitation ou langage de programmation. Il regroupe ainsi la quasi-totalité des alphabets existants (arabe, arménien, cyrillique, grec, hébreu, latin, ...) et est compatible avec le code ASCII. L'ensemble des codes Unicode est disponible sur le site <http://www.unicode.org>.

### Codage des instructions.

Les instructions comportent un certain nombre d'informations, notamment le code opération indiquant l'opération à effectuer, des adresses de registre ou de cellule mémoire, des conditions d'adressage spécifiant les calculs éventuels à faire subir aux adresses de cellule mémoire. Suivant le nombre et l'importance de ces informations, les instructions seront codées sur un ou plusieurs mots machines.

A chacun des éléments d'information composant l'instruction on associe généralement une zone formée d'un nombre de digits suffisant pour coder les différents états possibles de cette information. C'est ainsi qu'une zone de 6 bits allouée au code instruction permettra de coder jusqu'à 64 opérations différentes, qu'une zone de 4 bits allouée à une adresse de registre permettra de désigner un registre parmi 16 et qu'une zone de 16 bits permettra d'adresser une mémoire de 65 536 mots. Nous reviendrons au chapitre VI sur les problèmes posés par le codage des instructions, notamment en ce qui concerne les conditions d'adressage et les adresses. En ce qui concerne le codage des opérations, on pourrait associer arbitrairement les 2<sup>n</sup> configurations binaires du code opération de n bits aux diverses instructions.

On préfère souvent, afin de simplifier le décodage, diviser la zone code opération en sous-zones ou champs qui seront décodées indépendamment. Cette décomposition en champs peut recouvrir deux optiques différentes :

(1) chaque champ peut être associé à un organe de la machine, le décodage de l'information contenue dans le champ indiquant la fonction que doit remplir cet organe. Cette solution est coûteuse en nombre de bits, dans la mesure où tous les organes de la machine ne sont pas mis en jeu par toutes les instructions ;

(2) les instructions peuvent être divisées en classes d'instructions, un premier champ étant réservé en codage de la classe d'instructions. La signification à attribuer aux autres champs dépend de la classe de l'instruction. Ainsi le même champ peut désigner le type d'opération arithmétique à exécuter si l'instruction appartient à la classe des instructions arithmétiques, tandis qu'il désignera la condition de branchement si l'instruction appartient à la classe des branchements. Le champ réservé au code opération peut être de longueur variable en fonction de la classe d'instructions.

On appelle souvent indice un bit (ou un ensemble de bits) qui détermine la signification globale à attribuer à un champ. Ainsi par exemple, un même champ devra être reconnu comme contenant une adresse de registre si l'indice associé vaut 0, et comme contenant un complément au code opération si l'indice associé vaut 1.

## L'organisation et la recherche des informations en mémoire

### Notion de tableau, de liste et de pointeur

L'organisation la plus classique des informations en mémoire consiste à les ranger à la suite les unes des autres.

C'est très généralement le cas des instructions successives d'un programme, des tableaux de nombres en calcul scientifique, etc.

Il faut alors, pour rechercher les informations successives, disposer d'un pointeur qui prend successivement pour valeur les adresses des informations rangées en séquence (fig. 11).

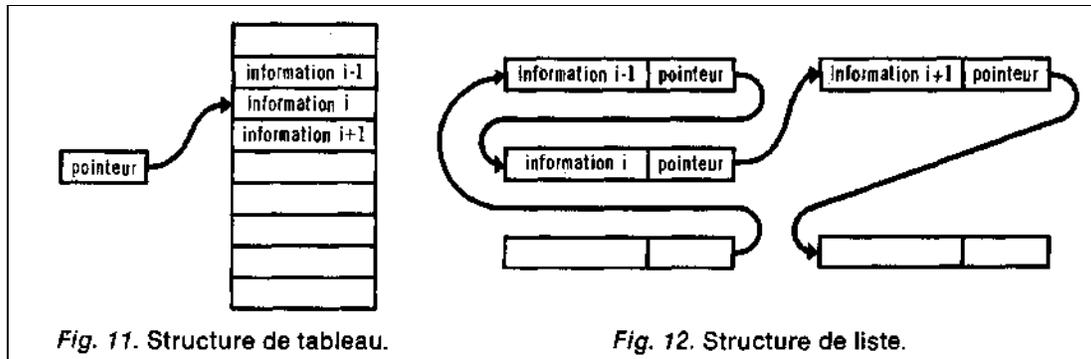
C'est évidemment le rôle du compteur ordinal en ce qui concerne les instructions. Un ensemble d'informations rangées en séquence est généralement appelé tableau.

Cette méthode de rangement, qui convient très bien dans le domaine des calculs scientifiques (rangement de tableaux matriciels par exemple), n'est pas bien adaptée à des manipulations de listes d'informations, telles que l'extraction ou l'insertion d'informations, la combinaison de listes d'informations, etc., car elles exigent des déplacements d'informations en mémoire.

On emploie alors une structure de liste (fig. 12) où chaque information est complétée par un pointeur qui contient l'adresse de l'information suivante.

Les opérations d'insertion, extraction, combinaison se font par simple manipulation de pointeurs.

On peut concevoir des structures de machines spécifiquement orientées vers le traitement des listes.



### Notion de table.

Une table établit une correspondance entre deux types d'information l'information significative que l'on recherche, et l'information d'entrée qui permet de la trouver. Cette dernière est également appelée adresse associative ou étiquette.

La mise en mémoire d'une table peut se réduire à la constitution d'un simple tableau si la position de l'information significative dans le tableau peut être calculée à partir de l'information d'entrée. C'est le cas, par exemple, de la table des logarithmes des  $n$  premiers entiers positifs. Le logarithme de  $i$  s'obtiendra directement en chargeant  $a + i$  dans le pointeur de tableau,  $a$  étant l'adresse d'implantation en mémoire du premier élément du tableau.

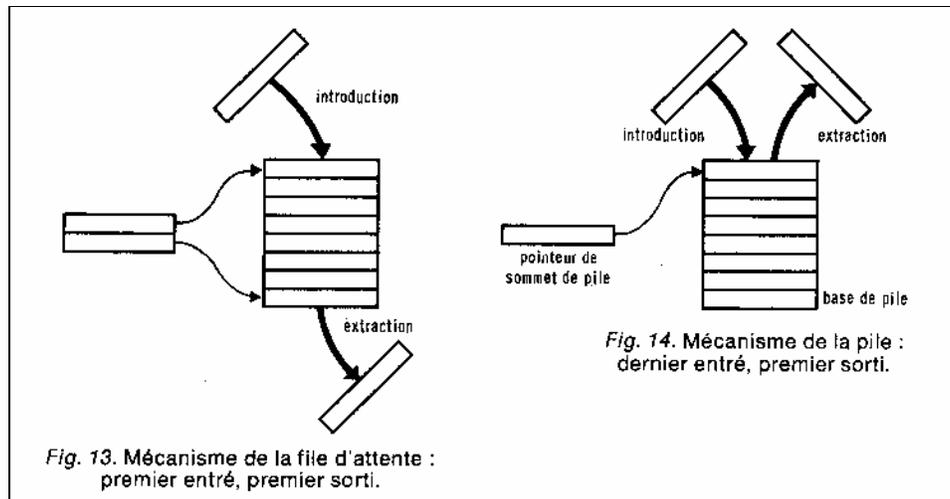
Dans le cas général, la table sera rangée sous forme d'un tableau dont chaque élément contiendra l'information d'entrée et l'information significative correspondante. Ce sera par exemple la table qui fait correspondre au code opération mnémotique, utilisé en langage assembleur, la représentation binaire machine du code opération.

La recherche en table nécessite les comparaisons successives de l'information d'entrée à toutes les informations d'entrée possibles jusqu'à ce qu'il y ait coïncidence. Cette recherche peut être accélérée soit à l'aide d'algorithmes de programmation, notamment lorsque des informations d'entrée sont classées par valeurs binaires croissantes, soit par l'utilisation de mémoires spéciales appelées mémoires associatives ou mémoires adressables par le contenu. Nous décrivons ces dernières au chapitre IV.

### Notion de pile et de file d'attente

Ces deux notions se réfèrent à des organisations particulières des données en mémoire dans lesquelles l'ordre d'utilisation des informations dépend de l'ordre dans lequel elles ont été introduites. La file d'attente fonctionne selon un principe analogue à celui que nous connaissons chez le boulanger ou le boucher : on a accès à l'information qu'a séjourné le plus longtemps selon le principe « premier arrivé, premier servi ». Le mécanisme de la pile rappelle la pile de chemises dans le placard : on a accès à la dernière information rangée selon le principe « dernier entré, premier sorti ».

Deux pointeurs sont nécessaires pour définir l'état d'une file d'attente, l'un pointant sur le dernier mot entré, l'autre sur le prochain mot à sortir, alors qu'un seul pointeur suffit pour définir l'état d'une pile, puisqu'il suffit de connaître l'emplacement du sommet de pile. (Fig. 13 et 14.)



Les piles et files d'attente peuvent être constituées en mémoire centrale par programmation ; leur gestion peut être simplifiée par des instructions spéciales d'insertion et d'extraction et éventuellement par l'utilisation de pointeurs câblés. Enfin il existe des petites mémoires spécialement câblées en pile ou en file d'attente.

### Notion de préfixe

Dans la plupart des machines, les informations sont reconnues comme instructions ou données par le contexte du programme. Ceci veut dire, par exemple, que la machine suppose que le contenu de la cellule mémoire adressée par une instruction d'arithmétique flottante est un nombre flottant et, en tout état de cause, l'interprète comme tel.

Dans certaines machines les informations sont dites préfixées. Le préfixe désigne sur un petit nombre de bits le type d'information contenu dans le mot machine (nombre fixe simple ou double précision, nombre flottant, instruction, pointeur, etc.). Les circuits de la machine sont capables d'analyser le préfixe lorsque le mot machine est utilisé, soit que le contexte ne permette pas de définir suffisamment le contenu de l'information, soit à titre de contrôle. Cette technique permet notamment d'adresser des structures de données complexes, l'adresse pointant non pas directement sur les données, mais sur un descripteur de cette structure qui contient notamment un pointeur vers la première donnée de la structure. Ce descripteur est reconnu comme tel par la machine grâce à son préfixe. Nous en verrons des exemples au chapitre IX.

### Eléments de langage machine.

Nous nous proposons ici de tenter une classification des machines de type Von Neumann en fonction des éléments d'information constitutifs de l'instruction et de décrire brièvement les principaux types d'instructions que l'on rencontre dans ces machines.

### Forme générale des instructions arithmétiques.

L'instruction arithmétique type doit fournir à l'unité de contrôle chargée de la décoder et de commander son exécution un certain nombre de renseignements dont les principaux sont

(1) la nature de l'opération à réaliser. C'est le rôle du code opération. Exemple : opération d'addition en virgule fixe simple précision;

(2) les données, ou opérandes, sur lesquelles porte l'instruction. Généralement l'instruction spécifie les adresses des opérandes. Exemple : l'opération d'addition en virgule fixe simple précision porte sur l'opérande qui est à l'adresse AD<sub>1</sub>, et l'opérande qui est à l'adresse AD<sub>2</sub>. Si l'instruction porte sur plus de deux opérandes, on s'arrange pour les ranger en séquence de telle sorte qu'il suffit de donner l'adresse du premier et, soit l'adresse du dernier, soit la longueur de la séquence ;

(3) l'emplacement où doit être rangé le résultat. Exemple : le résultat de l'addition en virgule fixe simple précision des deux opérandes d'adresse AD<sub>1</sub> et AD<sub>2</sub> doit être rangé à l'adresse AD<sub>3</sub>

(4) l'adresse de la prochaine instruction à exécuter. Exemple : après avoir exécuté l'addition, exécuter l'instruction d'adresse AD<sub>4</sub>.

CO	AD <sub>1</sub>	AD <sub>2</sub>	AD <sub>3</sub>	AD <sub>4</sub>
code opération	adresse du premier opérande	adresse du deuxième opérande	adresse de rangement du résultat	adresse de la prochaine instruction

Fig. 15. Instruction à quatre adresses.

Une machine dont l'instruction possède explicitement ces différents éléments est appelée **machine à quatre adresses** (ou encore à 3 + 1 adresses pour différencier des autres l'adresse de l'instruction suivante). Les machines à quatre adresses sont très rares, certaines adresses étant en général implicitement déterminées. La quatrième adresse qui pointe sur la prochaine instruction est évidemment implicite, en dehors des instructions de rupture de séquence, dans les machines où les instructions sont rangées séquentiellement en mémoire. Elle serait nécessaire dans une machine orientée vers le traitement des listes où le programme lui-même s'organiserait en liste, et a trouvé une justification technologique dans certaines machines de la première génération dont la mémoire centrale était constituée d'un tambour magnétique, par le fait qu'elle évitait d'avoir à attendre une rotation complète du tambour entre deux instructions.

On appelle **machine à trois adresses** les machines dont l'instruction spécifie les adresses des deux opérandes et l'adresse du résultat.

Dans les **machines à deux adresses**, beaucoup plus courantes, l'adresse du résultat est en général implicitement choisie égale à l'adresse du premier opérande.

Enfin, dans les **machines à une adresse**, ne figure plus que l'adresse du deuxième opérande, l'emplacement du premier opérande qui sera remplacé par le résultat en fin d'opération étant implicitement défini par le code opération. La plupart des calculateurs scientifiques de la deuxième génération fonctionnaient selon ce principe, toutes les opérations étant effectuées sur un même registre appelé **accumulateur**.

La classification des machines en machines à une, deux, trois, exceptionnellement quatre adresses devient caduque avec la troisième génération, d'autant que, compte tenu de la multiplication des registres dans les unités centrales, les opérandes peuvent se trouver aussi bien dans des registres qu'en mémoire. Voici quelques exemples : la CDC 6600 se comporte comme une machine à deux adresses pour les opérations de transferts entre mémoire centrale et registres (l'instruction désigne le registre et la cellule mémoire concernée), et comme une machine à trois adresses de registres pour toutes les instructions arithmétiques. L'IBM 360 admet essentiellement trois types d'instructions à deux adresses : registre à registre, registre à mémoire et mémoire à mémoire. Dans la CII 10070 ou le PDP 10 les registres sont

considérés comme les premières cellules de la mémoire, de telle sorte que le même champ de la même instruction peut adresser soit un registre, soit une cellule mémoire, etc.

### Le jeu d'instructions d'un calculateur.

Le jeu d'instructions peut varier d'une dizaine d'instructions sur les toutes petites machines utilisées en automatisme à une centaine d'instructions. Exceptionnellement la technique de codage par champ peut conduire à des jeux d'instructions plus importants.

En fait un jeu d'une dizaine d'instructions est suffisant pour réaliser un calculateur universel, c'est-à-dire un calculateur capable d'aborder n'importe quel type de problème dont la solution est du ressort du traitement digital et qui ne dépasse pas les capacités mémoire et les performances du calculateur. C'est ainsi que les instructions addition, complémentation, effacement, rangement et décalage sont suffisantes pour réaliser toutes les opérations arithmétiques : le chargement de l'accumulateur sera obtenu par effacement et addition ; la soustraction sera obtenue par addition du complément, la multiplication et la division par la programmation d'un algorithme utilisant des additions, soustractions et décalages.

C'est ici que nous voyons apparaître une certaine complémentarité entre les circuits et les programmes, entre le hardware et le software. Ce qui n'est pas réalisé par l'un doit être réalisé par l'autre. Il est clair que le câblage d'un grand nombre d'opérations sur un calculateur correspond à un accroissement notable de performances, contrebalancé évidemment par un accroissement correspondant du prix. Nous verrons au chapitre VII que les techniques de microprogrammation introduisent un concept intermédiaire entre le hardware et le software, le firmware qui permet d'augmenter notablement le nombre d'instructions d'un calculateur sans augmentation sensible du prix.

En ce qui concerne les opérations arithmétiques, le tableau de la figure 16 indique, à titre d'exemple, ce qui est normalement disponible sous forme câblée dans les différents types de calculateurs.

		calculateurs de bureau	calculateurs de gestion	petits calculateurs scientifiques	gros ordinateurs scientifiques
addition	fixe binaire				
soustraction					
multiplication	fixe binaire				
division					
arithmétique fixe décimale					ordinateurs mixtes scientifiques gestions
arithmétique flottante binaire					
arithmétique flottante décimale					
racine carrée décimale					
fonctions transcendantes classiques : log, sin, exp.		calculateurs de bureau très élaborés			
fonctions spéciales			*		

légende :  toujours câblé     éventuellement câblé     jamais câblé (ou réalisé par programme)

Fig. 16.

### Les différents types d'instructions.

Il n'est pas possible de passer en revue toutes les instructions machine qui apparaissent dans les manuels de référence des différents calculateurs. Nous nous contenterons de donner quelques indications très générales sur les classes d'instructions les plus répandues dans les calculateurs de type Von Neumann que nous illustrerons sur des exemples concernant un ordinateur à deux adresses, une adresse de registre pour le premier opérande et le résultat, une adresse en mémoire pour le deuxième opérande (ce ordinateur sera décrit au chapitre VI sous le nom de Superboulx).

Pour décrire l'effet des instructions, nous adopterons les conventions suivantes

- M désigne une adresse en mémoire ;
  - M + 1 l'adresse de la cellule suivante;
  - R désigne une adresse de registre ;
  - R + 1 l'adresse du registre suivant ;
  - (R) désigne le contenu du registre d'adresse R, tandis que
  - (R, R + 1) désigne le contenu en double longueur des registres R et R + 1 mis bout à bout;
  - (M) désigne le contenu de la cellule mémoire d'adresse M, et
  - (M, M + 1) le contenu des deux cellules mémoires successives d'adresses M et M + 1; enfin le signe
- placé entre deux membres indique que l'information définie dans le premier membre est rangée dans l'élément de mémoire (M ou R) défini par le second membre.

### Instructions de transfert

Il s'agit essentiellement de transferts de mots entre mémoire et registre ou entre registres. Citons notamment

chargement	CHA	R	M	(M) → R
rangement	RNG	R	M	(R) → M
transfert	TRA	R <sub>i</sub>	R <sub>j</sub>	(R <sub>i</sub> ) → R <sub>j</sub>
échange	ECH	R <sub>i</sub>	R <sub>j</sub>	(R <sub>i</sub> ) → R <sub>j</sub> et R <sub>j</sub> → (R <sub>i</sub> )

Il peut exister des instructions de transfert avec complémentarité, par exemple :

$$\text{CHAC} \quad R \quad M \quad (\overline{M}) \rightarrow R.$$

### Instructions arithmétiques en virgule fixe binaire.

Il s'agit des quatre opérations. On les trouve souvent en simple et double longueur exceptionnellement en demi, triple ou quadruple longueur. Donnons quelques exemples

simple longueur

addition	ADD R M	$(R) + (M) \rightarrow R$
soustraction	SUB R M	$(R) - (M) \rightarrow R$
multiplication	MUL R M	$(R) \times (M) \rightarrow R, R+1$
division	DIV R M	$(R, R+1) / (M) \rightarrow R$

double longueur

ADD2 R M	$(R, R+1) + (M, M+1) \rightarrow R, R+1$
SUB2 R M	$(R, R+1) - (M, M+1) \rightarrow R, R+1$

### Instructions arithmétiques en virgule flottante binaire

Il peut exister plusieurs formats flottants sur un, deux ou trois mots calculateurs. Dans ce cas on aura pour chaque opération un code par format. En simple longueur, nous aurons par exemple

ADDF 1 R M	$(R) + (M) \rightarrow R$
NORM 1 R	normalise le contenu du registre R.

### Instructions logiques

La complémentation d'un opérande consiste à remplacer les 1 par des 0 et les 0 par des 1. Le produit logique de deux opérandes comporte des bits 1 uniquement aux emplacements où les deux bits correspondants des opérandes sont à 1. La somme logique de deux opérandes comporte des bits zéro uniquement aux emplacements où les deux bits correspondants des opérandes sont à zéro.

Nous les notons ainsi

complémentation	CMP R ( $\bar{R}$ )	$\rightarrow R$
produit logique	ET R M	$(R) \text{ et } (M) \rightarrow R$
somme logique	OU R M	$(R) \text{ ou } (M) \rightarrow R$

### Instructions de décalage

Les décalages peuvent avoir lieu sur un mot ou sur deux mots. Ils peuvent être ouverts et, dans ce cas, on remplit généralement de zéros les places laissées libres, ou fermées - on dit encore circulaires -, les bits perdus d'un côté réapparaissant de l'autre.

Les décalages arithmétiques permettent de décaler la valeur d'un opérande fixe binaire sans toucher au signe. En cas de décalage à droite, le signe est répété dans les portions laissées libres si les nombres négatifs sont représentés en complément.

Exemple d'instruction de décalage : l'instruction DECD R n décale le contenu du registre R de n positions à droite.

### Instructions arithmétiques décimales fixes

Elles portent sur des chaînes de chiffres décimaux codés sur des caractères de 4, 6 ou 8 bits qui sont de longueur variable. Certaines machines admettent un accumulateur décimal, ou le simulent en mettant bout à bout plusieurs registres ; l'instruction est alors de la forme DADD R M où R désigne un registre qui contient le nombre de caractères (ou de mots) du deuxième opérande et M l'adresse du dernier caractère (ou mot) de cet opérande. D'autres machines opèrent en série caractère par caractère en utilisant des marques de début ou fin de nombre ; l'instruction prend alors la forme DADD M, M2 où M, et M2 sont les adresses de fin des opérandes (on choisit de préférence les adresses de fin des opérandes puisque l'on doit exécuter les opérations à partir des poids faibles).

### **Instructions décimales flottantes.**

Ces instructions travaillent avec accumulateurs sur des formats flottants de longueur fixe. On ne les trouve que sur de très rares machines initialement prévues pour la gestion puis modifiées pour le calcul scientifique (comme par exemple le Gamma 30, modèle scientifique).

### **Instructions de conversion**

Ces instructions permettent de convertir les nombres d'un format à un autre format, par exemple un nombre fixe binaire en décimal.

### **Instructions de mouvement de chaînes de caractères**

Ces instructions permettent de transférer des chaînes de caractères de longueur variable d'un emplacement de mémoire à un autre emplacement, de comparer deux chaînes de caractères, etc. Elles s'exécutent normalement en série caractère par caractère. L'instruction doit adresser les débuts (ou les fins) des deux chaînes opérandes et éventuellement contenir soit des indications de longueur de chaînes soit une indication de caractère d'arrêt, à moins que les chaînes de caractères ne se terminent par des marques de fin de chaîne.

### **Instructions de branchement**

Ces instructions permettent d'exécuter des ruptures de séquence soit inconditionnelles, soit dépendant de conditions portant sur un certain nombre d'indicateurs caractérisant le résultat de la dernière opération effectuée, par exemple : branchement si accumulateur positif, négatif, nul, branchement si débordement, branchement si identité entre deux chaînes de caractères, etc. Des branchements spéciaux permettent d'appeler des sous-programmes ou de terminer des boucles de programme.

L'instruction type de branchement s'écrit : BR COND M, où le champ COND représente la condition pour qu'il y ait branchement et M l'adresse de la prochaine instruction à exécuter si la condition est réalisée; si elle ne l'est pas, le programme se continue en séquence.

### **Instructions portant sur des structures d'informations élaborées**

Certaines machines possèdent des instructions capables de gérer des listes (insertion ou extraction d'éléments, concaténation de listes) ou des piles (insertion et extraction d'éléments). Ces instructions remettent à jour les différents pointeurs.

### **Instructions de commande et d'état**

Ces instructions permettent soit de commander la mise d'un organe de la machine dans un certain état, soit de tester l'état dans lequel se trouve cet organe.  
 Dans cette classe entrent notamment les instructions commandant les organes d'entrée-sortie, le système d'interruption, le système de protection mémoire, etc. Elles seront étudiées dans les chapitres où sont décrits les organes qu'elles commandent.

### Notions élémentaires sur la structure des programmes.

Ce paragraphe n'est pas une introduction à la programmation ; il se contente d'exposer les éléments nécessaires à la compréhension des chapitres suivants (VI, IX et X notamment), en particulier les notions de récurrence, sous-programme, réentrance et récursivité. Il permettra au passage de montrer des exemples d'application du calcul d'adresse que nous exposerons au chapitre VI.

#### Calcul d'une expression arithmétique.

Soit l'expression : 
$$a = \frac{b + 3d}{c + e}$$

Nous allons la programmer successivement dans le langage machine de Boulix, en langage assembleur pour ce même calculateur, enfin en langage évolué (type Fortran).

### Programmation en langage machine.

Bien que la machine travaille au binaire, nous écrirons notre élément de programme dans la représentation octale pour en faciliter la lecture.

Nous supposons

- (1) que l'instruction est codée sur 18 bits : 6 bits, soit 2 chiffres octaux, pour le code opération et 12 bits, soit 4 chiffres octaux, pour l'adresse.
  - (2) que les valeurs binaires de b, c, d, e se trouvent dans les mémoires d'adresse octale 40, 41, 42, 43; que le résultat a doit être rangé à l'adresse 100 octal ; qu'à l'adresse 77 octal est rangée la valeur 3 en binaire ; que les mémoires d'adresse 101, 102, ... sont libres.
  - (3) que les codes opérations sont codés selon le tableau suivant
- |                              |     |    |
|------------------------------|-----|----|
| chargement de l'accumulateur | CHA | 00 |
| rangement de l'accumulateur  | RNG | 01 |
| addition                     | ADD | 02 |
| soustraction                 | SUB | 03 |
| multiplication               | MUL | 04 |
| division                     | DIV | 05 |
- (4) que l'élément de programme commence à l'adresse 15 (en octal).

Le programme s'écrit

EMPLACEMENT	INSTRUCTIONS	COMMENTAIRES
N	N	RES

	C.	ADRESSE		
	O.			instruction précédente
...	...	...		
0015	00	0043		charge e dans l'accumulateur
0016	02	0041		additionne c à
0017	01	0101		range c + e dans mémoire
0020	00	0042		charge d dans l'accumulateur
0021	04	0077		multiplie par 3
0022	02	0040		additionne b
0023	05	0101		divise par c + e
0024	01	0100		range le résultat a
0025	...	...		instruction suivante
...	...	...		...
0040				valeur` de b
0041				valeur de c
0042				valeur de d
0043				valeur de e
0077	00	0003		valeur 3
0100				valeur de a
0101				mémoire de travail
...	..	..		...

### Programmation en langage assembleur

En langage assembleur (encore appelé langage symbolique), le programmeur écrit les mêmes, instructions qu'en langage machine mais en utilisant des codes mnémoniques pour les codes opérations et en ne se préoccupant plus des emplacements en mémoire, les adresses des opérands ou des instructions étant notées sous forme symbolique. Voici le même élément de programme écrit en langage assembleur

ADRESSE SYMBOLIQUE	INSTRUCTION	ADRESSE SYMBOLIQUE	INSTRUCTION
	CHA E	B	RES 1
	ADD C	C	RES 1
	RNG TRA 1	D	RES 1
	CHA D	E	RES 1
	MUL TROIS	...	
	ADD B B	TROIS	DEC 3
	DIV TRA 1	A	RES 1
	RNG A	TRA 1	RES 1

**L'assembleur** est un programme chargé de traduire en langage machine les instructions écrites en langage symbolique. Des instructions telles que RES 1, qui veut dire réserver une position mémoire pour la donnée dont l'adresse symbolique est indiquée en zone adresse, ou encore DEC 3, qui veut dire attribuer la valeur 3 à la constante dont l'adresse symbolique est indiquée en zone adresse, ne sont pas des instructions exécutables par la machine, mais des

directives pour l'assembleur, appelées pseudo-instructions. L'assembleur fournit, à partir du programme source écrit en langage symbolique, un programme objet transcrit instruction par instruction en langage machine. Pour pouvoir être exécuté, ce programme objet doit être rangé en mémoire à l'aide d'un chargeur. On distingue les assembleurs non translatables qui transcrivent directement les adresses symboliques en adresses absolues, et les assembleurs translatables qui permettent le chargement du programme à partir de n'importe quelle adresse de la mémoire. Dans cette dernière hypothèse, toutes les adresses symboliques sont transcrites par l'assembleur comme si le programme devait être implanté à partir de l'adresse zéro et c'est le chargeur qui se charge d'exécuter les translations nécessaires.

### Programmation en langage évolué

Dans la plupart des langages évolués, orientés vers le calcul scientifique, l'expression s'écrit sous sa forme algébrique normale

$$A = (B + 3 * D) / (C + E)$$

Ici, A, B, C, D, E ne sont plus des adresses symboliques, mais des identificateurs permettant de repérer les variables a, b, c, d, e. Les variables peuvent être, suivant les langages, des nombres fixes, flottants, décimaux, des caractères, des bits, ou dans certains cas des structures plus complexes : tableaux, listes, pointeurs...

L'instruction écrite ci-dessus assigne à la variable identifiée par l'identificateur A la valeur obtenue après calcul de l'Expression du second membre portant sur les variables identifiées par B, C, D, E. (Rien n'interdit d'écrire :  $A = A + B$ , ce qui exprime que l'on ajoute les valeurs des variables identifiées par A et B et qu'ensuite l'on assigne la valeur du résultat obtenu à la variable identifiée par A.)

On appelle **compilateur** le programme qui analyse un tel langage et fournit un code objet semblable à celui fourni par un assembleur, c'est à dire susceptible d'être chargé par un chargeur en vue de son exécution.

### Notion de récurrence; boucles, indices, registres d'index.

Soit à faire la somme de n grandeurs rangées sous forme d'un tableau.

### Gestion de boucle en langage évolué.

On utilise un identificateur pour désigner l'ensemble du tableau : soit TAB cet identificateur. Un élément du tableau est désigné par l'identificateur du tableau suivi du numéro de l'élément dans le tableau placé entre parenthèses. Ce numéro peut lui-même être remplacé par une variable connue par son identificateur (ou même par une expression) : on l'appelle alors indice. Si N est l'identificateur associé au nombre d'éléments du tableau, on peut définir TAB (1) comme le premier élément du tableau et TAB (N) comme le dernier élément du tableau. Dans un langage évolué très simplifié, le programme calculant la somme des éléments du tableau peut s'écrire de la façon suivante, I étant l'identificateur de l'indice et S l'identificateur du résultat

```

I = 1
S = 0
10  S = S + TAB (I)
    I = I + 1
    SI (I — N) ≤ 0, ALLER A 10

```

Le programme est dit récurrent; on dit aussi qu'il contient une boucle. Les deux premières instructions initialisent la boucle, l'instruction étiquetée 10 représente à elle seule le corps de la boucle : chaque fois qu'on l'exécute on ajoute au résultat précédent un nouvel élément du tableau. Les deux dernières instructions contrôlent la boucle, la première en faisant progresser l'indice, la deuxième en testant si la boucle est terminée.

### Gestion de boucle, en langage assembleur

Nous indiquerons ici deux méthodes pour adresser les éléments d'un tableau dans une boucle

**Méthode par modification d'instruction.** On adresse les éléments successifs du tableau en incrémentant de 1, à chaque passage dans la boucle, l'adresse de l'instruction chargée d'additionner le nouvel élément du tableau. Dans la quasi-totalité des machines à une adresse, l'adresse se trouve placée en fin d'instruction, de sorte qu'il suffit d'ajouter 1 à cette dernière pour faire progresser l'adresse.

Ajoutons trois instructions au code opération de BOULIX

RAZ	remise à zéro de l'accumulateur
INCR	incrémentant de l'accumulateur
BRN	branchement si accumulateur négatif

Le programme de calcul de la somme des éléments du tableau s'écrit

	RAZ		} initialisation
	RNG	I	
	RNG	S	
BOUCLE	CHA	S	} addition d'un élément du tableau à la somme
ALFA	ADD	TAB	
	RNG	S	

CHA	ALFA	} progression de l'adresse de l'élément
INCR	ALFA	
RNG	ALFA	
CHA	1	} comptage
INCR	1	
RNG	1	
SUB	N	} bouclage si $i < n$
BRN	BOUCLE	

**Méthode par registre d'index.** On évite de modifier l'instruction en ne modifiant l'adresse de l'élément du tableau qu'au moment de l'exécution de l'instruction. Ceci nécessite l'adjonction d'un registre supplémentaire appelé registre d'index. L'instruction dont l'adresse doit être modifiée en

cours d'exécution est dite indexée; ici nous l'écrivons ADD TAB (X), le (X) désignant le registre d'index. Lors de l'exécution, le contenu du registre d'index est ajouté au contenu de la partie adresse de l'instruction indexée. Il suffit alors pour adresser successivement les divers éléments du tableau de faire progresser le contenu du registre d'index d'une unité à chaque passage dans la boucle.

Ajoutons les instructions nécessaires à la gestion du registre d'index TAX transfert du contenu de l'accumulateur dans le registre d'index

TXA                    transfert du contenu du registre d'index dans l'accumulateur

INCX                    incrémentation de l'index.

Le programme peut alors s'écrire

	RAZ		} initialisation à zéro du registre d'index et de la somme.
	TAX		
	RNG	S	} addition d'un élément du tableau à la somme.
BOUCLE	CHA	S	
	ADD	TAB(X)	
	RNG	S	
	INCX		comptage dans le registre d'index.
	TXA		} bouclage si le contenu de l'index est inférieur à n.
	SUB	N	
	BRN	BOUCLE	

REMARQUE : chaque machine a sa méthode d'utilisation du registre d'index :

certaines comptent de -n à zéro (ou encore comptent à rebours de n à 0) de façon à utiliser une seule

instruction de fin de boucle qui à la fois incrémente (ou décrémente) l'index et le compare à zéro;

certaines utilisent deux registres, l'un contenant l'index, l'autre sa valeur maximum,

l'instruction de fin de boucle incrémentant le contenu du premier registre avant de le comparer à celui du second ; etc.

### Notion de sous-programme.

Si l'on a plusieurs fois à exécuter le même traitement au cours d'un programme, on peut ne le programmer qu'une fois sous forme de **sous-programme**.

La figure 17 représente deux appels du sous-programme SP, dans le corps du programme (souvent appelé programme principal).

Ainsi pourra-t-on mettre sous forme de sous-programme le calcul d'une fonction, la résolution d'une équation algébrique, la résolution d'un système linéaire, l'inversion d'une matrice. Un sous-programme est écrit en toute généralité, c'est-à-dire sans faire aucune hypothèse ni sur le programme qui l'appelle et son emplacement en mémoire ni sur les emplacements des données de ce programme qu'il aura à utiliser. Aussi deux problèmes se posent

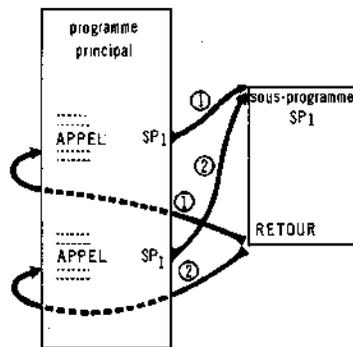


Fig. 17. Appel du sous-programme SP à partir de deux emplacements du programme principal.

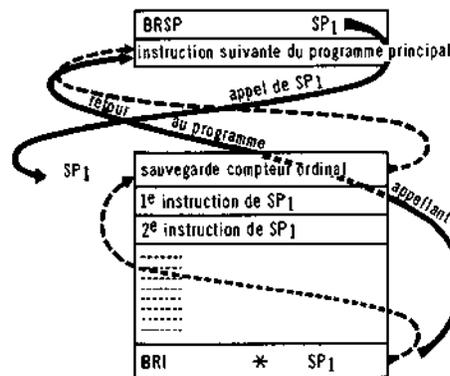


Fig. 18. Instruction d'appel de sous-programme et de retour au programme principal.

### Sauvegarde de l'adresse de retour

L'appel d'un sous-programme SP, se fait par une instruction BRSP SP, où le code instruction BRSP signifie branchement vers sous-programmes et la partie adresse SP, représente l'adresse du début du sous-programme. Si cette instruction n'exécutait qu'un simple branchement à l'adresse SP, le sous-programme pourrait se dérouler mais aurait perdu toute trace du programme qui l'a appelé ; en particulier il ne pourrait y revenir. L'instruction BRSP est donc un branchement spécial qui préserve l'adresse de retour, c'est-à-dire la valeur du compteur ordinal au moment de l'appel. Cette sauvegarde de l'adresse de retour de l'instruction suivant le BRSP est généralement faite soit dans un registre spécialisé, soit dans le premier mot du sous-programme.

Dans cette dernière hypothèse très fréquemment utilisée, l'instruction BRSP ajoute 1 au compteur ordinal, range son contenu à l'adresse SP, puis recharge le compteur ordinal avec la valeur SP, + 1 qui représente l'adresse de la première instruction du sous-programme (fig. 18). L'instruction de retour doit être un branchement à la cellule mémoire dont l'adresse est contenue dans le premier mot du sous-programme. On utilise généralement une nouvelle possibilité d'adressage que l'on trouve dans un grand nombre de machines. Il s'agit de l'adressage indirect. Dans l'adressage indirect, l'adresse qui figure dans la partie adresse de l'instruction ne pointe pas directement sur l'information cherchée, mais sur une cellule mémoire qui pointe elle-même sur l'information cherchée. Ainsi la dernière instruction du sous-programme sera un branchement inconditionnel (BRI) avec adressage indirect (ce que l'on note souvent par une étoile) portant sur l'adresse symbolique SP, : BRI \* SP,.

### Transmission des arguments

Le programme principal fournit un certain nombre de données au sous-programme et escompte que ce dernier lui restitue les résultats correspondants. Données et résultats sont appelés **arguments** du sous-programme. Le sous-programme ne connaissant pas les adresses symboliques des arguments dans le programme appelant, il y a un problème connu sous le nom de problème de transmission ou passation des arguments. Cette opération peut être constituée soit par la transmission directe des valeurs des arguments (transmission par valeur), soit par la transmission de leurs adresses (transmission par nom).

On lui connaît différentes solutions

### Transmission par registres.

C'est la solution retenue lorsque le nombre d'arguments est très faible. Par exemple pour un sous-programme calculant une fonction d'une variable, on pourra utiliser l'accumulateur ce dernier contiendra la valeur de la variable lors de l'appel ; le sousprogramme y rangera la valeur de la fonction avant le retour.

### Transmission par zone commune fixe.

On peut décider qu'une zone fixe placée généralement en début de mémoire est réservée à la transmission des arguments de tous les sous-programmes. Cette solution est rarement retenue, car malcommode si on utilise plusieurs niveaux de sous-programme (un sous-programme pouvant être sous-programme d'un autre sous-programme) et surtout si on travaille avec interruptions, un sous-programme pouvant être interrompu par un programme qui appelle lui-même un sous-programme.

### Transmission par séquence d'appel.

L'instruction BRSP est suivie soit du tableau des valeurs des arguments successifs dans le cas d'une transmission par valeur, soit du tableau des adresses des arguments successifs dans le cas d'une transmission par nom.

#### TRANSMISSION PAR VALEUR

BRSP SP<sub>1</sub>

ARG<sub>1</sub>

ARG<sub>2</sub>

...

ARG<sub>n-1</sub>

ARG<sub>n</sub>

#### TRANSMISSION PAR NOM

BRSP SP<sub>1</sub>

ADARG<sub>1</sub>

ADARG<sub>2</sub>

...

ADARG<sub>n-1</sub>

ADARG<sub>n</sub>

On remarque que l'adresse sauvegardée par l'instruction BRSP pointe sur le premier argument dans la transmission par valeur, sur l'adresse du premier argument dans la transmission par nom. Si l'adresse est sauvegardée dans la première cellule mémoire du sous-programme, les instructions du sous-programme pourront atteindre soit les arguments, soit leurs adresses par une combinaison de l'adressage indirect, et de l'adressage indexé, le registre d'index devant contenir le numéro de l'argument.

Ainsi l'instruction CHA \* SP<sub>1</sub> (X), le registre d'index contenant 4 permettra de charger dans l'accumulateur soit la valeur du quatrième argument, soit son adresse. On constate ici la nécessité de pouvoir combiner les types d'adressage indirect et indexé, l'indexation devant s'effectuer après l'indirection.

### Transmission par registre de base

Nous réservons cette solution pour un paragraphe ultérieur consacré aux sous-programmes réentrants.

## Invariance, réentrance et récursivité.

### Notion d'invariance.

Un programme en langage machine obtenu soit par assemblage, soit par compilation est dit invariant s'il ne se modifie pas en cours d'exécution. Ceci exclut notamment toute modification de la partie adresse des instructions (cf. 6. 2. 2. 1.). Il est clair cependant que les données sur lesquelles travaille le programme ne peuvent être toutes invariantes : elles seront donc implantées dans une autre zone de la mémoire.

Sans être forcément invariant, un programme peut être réutilisable après exécution ou encore réutilisable en série si des modifications intervenant au cours de son exécution se trouvent corrigées en fin d'exécution.

### Notion de réentrance

Un sous-programme est dit **réentrant** si plusieurs programmes peuvent l'utiliser « simultanément ». Par « simultanément » nous entendons des situations telles que la suivante : du fait du jeu des interruptions, un sous-programme peut avoir été commencé avec certains arguments sur l'appel d'un premier programme A ; une interruption peut avoir passé la main à un deuxième programme B qui lui-même appelle le sous-programme avec de nouveaux arguments ; lorsque le programme B sera terminé et que le contrôle sera redonné au programme A, il faut que le sous-programme puisse se terminer avec les arguments de A.

La réentrance suppose deux conditions

(1) Le code machine d'un sous-programme réentrant doit être invariant.

(2) Les données du sous-programme (arguments, adresses de retour et mémoires de travail) doivent être complètement séparées des instructions et occuper des zones de mémoire différentes pour les différents programmes utilisateurs du sous-programme. En général, les données du sousprogramme relatives à un programme appelant sont implantées dans une zone du programme appelant.

Pour cela il est commode de disposer d'un nouveau type d'adressage l'adressage par **base et déplacement**. Le calculateur possède un registre de base que le programme appelant charge avec l'adresse de la zone qu'il réserve aux données du sous-programme. Dans les instructions du sousprogramme adressant des données, la partie adresse de l'instruction contient non pas l'adresse de la donnée car elle est inconnue du sous-programme, mais le numéro de la donnée dans la zone de donnée. Ce numéro est appelé déplacement. L'adresse de la donnée est obtenue par addition du déplacement au contenu du registre de base. (Nous verrons que le registre de base doit être sauvegardé au moment d'une interruption, restauré par la suite.

### Notion de récursivité

On dit qu'un sous-programme est récursif s'il peut s'appeler lui-même.

Prenons un exemple type : le calcul de factorielle N pour N entier positif selon les formules

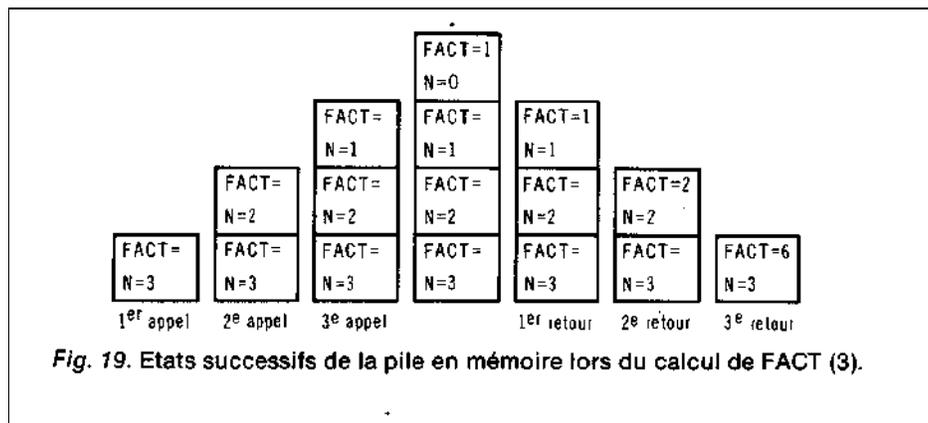
$$\text{FACT}(N) = N * \text{FACT}(N - 1)$$

$FACT(0) = 1$

Supposons que nous entrons dans le sous-programme  $FACT$  avec  $N = 3$  ; arrivé sur l'instruction  $FACT(N) = N * FACT(N - 1)$  il s'appelle lui-même avec  $N = 2$  ; mais dans ce nouveau passage, il s'appelle avec  $N = 1$  ; dans ce nouveau passage, il s'appelle avec  $N = 0$ . Après ces appels successifs, il y a les retours successifs : pour  $N = 0$ , la formule d'initialisation donne  $FACT = 1$  ; mais il y a retour au sous-programme avec  $N = 1$ , ce qui donne encore  $FACT = 1$  ; puis retour avec  $N = 2$ , ce qui donne  $FACT = 2$  ; puis retour avec  $N = 3$ , ce qui donne le résultat  $FACT = 6$ , enfin retour au programme appelant qui pourra utiliser ce résultat.

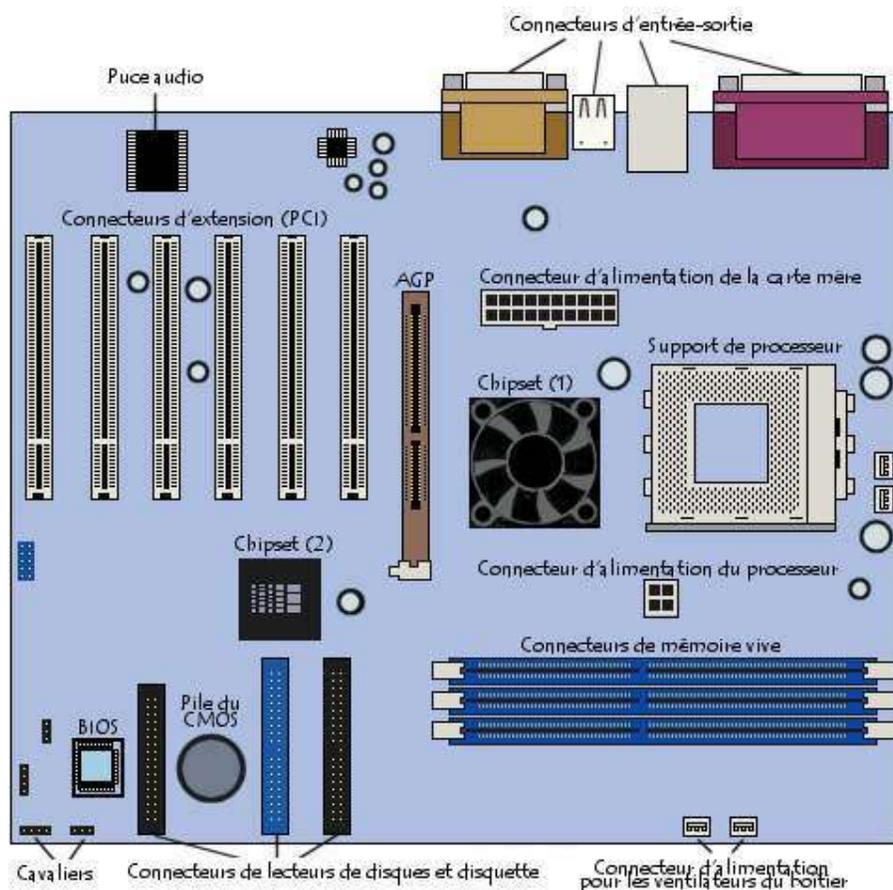
Tout le problème de la récursivité consiste à sauvegarder les valeurs des données sur lesquelles travaille le sous-programme à chaque nouvel appel, et à savoir les retrouver au retour.

La figure 19 montre comment l'utilisation d'une pile en mémoire permet de résoudre ce problème.



# La carte mère

L'élément constitutif principal de l'ordinateur est la **carte mère** (en anglais « *mainboard* » ou « *motherboard* », parfois abrégé en « *mobo* »). La carte mère est le socle permettant la connexion de l'ensemble des éléments essentiels de l'ordinateur.



Comme son nom l'indique, la carte mère est une carte maîtresse, prenant la forme d'un grand circuit imprimé possédant notamment des connecteurs pour les cartes d'extension, les barrettes de mémoires, le processeur, etc.

## Caractéristiques

Il existe plusieurs façons de caractériser une carte mère, notamment selon les caractéristiques suivantes :

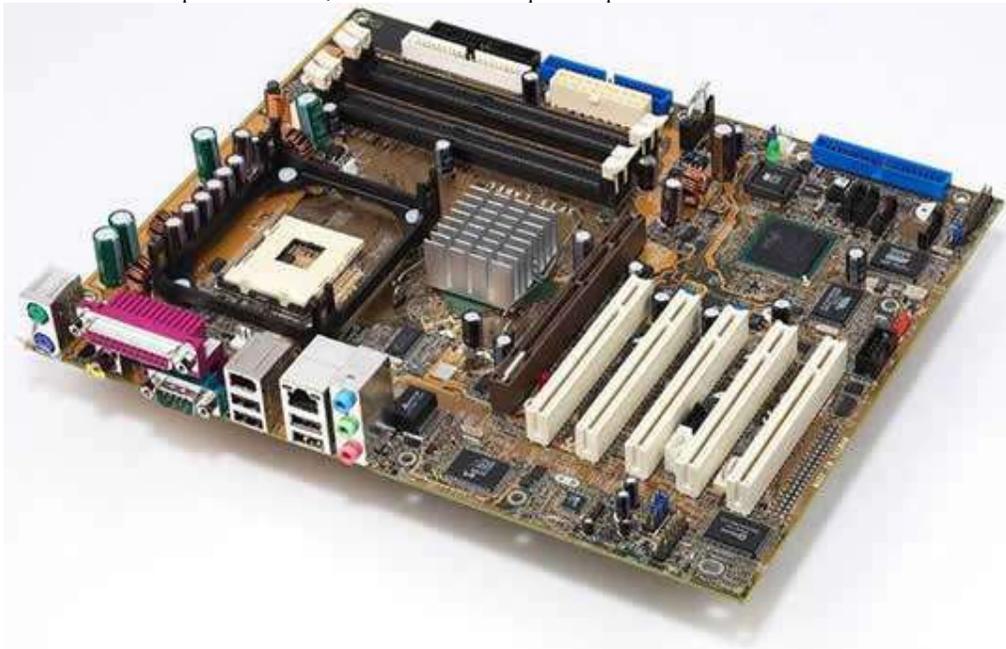
- le facteur d'encombrement,
- le chipset,
- le type de support de processeur,
- les connecteurs d'entrée-sortie.

## Facteur d'encombrement d'une carte mère

On désigne généralement par le terme « **facteur d'encombrement** » (ou *facteur de forme*, en anglais *form factor*), la géométrie, les dimensions, l'agencement et les caractéristiques électriques de la carte mère. Afin de fournir des cartes mères pouvant s'adapter dans différents boîtiers de marques différentes, des standards ont été mis au point :

Des **standards** ont été mis en place pour que les cartes mères puissent s'adapter dans des boîtiers (châssis de l'unité centrale) de marques différentes.

- Le format **AT** (305x305 mm) et son dérivé Baby-AT ont été utilisés à partir de 1984 sur les premiers PC, avant d'être remplacés par le format ATX .



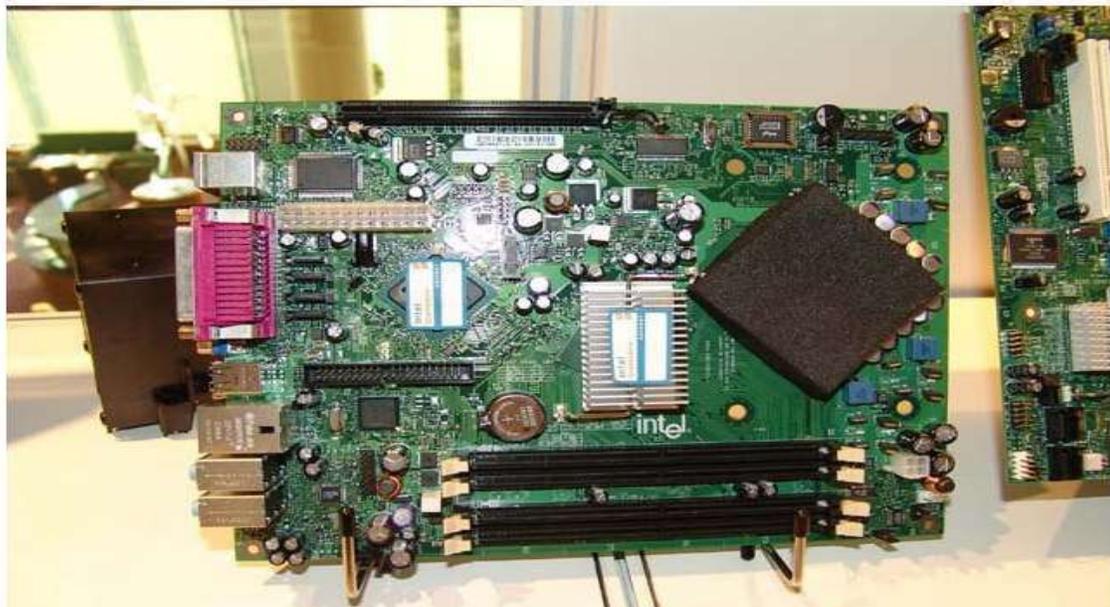
- Le format **ATX** (305x244 mm) est une évolution du format Baby-AT apparu en 1995 et doté d'une meilleure ergonomie  
Ses dérivés sont le Micro-ATX (244x244 mm), le Flex-ATX (229x191 mm) et le Mini-ATX (284x208 mm).



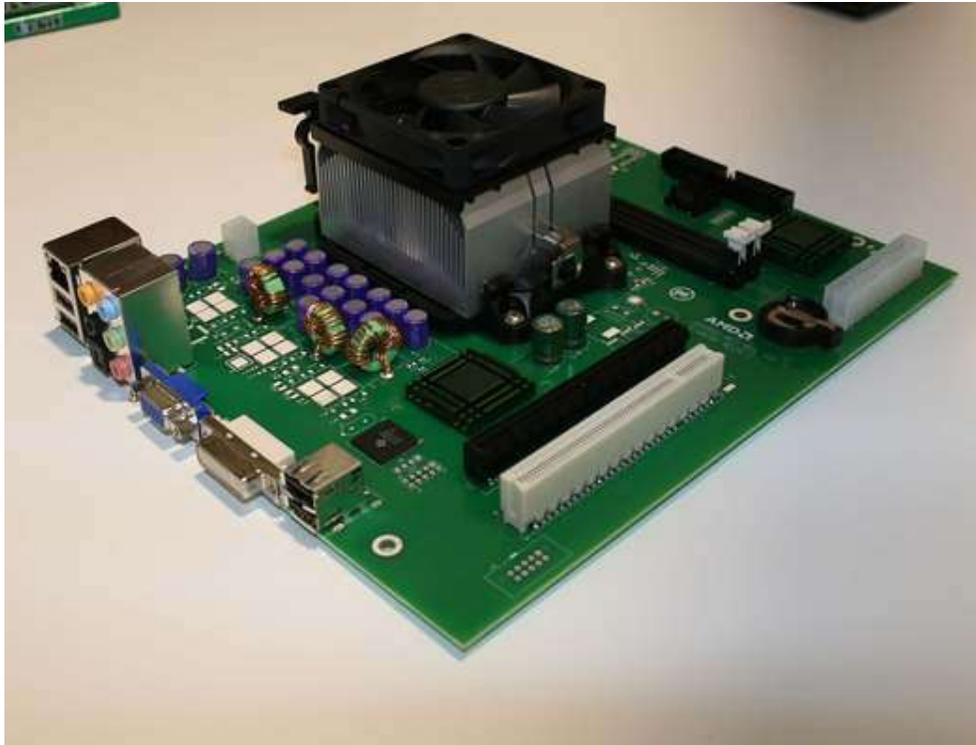
- Le format **ITX** (Information Technology eXtended, lancé par la société Via en 2001) est extrêmement compact (215x195 mm) et a pour dérivés le Mini-ITX (170x170 mm) et le Nano-ITX (120x120 mm).



- Le standard **BTX** (Balanced Technology eXtended) lancé par la société Intel en 2005 correspond à trois formats : BTX standard (325x267 mm), Micro-BTX (264x267 mm) et Pico-BTX (203x267 mm).



- Le format **DTX**, lancé en 2007 par AMD, a pour dimensions 248 x 203 mm tandis que le MiniDTX présente des dimensions réduites à 170 x 203 mm.



### **Format XT.**

Le format XT date des ordinateurs 8088 et 8086. Comme peu d'entre vous en rencontreront à part dans un musée, je passe. Ce format est spécifique à ces boîtiers (de même que les fixations alimentations). De toute façon, **Tous les composants** sont spécifiques à ces machines: disques durs et contrôleurs, cartes écrans, ... En cas de panne, vous ne pouvez rien faire, même pas récupérer les données du disque via un autre PC. En plus, ces machines utilisent généralement le DOS 3.0 ou 3.3 dont la FAT n'est pas lisible par les DOS et forcément Windows supérieurs.

### **Format AT – Baby AT.**

Avec l'introduction du 286 en version assembleur, les constructeurs de cartes mères ont normalisé les fixations mécaniques de ces cartes. Ceci permet d'insérer n'importe quelle carte mère AT dans un boîtier de type AT.

Le format des cartes mères AT est de 22 cm de large pour 33 cm de long. Il fut très utilisé avec les cartes pour 286, 386, 486 et Pentium. Les fabricants de carte mère ont suivant réduit la longueur au fur et à mesure de l'évolution dans l'intégration des circuits intégrés dans un seul boîtier et de l'évolution du multi-couche des cartes mères.

### **Format ATX**

Basé sur une spécification d'INTEL de 1997, les cartes mères ATX diffèrent nettement des cartes AT. L'ergonomie a été complètement remaniée. Le processeur a été rapproché de l'alimentation électrique. Désormais, les ports série, parallèle, contrôleurs IDE et FD, PS2 (clavier et souris), USB sont directement intégrés sur la carte mère. Déjà dans les cartes Pentium, les ports IDE et

FD étaient intégrés sur la carte mère. Ceci est dû à l'intégration de la fonction IDE et FD dans le chipset.

Le connecteur a été complètement revu. Ceci affecte les spécifications du régulateur de tension, maintenant intégré dans l'alimentation, alors que les circuits et leurs refroidisseurs étaient autrefois intégrés sur la carte mère. De ce fait, la fabrication d'une carte mère est moins coûteuse. Ces cartes restent sous tension lorsque le PC est éteint. Un simple signal via un interrupteur redémarre la carte mère.

### **Format LPX**

Ce format de carte mère a été utilisé par Western Digital en 1987. Les spécificités n'ont jamais été réellement publiées. Compaq et Packard Bell ont été les premiers utilisateurs. Les slots d'extension ne sont pas fixés sur la carte comme dans les formats AT et ATX, mais à l'aide d'une carte fixée perpendiculairement à la carte mère. Ceci explique que les boîtiers utilisant ce format sont souvent de faible hauteur. Actuellement, à part pour certaines machines spéciales (pensez au serveur "en rack"), ce format n'est plus utilisé.

### **Format NLX**

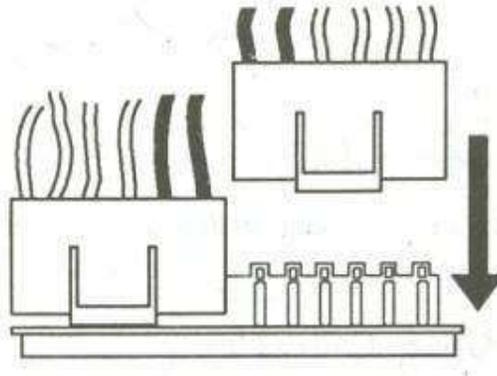
Lancé en 1997, ce format de carte mère est basé sur le LPX. Il intègre des caractéristiques mécaniques identiques, avec 2 cartes distinctes. La première intègre le processeur, la mémoire cache et la mémoire vive ainsi que les ports intégrés. La carte fille intègre les slots d'extension. Lors du démontage du boîtier, seule la carte principale est retirée, la carte fille restant en place. Les technologies actuelles sont toutes intégrées sur ces cartes, AGP, USB

Ainsi, du choix d'une carte mère (et de son facteur de forme) dépend le choix du boîtier. Le tableau ci-dessous récapitule les caractéristiques des différents facteurs de forme :

<b>Facteur de forme</b>	<b>Dimensions</b>	<b>Emplacements</b>
ATX	305 mm x 244 mm	AGP / 6 PCI
microATX	244 mm x 244 mm	AGP / 3 PCI
FlexATX	229 mm x 191 mm	AGP / 2 PCI
Mini ATX	284 mm x 208 mm	AGP / 4 PCI
Mini ITX	170 mm x 170 mm	1 PCI
Nano ITX	120 mm x 120 mm	1 MiniPCI
BTX	325 mm x 267 mm	7
microBTX	264 mm x 267 mm	4
picoBTX	203 mm x 267 mm	1

### **Alimentation PC et boîtiers associés**

## Alimentation AT

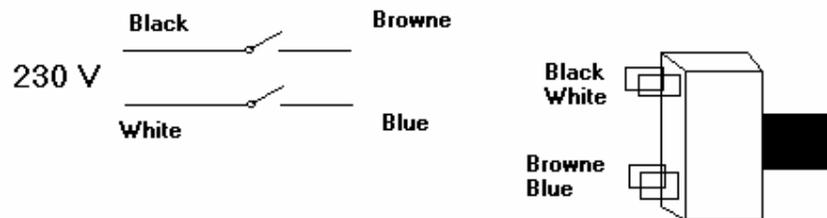


Les premiers XT utilisaient leurs propres alimentations (power supply). Si la forme et les fixations ont changées dans les 286 de type clones, les tensions présents sur les connecteurs sont restées les mêmes. Ces alimentations PC (incluent dans les boîtiers) sont appelées de type AT. Pour la connexion sur la carte mère, les fils noirs de chaque connecteur (les masses) doivent se trouver au milieu

### Brochage d'une alimentation AT

Pin	Description	Pin	Description
12	+ 5V	6	Masse
11	+ 5V	5	Masse
10	+ 5V	4	- 12 V
9	+ 5V	3	+ 12 V
8	Masse	2	+ 5 V
7	Masse	1	Alimentation correcte

Le connecteur 230 V passe directement de la prise électrique extérieure vers l'interrupteur (généralement en face avant). Il revient vers l'alimentation ensuite. Bricoler sur l'interrupteur nécessite de **RETIRER LA PRISE ELECTRIQUE**.

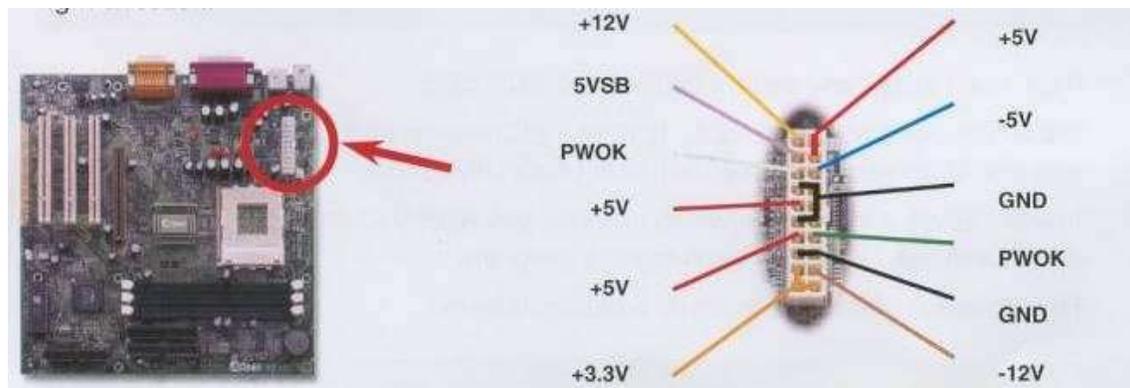


Les fils Black (noir) et White (blanc) viennent du réseau électrique (en passant directement à travers l'alimentation) puis repartent à l'alimentation par les fils brown (brun) et blue (bleu). Ces couleurs peuvent changer d'une alimentation à l'autre, les couleurs sont la plus part du temps imprimées sur l'alimentation AT.

## Alimentation ATX pour PC

Depuis les premiers Pentium II, les alimentations (et donc les boîtiers) sont de type ATX. Dans ce cas, l'interrupteur n'est plus relié sur le réseau 220V, mais sur la carte mère. L'alimentation alimente en permanence la carte mère avec une tension de vérification. En appuyant sur l'interrupteur, le signal est envoyé sur la carte mère qui demande à l'alimentation de démarrer. Ceci n'est pas sans risques. Comme la carte mère est toujours sous tension, même PC éteint, une sur-tension sur le réseau électrique provoque généralement des pannes irréversibles. De plus, en cas de sur-tension, il arrive que le PC ne démarre, ni ne s'éteigne plus. Il suffit de **retirer la prise électrique** quelques minutes et de la remettre avant de redémarrer le PC. Vous trouverez ci-dessous le brochage du connecteur ATX sur la carte mère.

L'avantage de l'ATX, le PC se coupe tout seul (sous Windows 95/98 et Windows 2000) et peut redémarrer directement par un signal modem par exemple.



### Brochage d'une alimentation ATX

Certaines cartes mères dans la période de transition (1998 – 1999) les possédait 2 modes de connections.

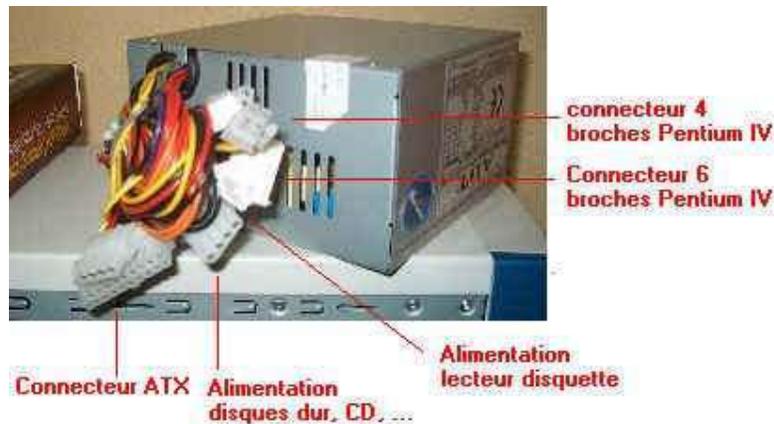
### Alimentation ATX pour Pentium 4.

Les ordinateurs Pentium IV nécessitent deux connecteurs d'alimentation supplémentaires.

Le premier fournit une tension électrique de + 12 V DC sur un connecteur de 4 pins

Le deuxième fournit des tensions de + 3,3V et 5V sur un connecteur 6 pins.

Cette alimentation est référencée sous la norme ATX 2.03, aussi appelée ATX12V. Un Pentium IV nécessite minimum une alimentation de 300 W pouvant fournir 20 A sur le +5 et 720ma sur le +5SVB. Avec des configurations courantes, les Pentium IV fonctionne sans les connecteurs d'alimentation auxiliaires, mais mieux vaut prévenir que guérir.



### Autres caractéristiques des alimentations.

On retrouve d'autres caractéristiques des alimentations AT et ATX:

La puissance. Si les premiers 486 utilisaient des alimentations 120 W, le standard était plutôt de 200 W. Elle est de minimum 300 W conseillée pour les Pentium IV. Plus vous insérez de périphériques dans un PC, plus l'alimentation doit être puissante. Dans le cas de 4 disques durs, l'alimentation doit minimum être de 300 W.

Certaines alimentations sont de type autorégulées. Ces alimentations utilisent une sonde de température interne qui fait tourner le ventilateur uniquement en cas de besoin. Ces alimentations sont plus silencieuses

### Composants intégrés

La carte mère contient un certain nombre d'éléments embarqués, c'est-à-dire intégrés sur son circuit imprimé:

- Le chipset, circuit qui contrôle la majorité des ressources (interface de bus du processeur, mémoire cache et mémoire vive, slots d'extension,...),
- L'horloge et la pile du CMOS,
- Le BIOS,
- Le bus système et les bus d'extension.

En outre, les cartes mères récentes embarquent généralement un certain nombre de périphériques multimédia et réseau pouvant être désactivés :

- carte réseau intégrée ;
- carte graphique intégrée ;
- carte son intégrée ;
- contrôleurs de disques durs évolués.

### Le chipset

Le **chipset** (traduisez *jeu de composants* ou *jeu de circuits*) est un circuit électronique chargé de coordonner les échanges de données entre les divers composants de l'ordinateur (processeur, mémoire...). Dans la mesure où le chipset est intégré à la carte mère, il est important de choisir

une carte mère intégrant un chipset récent afin de maximiser les possibilités d'évolutivité de l'ordinateur.

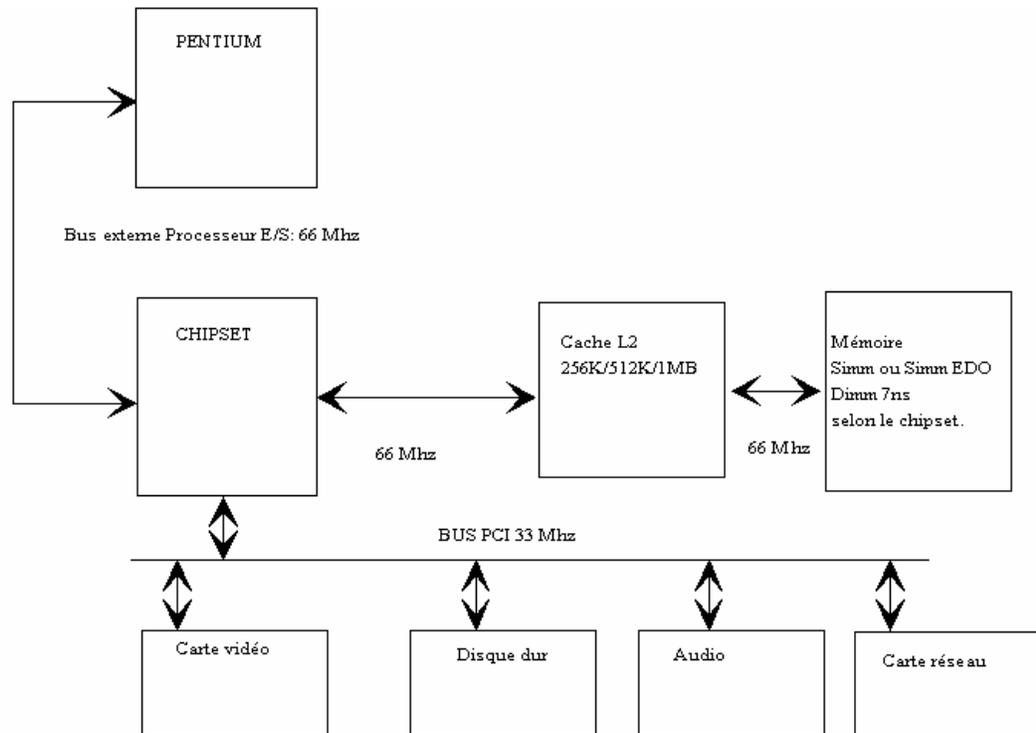
Certains chipsets intègrent parfois une puce graphique ou une puce audio, ce qui signifie qu'il n'est pas nécessaire d'installer une carte graphique ou une carte son. Il est toutefois parfois conseillé de les désactiver (lorsque cela est possible) dans le setup du BIOS et d'installer des cartes d'extension de qualité dans les emplacements prévus à cet effet.

L'avancée technologique des processeurs, des fréquences associées (on double la vitesse des processeurs chaque année) est limitée par des goulets externes: les périphériques, mémoires, spécifications des bus sont loin d'atteindre les vitesses des processeurs. Les fabricants de processeurs se sont donc beaucoup attachés au circuit de gestion externe d'interface processeur - périphériques: le **CHIPSET**. Si une carte mère 8088 comportait une centaine de circuits électroniques, le nombre se restreint par l'utilisation de Chipset qui reprennent l'ensemble des fonctions de contrôles internes (bus et disques durs) à celui des périphériques externes (USB, série, parallèle, ...). Le chipset sert d'interface entre le processeur et ses périphériques (mémoire, entrées / sorties). C'est le composant essentiel d'une carte mère.

Le chipset n'est généralement pas constitué d'un seul circuit, mais de 8 pour les plus anciens et 2 dans les circuits modernes (le Northbridge et le Southbridge). En gros, le **Northbridge** sert de lien entre le processeur et la mémoire, tandis que le **southbridge** gère les entrées / sorties (bus PCI et AGP, ...). Tous les fabricants actuels (VIA, NVIDIA, ULI et INTEL) utilisent le même principe: un northbridge adapté à chaque modèle de processeur lié à des southbridges interchangeables suivant le niveau de périphériques souhaités. Cette solution permet de ne développer qu'un seul circuit lorsque l'on change de caractéristique pour un processeur. De même, cette solution permet de fabriquer des cartes mères de différentes gammes en utilisant différents modèles de southbridges.

### Architecture Pentium

Voici la structure d'une carte mère pour processeur PENTIUM, PENTIUM MMX, K5 et K5 MMX (pas K6-2 ou K6-3). Le bus ISA est dérivé du bus PCI.



Selon la vitesse du processeur en interne, la vitesse externe peut être différente. On trouve sur les cartes mères des vitesses de 50, 60, 66 et 75 Mhz. Plus la vitesse externe est importante, plus la vitesse globale du PC est rapide. Les caractéristiques de départ sont données par le fournisseur du processeur, "overclocker" la vitesse externe peut apporter quelques problèmes. Pour les Pentium INTEL, la vitesse externe, appelée FSB pour Front Side Bus, est de 66 Mhz maximum! Les Simm sont limitées à cette vitesse.

Les CHIPSET utilisés pour ces processeurs sont la série INTEL 430, suivi de 2 lettres.

- **Le 430 LX** Sorti en 1993, c'est le premier chipset pour Pentium 60 Mhz et 66 Mhz. Pour rappel, les Pentium 60 et 66 utilisent un socket (et donc une carte mère) spécifique.
- **Le 430 NX** identique au 430 LX mais supporte les Pentium à 90 Mhz et 100 MHz.
- **Le 430 MX**, version 430 FX spécifique aux premiers ordinateurs portables à base de Pentium. Son contrôleur de disque ne gère pas la norme ATAPI donc les CD-ROM IDE.
- **Le 430 FX** appelé Triton, fut le tout premier jeu de composants PCI d'Intel à se rapprocher des limites théoriques du bus PCI v2.0 à 133 Mo/s. Attention, limitation des disques durs à 2,1 GB maximum
- **Le 430 VX** aussi appelé Triton II. Elle permet de partager avec le contrôleur graphique une partie de la mémoire vive du PC. Elle est destinée à mettre au point des PC à faible coût. Par contre, ce chipset permet de gérer les mémoires Dimm SDRAM. Ceci explique son succès chez les assembleurs
- Les **430 HX** ( Triton III ) apportent d'une part, le support du bus série **USB** ( Universal Serial Bus) et quelques spécificités de partage de bus ISA / PCI.
- **Le 430 TX** aussi appelé Triton IV (disques durs Ultra-ATA, optimisation de la gestion des mémoires SDRAM Dimm), supporte la présence d'un second processeur et la mémoire autocorrective **ECC**. En outre, il autorise jusqu'à 512 Mo de mémoire vive, contre 128 Mo seulement pour le modèle VX. Petite remarque, le 430 HX peut cacher 512 Mo à condition d'ajouter un composant à la carte mère appelé un Tag Ram. Le bus

USB est reconnu (en partie, surtout dû aux connecteurs et à Windows 95 B qui ne gère pas ce bus en pratique). Attention problème si mémoire Ram supérieure à 64 MB – elle n'est pas gérée en cache L2.

Les cartes mères à base de chipset FX et HX recevaient une **mémoire cache externe** de 256k ou 512K de type insertion. Même si les circuits sont équivalents de par leur forme, les caches ne sont pas compatibles entre-eux. En cas d'instabilité sur ces cartes, désactivez le cache externe au niveau du Bios.

### Architecture PENTIUM II 233 – 450 Mhz, Pentium III jusqu'à 600 Mhz, Pentium Pro

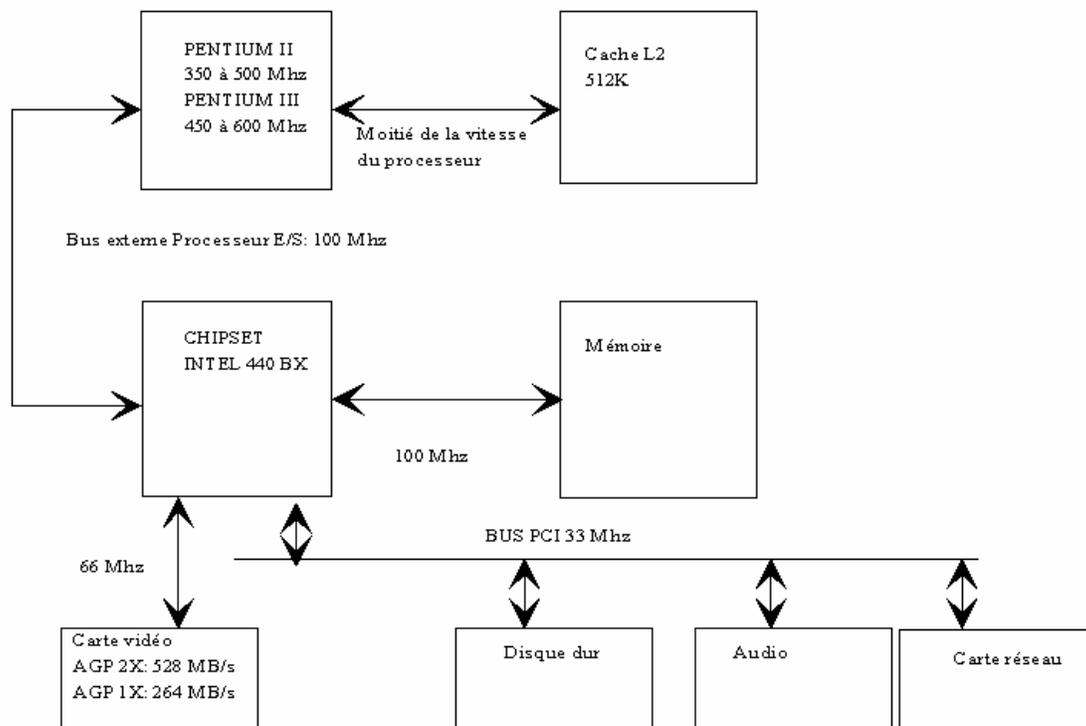
#### 440 LX: Pentium II 233 -333 Mhz

Les CHIPSET utilisés pour les premiers processeurs Pentium II sont la série INTEL 440, suivi de 2 lettres. Le 440 LX est le premier chipset sorti en grande série. Il est identique au 440 BX ci-dessus mais avec un bus processeur externe de 66 Mhz

L'architecture des CELERON est identique, excepté le cache. Notons que le 440EX est une version bridée du 440 LX destinée au CELERON

#### Architecture Pentium II de 350 à 450– premiers Pentium III 450 à 600 Mhz

La grande évolution ici vient du chipset **INTEL 440 BX**. Ce circuit d'interface gère les entrées sorties à 100 Mhz. Remarquez que vous pouvez toujours utiliser une carte 440 BX pour des Pentium II inférieurs ou des CELERON (attention au nouveau socket 370 pour les CELERON).



440 ZX est pratiquement identique au 440 BX. Il ne gère que 2 dimm

## Pentium Pro

Les premiers Pentium Pro (FSB 66) utilisaient des jeux de composants équivalents

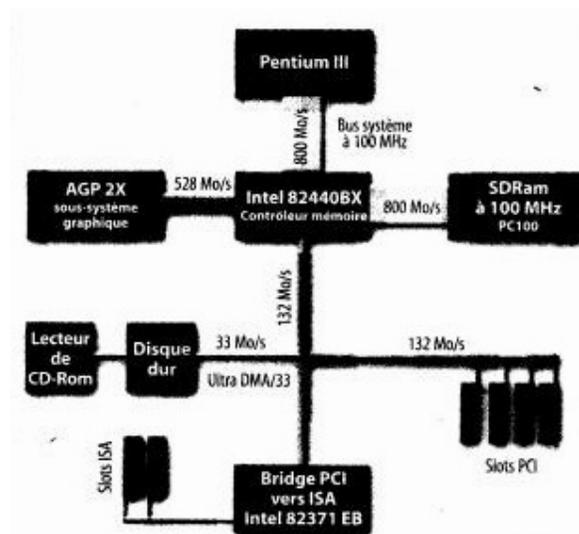
**Le 440 FX** équivalents au 430 HX, apporte la gestion de la mémoire ECC, c'est à dire de la mémoire autocorrective. Ce chipset a aussi été utilisé sur les premières cartes mères intégrant un Pentium II.

**Le 450 KX et le 450 GX** sont des chipsets spécifiques pour Pentium Pro. Ils supportent 2 processeurs simultanés. Le 450 GX permet de gérer 2 bus PCI et 2 Bus mémoire, un pour chaque couple de processeur. De plus il offre la possibilité de gérer un bus EISA-VLB (processeurs 486), principalement pour conserver une compatibilité avec les anciens serveurs, mais sans réel intérêt technique.

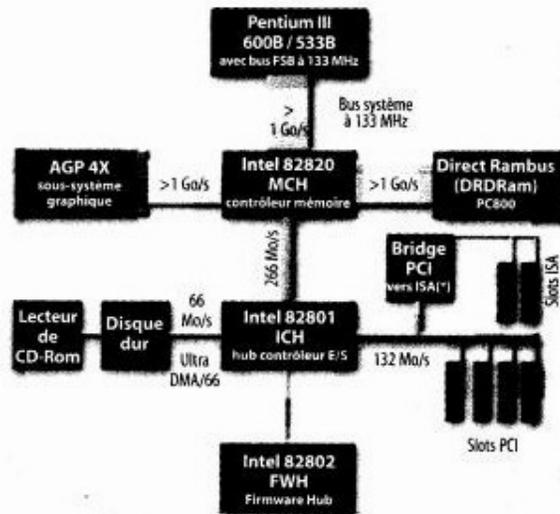
## Le Camino INTEL I820

Sorti fin 1999, le Camino permet au bus de donnée du processeur de fonctionner à une fréquence externe de 133 Mhz (Front Side Bus), pour 100 Mhz pour le 440 BX. Ceci apporte un gain de 33 % de performances. Il est utilisé avec les processeurs INTEL Pentium III de la série B (533B et 600B). Le i820 gère la Direct Rambus (aussi appelée **DRDRam**), un nouveau type de mémoire vive qui tourne à 300 ou 400 Mhz, plus chère que la PC100 Dimm. L'Intel 820 **n'est pas prévu** pour gérer la SDRam. Intel commercialise en option pour les cartes mères à base de i820 le *Memory Translator Hub* (ou MTH) qui permet de gérer la SDRam PC 100. Un des principaux intérêts de la DRDRam réside dans son débit suffisant pour l'AGP 4X, l'une des nouveautés de l'Intel 820. L'AGP 4X offre une bande passante supérieure à 1 GB/s - contre 528 MB/s pour l'AGP 2X. Le chipset i820 apporte d'autres améliorations, comme l'architecture *accelerated I-Hub* qui fournit des connexions directes à un débit de 266 MB/s (133 MB/s pour le 440BX) entre le chipset et les sous-systèmes contrôleurs IDE, USB et PCI. En même temps, l'interface Ultra DMA/66 double le débit des disques durs: 66 MB/s contre 33 MB/s pour l'Ultra DMA/33.

L'Intel 820 supporte le traitement biprocesseur en mode symétrique SMP (*symétrie multiprocessing*). En dernier, le i820 dispose d'un générateur de nombres aléatoires qui, selon Intel, améliore les opérations de cryptage, sécurité et authentification.



FSB 100 Mhz, mémoire PC 100 SDRam, AGP 2X, UDMA 33



FSB à 133 Mhz, mémoire Rambus DRDRAM, UDMA 66, AGP 4X. Le pont (bridge) vers un bus ISA est optionnel

### ALI Aladdin TNT2.

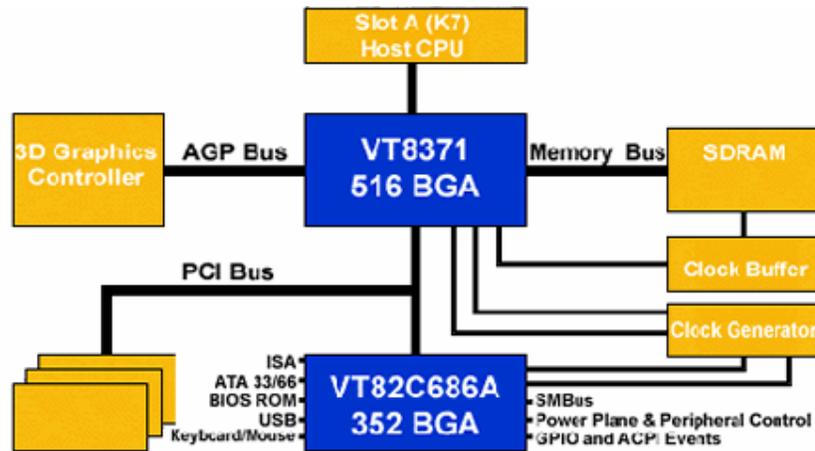
Reprenant les fonctionnalités du chipset P6ALI, il intègre le processeur graphique RIVA TNT2 de nVIDIA.

### VIA Apollo Pro 133 et Apollo VCM133. (Pentium III)

Ce circuit pour Pentium III est sorti fin 1999 (avant le I820) gère la mémoire PC133 (comme les Intel 810 et 815). C'est une des bases des circuits VIA: un northbridge spécifique suivant le processeur (notamment INTEL ou AMD) et le même southbridge. Les KX133 ci-dessous sont pour les AMD. Ses performances par rapport à une configuration identique à base de 440BX sont de 4 à 7 % supérieures. Selon le type de mémoire (PC133 ou VCM133), le chipset est différent.

### VIA Apollo KX133 et Apollo KT133 (Athlon et Duron)

VIA a sorti avant AMD le premier circuit d'interface spécifique Athlon, l'Apollo KX133. Celui-ci gère le bus EV6 200/266 Mhz développé par alpha, ainsi que l'AGP 4X, la SDRAM PC 133 et PC100 et l'UDMA 66. Il est utilisé autant en socket 462 qu'en slot A (Athlon et Duron).



Sorti courant 2000, le VIA Apollo KT133 est livré en parallèle avec le KX133. Il gère le socket 462 pour les Duron et Tunderbird. Identique au KX sur bien des points, il ne gère plus les bus ISA, et peut être vu comme une version allégée du KX.

### VIA ProSavage PM133. (Celeron / Pentium II / Pentium III)

Le marché des ordinateurs bon marché est probablement l'avenir. Pour concurrencer le i815 d'INTEL, VIA sort le ProSavage PM133. Ce chipset inclue un contrôleur Apollo Pro133 (donc CELERON et PENTIUM II / Pentium III), un contrôleur réseau Ethernet 10/100, un processeur graphique S3 Savage 4 (3D) et S3 Savage 2000 (2D). Il accepte la SDRAM 133. Il est également compatible avec les VIA Cyrix III.

### INTEL i840 (XEON)

Ce chipset est spécifique aux Xeon, il est de structure identique au i820, mais accepte le multi-processeur

### INTEL i820E

Avec le succès mitigé du i820 pour Pentium III (mémoire Rambus), INTEL sort mi-2000 le i815 pour le bas de gamme. Pour garder une certaine supériorité sur VIA, INTEL sort également le i820E. Celui-ci est identique au i820, mais accepte deux canaux (séparés) USB, une interface réseau, un double contrôleur ATA / 100 et un contrôleur audio 6 canaux.

De plus, il inclut un slot CNR incompatible avec le slot AMR. Ces techniques permettent de connecter des systèmes audio / modem. Dans ce dernier cas, celui-ci permet de redémarrer complètement le PC suivant un signal modem.

### Le I850 (Pentium IV)

Le i850 d'Intel est associé avec les premiers Pentium IV. C'est une évolution du i820 (pour Pentium III). Il autorise uniquement la RamBus. Apparemment, c'est lui qui est responsable des mauvaises performances des Pentium IV (en plus du pipeline un peu long)

L'i850 est composé de deux chips, à savoir un MCH (Memory Controller Hub) i82850 et un ICH (I/O Controller Hub) i8201BA. Le i8201BA porte également le nom ICH2, on le retrouve sur l'i815E par exemple.



Le Northbridge i82850 gère les bus suivants :

- Un bus processeur 100 MHz Quad Bumped pour Pentium 4
- Un bus mémoire doté de deux canaux de RDRAM en dual-channel.
- Un bus AGP 4x
- Un bus Interlink pour le Southbridge

L'i82850 ne gère que la RDRAM (Rambus). Il dispose de deux canaux RDRAM (tout comme l'i840 pour Xeon), et peut donc accéder à deux barrettes de Rambus simultanément, cumulant la bande passante. Ainsi, avec deux barrettes de RDRAM PC800 on arrive à une bande passante de 2.98 Go /s. Néanmoins, au prix de la mémoire RAMBUS, cette solution performante est très chère.

Le southbridge gère deux canaux Ultra ATA 100, intègre deux contrôleurs USB 1.1 (soit une bande passante de 2x12 Mb /s), un support réseau 10/100 Mbits et le son AC'97 sur 6 canaux.

### **AMD-760 (Athlon et Duron)**

AMD annonce en décembre 2000 la sortie de ce chipset qui gère la mémoire SDRAM DDR (Double Data Rate) en flancs montants et descendants. AMD a sorti des processeurs Athlon 1 GHz, 1,1 et 1,2 Ghz répondant à cette technologie. Une version spécifique est dédié aux Athlons MP (acceptant de travailler en bi-processeurs. A ma connaissance, AMD ne produit plus de chipset, laissant ce travail à des sociétés spécialisées comme VIA ou nForce.

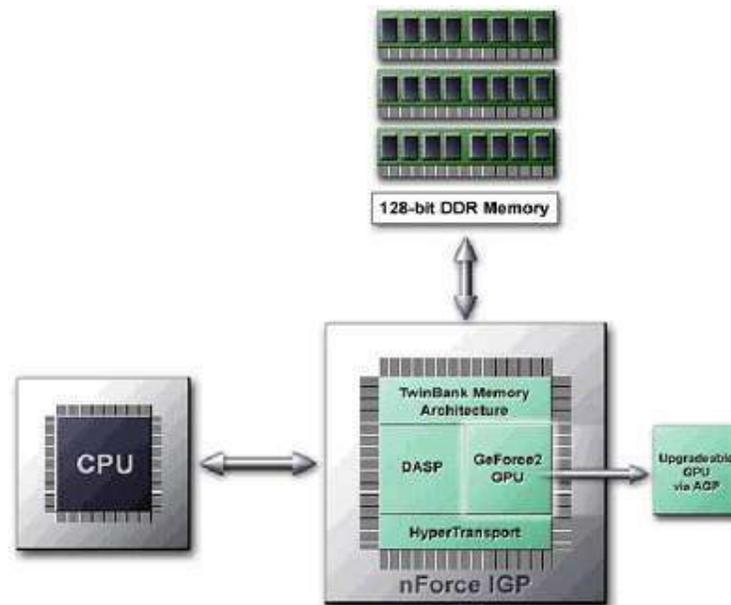
### **NVIDIA nForce pour Athlon et Duron.**

NVIDIA, initialement constructeur de circuits électroniques pour cartes graphiques, sort en juin 2001 une nouvelle architecture qui sera intégrée dans les prochains chipsets de la marque.

Le 82850 est intégré dans un package de type OLGA (Organic Land Grid Array). Ce package, plus petit et qui permet une meilleure dissipation thermique, avait été introduit par Intel avec le Pentium III. Le 82850 est surplombé d'un radiateur.



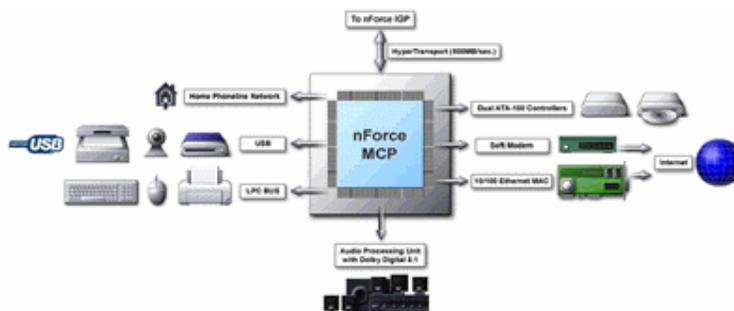
Comme la plupart des chipsets récents, le nForce est composé de deux chips : Le nForce Integrated Graphics Processor (IGP) comme Northbridge et le nForce Media and Communications Processor (MCP) comme Southbridge.



L'IGP (Northbridge) intègre deux contrôleurs mémoires indépendants, MC1 et MC2, ce qui permet de doubler la bande passante disponible lorsque l'on dispose de deux barrettes de mémoires de même capacité (ce qui double le taux de transfert effectif). Avec 2 barrettes de PC2100, on peut disposer de 3.96 Go /s grâce à cette architecture TwinBank, contre 2.98 Go /s pour les deux canaux de RDRAM PC800 gérés par les chipsets i840, i850 et i860. L'IGP intègre une carte graphique GeForce2. Proche du GeForce2 MX, ce processeur graphique dispose d'une interface interne de type AGP 8x. Le GPU peut être désactivé pour accueillir une carte graphique externe en AGP 4X.

Autre technologie intégrée dans l'IGP, le DASP (Dynamic Adaptive Speculative Pre-Processor). Il s'agit en fait d'un cache intelligent intégré dans le northbridge qui profite de la bande passante mémoire inutilisée pour charger des données qui devraient être requises plus tard par le processeur. Les résultats sont flagrants, puisque l'activation du DASP sur une plateforme de type Athlon 1.2 GHz + DDR permet de gagner de 20 à 30% de performances dans des benchs mémoires et de 6 à 8% dans les applications.

## Le MCP



IGP et MCP sont reliées via un bus utilisant la technologie HyperTransport de Digital - AMD. La bande passante offerte par ce bus est de 800 Mo /s, contre 'seulement' 266 Mo /s pour le V-Link de VIA et l'Interlink d'Intel.

Le MCP (le Southbridge) gère un bus PCI 32 bits, deux canaux Ultra ATA 100, deux contrôleurs USB 1.1 et un processeur audio hardware (256 voix en 2D et 64 en 3D), qui gère de 2 à 6 enceintes avec un encodeur Dolby Digital.

De plus, le MCP dispose d'un contrôleur Ethernet 10/100 Base-T hardware et d'un modem 56K émulé en software. Le gros avantage de contrôleur réseau 10/100 est d'être interfacé avec l'HyperTransport, qui gère les flux de données isochrones, ce qui permet aux données d'être délivrées avec une bande passante et un temps de latence garantis (NVIDIA appelle ça l'architecture StreamThru). Ceci offre une bande passante réelle de 10% supérieure à celle offerte par des cartes 10/100 3Com, Intel ou NetGear.

Les premières implémentations de l'architecture nForce seront le nForce 220 et le nForce 420, qui se distinguent par le support ou non de l'architecture TwinBank (DUAL-Channel Memory).

### **INTEL i845.**

Confronté aux prix de la mémoire RAMbus, INTEL sort fin décembre 2001 un chipset équivalent au 850 pour Pentium IV mais gérant les mémoires Dimm PC133/PC100, l'i845. La version i845D, sortie ensuite, est identique mais gère les mémoires DDR. Suivent plusieurs versions équivalentes selon les avancées technologiques. Remarquez que les 845 **ne gèrent pas les** ATA-133, uniquement les ATA-100.

La version 845G et suivantes sont des versions améliorées du 845E au niveau des fonctions graphiques améliorées dénommées Extreme Graphic. Suivant le Southbridge utilisé, certaines cartes mères (utilisée pour les CELERON) pourraient utiliser des mémoires PC133.

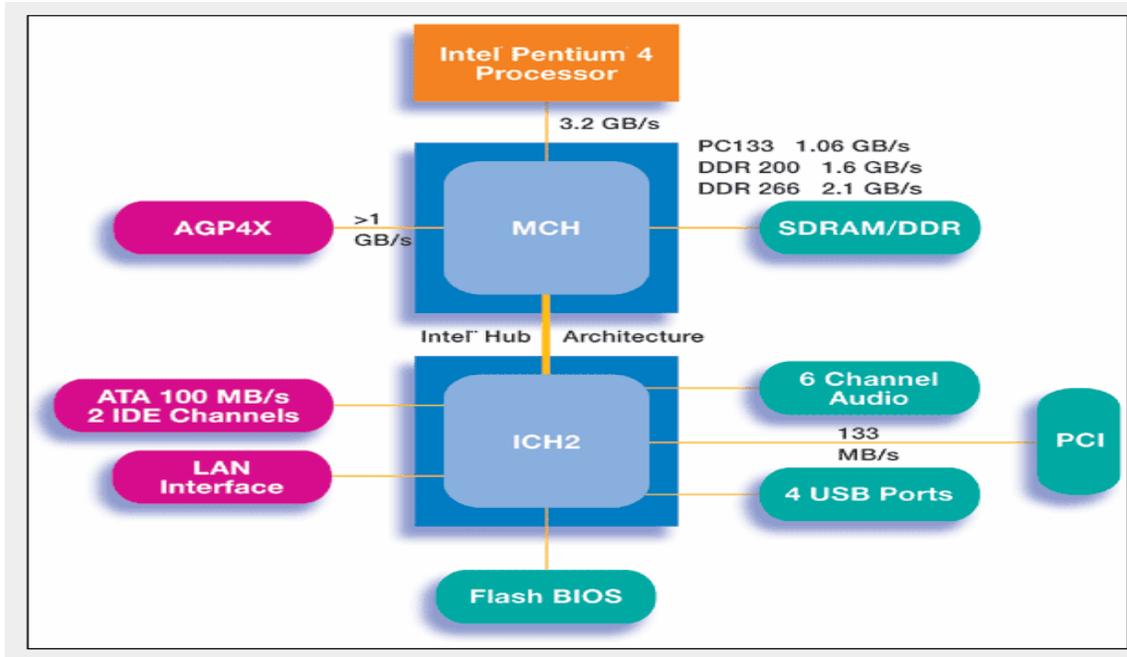


Schéma et caractéristiques du i845

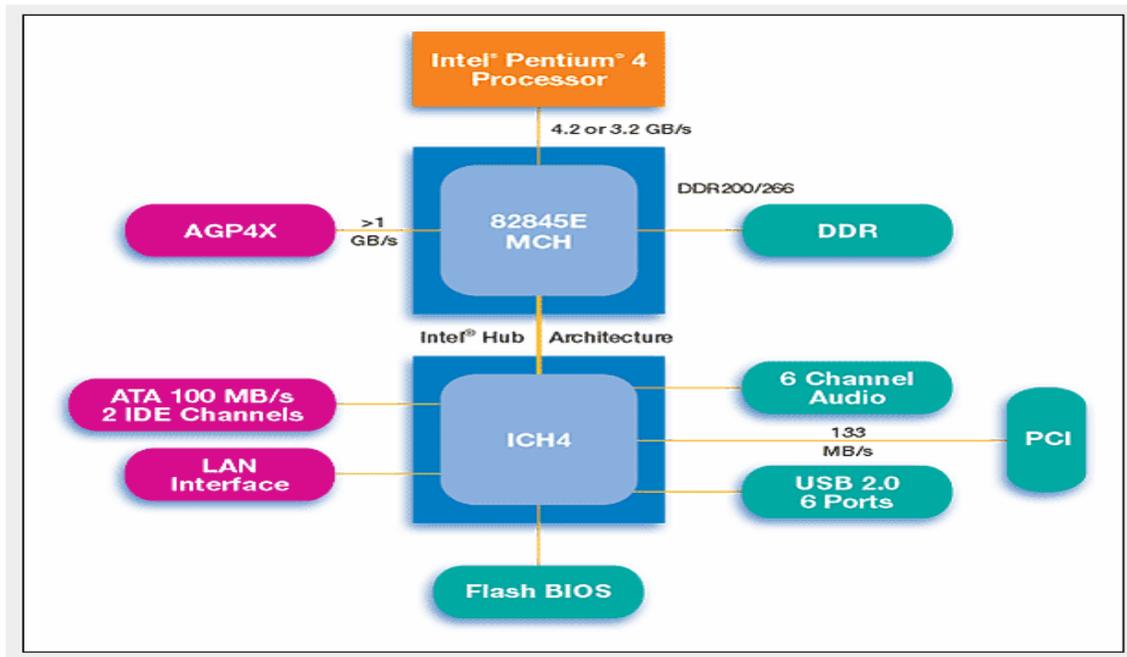
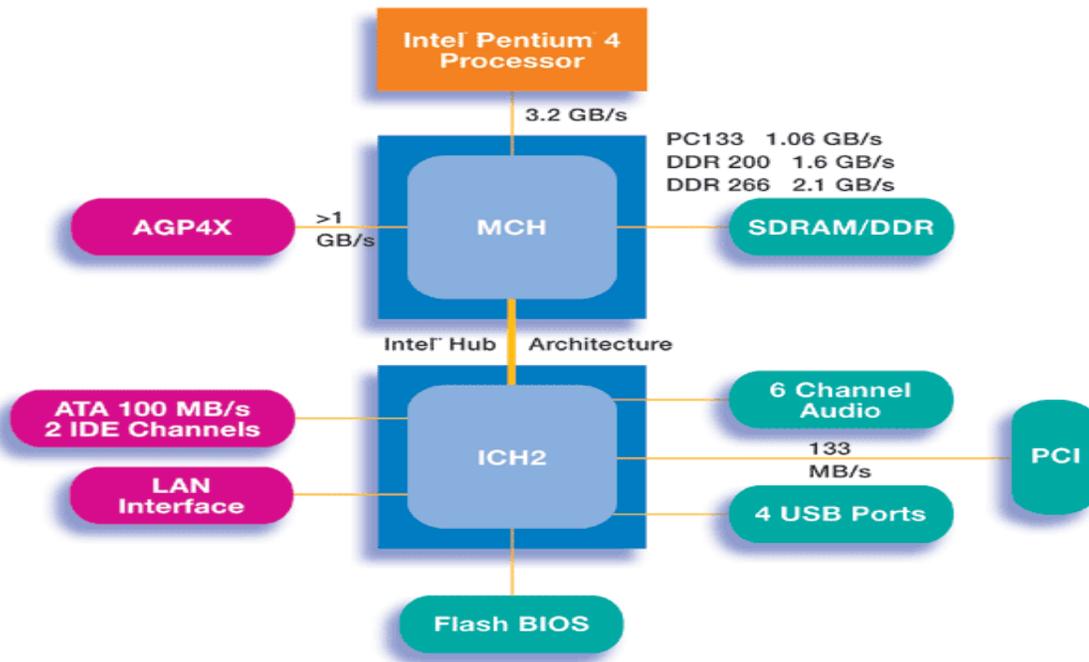


Schéma et caractéristiques du i845D



**Schéma et caractéristiques du i845E**

Confrontés aux performances des processeurs, INTEL comme d'autres constructeurs de chipsets s'est lancé dans le DUAL-Channel (aussi appelé Twin Bank). Cette technologie utilise 2 bancs de mémoires DDR333 ou DDR400 en parallèle et augmente de manière significative les performances mémoires du processeur. En même temps, les dernières technologies sont implantées sur la carte mère comme que le SERIAL ATA, l'AGP 8X et une technologie RAID dans certaines versions. Le i875PE accepte les mémoires auto-correctives ECC. Cette version est donc plus spécifiques pour des PC de haut de gamme de type serveur PC par exemple.

L'évolution du Chipset va de paire avec celle du microprocesseur. Comme les bus internes, périphériques et mémoires ne sont pas assez performantes pour suivre les vitesses actuelles des processeurs, le CHIPSET est le composant principal d'une carte mère. Depuis les Pentium, ils intègrent la majorité des interfaces (mémoire, bus, contrôleur disque dur et disquette, ports série et parallèle, USB, ...).

Dans une même famille de processeur, il est possible d'utiliser un chipset supérieur, en se méfiant des multiplicateurs de fréquence puisque généralement, les nouvelles cartes mères ne gèrent que les fréquences de base FSB supérieures. A l'inverse, une ancienne carte mère dans une famille de microprocesseurs ne gèrent pas souvent les processeurs les plus modernes, notamment au niveau des FSB du chipset. Si le FSB de la carte est compatible avec le FSB du processeur (mais facteur de multiplication non reconnu), un flashage du BIOS peut corriger le problème.

Chipset Processeur Pentium Intel						
Code	Fabricant	Processeurs	Mémoire	Norme IDE	Ports	Remarque

430 FX, <b>triton</b>	Chipset INTEL	Pentium	Simm 72c	2,1 GB Maximum	PCI v2.0 à 133 Mo/s	<b>mémoire cache L2</b> de 256k ou 512K de type insertion, incompatible avec les équivalents HX
430 HX, <b>triton III</b>	Chipset INTEL	Pentium	Simm 72c		bus série <b>USB</b> , PCI 2.0	quelques spécificités de partage de bus ISA / PCI. Cartes FX recevaient une <b>mémoire cache externe</b> de 256k ou 512K de type insertion, incompatible avec les équivalents HX
430 LX	Chipset INTEL	Spécifique Pentium 60 et 66 Mhz	Simm 72c			
430 MX	Chipset INTEL	Pentium portable		IDE		Pas de lecteur CD
430 NX	Chipset INTEL	Pentium	Simm 72c			Identique au 430LX, mais supporte jusqu'au Pentium 100
430 TX, <b>triton IV</b>	Chipset INTEL	Pentium	Auto- corrective <b>ECC</b> , jusqu'à 512 Mo Problème de gestion cache L2 si mémoire > à 64 MB – pas	Ultra-ATA	USB	optimisation de la gestion des mémoires SDRAM, supporte la présence d'un second processeur Peut cacher

			gérée en cache)			512 Mo à condition d'ajouter un composant (Tag Ram) à la CM.
<b>430 VX, Triton II</b>	Chipset INTEL	Pentium	Dimm SDRAM.et Simm 72c, 128 MB maximum			permet de partager avec le contrôleur graphique une partie de la mémoire vive
<b>Chipset Pentium II, Celeron et Pentium III</b>						
<b>440 BX</b>	Chipset INTEL	Pentium II, premiers Pentium III (<850 Mhz) et Celeron	3 Dimm 100, limitation par mémoire à 128 MB	UDMA-33	USB 1.1 AGP 2X	FSB 100 Mhz
<b>440 EX</b>	Chipset INTEL	Premiers CELERON		UDMA-33		Version bridée du 440 LX Fin 1998
<b>440 LX</b>	Chipset INTEL pour Pentium II et CELERON	Pentium II et CELERON	Dimm 66 et 100 (gérées en 66), limitation par mémoire à 128 MB (soit 512 MB maximum)	UDMA-33		FSB 66 Mhz
<b>440 ZX</b>	Chipset INTEL	Pentium II et CELERON	Identique au 440 BX mais seulement 2 Dimm	UDMA-33	USB 1.1 AGP 2X	Version allégée du 440 BX
<b>440 FX</b>	Chipset INTEL	Utilisé sur les premières cartes à base de Pentium II et CELERON	gestion de la mémoire ECC	UDMA-33		Equivalents au 430 HX (Pentium)

<b>810 (810i)</b>	Chipset INTEL	Celeron	DDR	Ultra DMA/66	AGP 4X mais bus non reconnu	i820 avec carte graphique i740 intégrée
<b>815 (815i)</b>	Chipset INTEL	Pentium III	SDRam PC100/PC133	UDMA 66	AGP 4X	09/2000 Remplace le i810 Equivalent au i820 en Dimm (sauf SMP)
<b>815E</b>	Chipset INTEL	Pentium III	SDRam PC100/PC133	UDMA 100	AGP 4X	815 + UDMA-100
<b>(Camino) 820</b>	Chipset INTEL	Pentium III de la série B (533B et 600B) et supérieur	Direct Rambus ( <b>DRDRam</b> ) En option le Memory <i>Translator Hub</i> (ou MTH) gère la SDRam PC 100	Ultra DMA/66	AGP 4X	12/1999 FSB 133 supporte le traitement biprocesseur en mode symétrique <a href="#">SMP</a>
<b>820E</b>	Chipset INTEL	Pentium III de la série B (533B et 600B) et supérieur	Direct Rambus ( <b>DRDRam</b> ) En option le Memory <i>Translator Hub</i> (ou MTH) gère la SDRam PC 100	Ultra DMA/100	AGP 4X slot CNR pour Modem 2 canaux USB 1.1 Ethernet 10/100 audio 6 canaux	06/2000 Version améliorée du i820
<b>ALLadin TNT2</b>	Chipset NVidia	Pentium III et Celeron				Carte graphique intégrée TNT2, concurrent du i810
<b>PRO Savage PM133</b>	Chipset VIA	Pentium II, Pentium III et Celeron VIA Cyrix III.	Dimm SDRAM 133		Contrôleur Ethernet 10/100	Carte écran intégrée S3 Savage 4 (3D) et S3 Savage 2000

						(2 D) Concurrent du i815
<b>apollo pro 133</b>	Chipset VIA	Pentium III	Dimm PC133	Ultra DMA/66	AGP 4X	11/1999
<b>Apollo vcm 133</b>	Chipset VIA	Pentium III	VCM133	Ultra DMA/66	AGP 4X	Déclinaison apollo KT133 pour AMD
<b>Chipset Pentium Pro</b>						
<b>450 GX</b>	Chipset INTEL	Pentium PRO	2 Bus mémoire, un pour chaque couple de processeur		2 bus PCI gère un bus EISA (486)	Gère 2 bus PCI et 2 Bus mémoire, un pour chaque couple de processeur.
<b>450 KX</b>	Chipset INTEL	Pentium PRO	2 Bus mémoire, un pour chaque couple de processeur			
<b>Chipset Intel XEON</b>						
<b>840</b>	Chipset INTEL multiprocess eur	XEON	2 canaux Direct Rambus <b>(DRDRam)</b> En option le Memory <i>Translator Hub</i> (ou MTH) gère la SDRam PC 100	Ultra DMA/66	AGP 4X	Identique au i820 mais multi- processeurs
<b>Chipset INTEL Pentium IV</b>						
<b>845</b>	Chipset INTEL	Pentium IV FSB 400, socket 478	Dimm PC100 et PC133	ATA-100	AGP 4X 4 ports USB 1.1 Eth. 10/100 Audio 6c.	

845D	Chipset INTEL	Pentium IV FSB400, socket 478	mémoire DDR 200 - 266	ATA-100	AGP 4X 4 ports USB 1.1 Eth. 10/100 Audio 6c.	modification i845
845E	Chipset INTEL	Pentium IV FSB 533- 400, socket 478	mémoire DDR 200 - 266	ATA-100	AGP 4X 6 ports USB 2.0 Eth. 10/100 Audio 6c.	845 D + USB 2.0
845G	Chipset INTEL	Pentium IV FSB 533- 400, socket 478	mémoire DDR 200 - 266	ATA-100	AGP 4X 6 ports USB 2.0 Eth. 10/100 Audio 6c.	Hyper- Treading Processeur graphique sur carte mère en option
845GL 845GV	Chipset INTEL	CELERON FSB 400 (GL) et FSB 533 (GV), socket 478	mémoire DDR 200 - 266	ATA-100	AGP 4X 6 ports USB 2.0 Eth. 10/100 Audio 6c.	Hyper- Treading Processeur graphique sur carte mère
845P	Chipset INTEL	Pentium IV FSB 533- 400, socket 478	mémoire DDR 200 - 266	ATA-100	AGP 4X 6 ports USB 2.0 Eth. 10/100 Audio 6c.	Hyper- Treading
<b>850</b>	Chipset INTEL	Pentium IV	Dual Channel RamBus		AGP 4X 2 canaux USB 1.1	Evolution du i820 bus processeur 100 MHz Quad Bumped
<b>865P</b>	Chipset Intel	Pentium IV, FSB 533- 400, socket 478	DDR ECC Dual Channel 333/400	UDMA-100	AGP 8X 8 ports USB 2.0 Audio AC97 Carte réseau 10/100	Hyper- Treading
<b>865PE</b>	Chipset Intel	Pentium IV, FSB 800- 533, socket 478	DDR ECC Dual Channel 333/400	UDMA-100 Raid en option	AGP 8X 8 ports USB 2.0 Audio AC97 Carte réseau	Hyper- Treading

					10/100	
<b>875</b>	Chipset Intel	Pentium IV, FSB 800-533, socket 478	DDR ECC Dual Channel 333/400 avec ECC	UDMA-100, RAID en option	AGP 8X 8 ports USB 2.0 Audio AC97 Carte réseau 1Gh	Hyper-Treading
<b>PA-X266</b>	chipset VIA	Pentium IV, socket 478	DDR ou SDRam PC100/PC133 max.4 GB		AGP 4X	06/2001
<b>SIS 645</b>	Chipset SIS	Pentium IV socket 478	DDR			
<b>Chipset Duron et Athlon</b>						
<b>750</b>	Chipset AMD	Athlon et Duron Slot et socket A	Dimm PC 100, max. 768 MB	UDMA-66	AGP 2 X	FSB 200
<b>760</b>	Chipset AMD	Athlon à partir 1 Ghz (double data rate) Socket A	DDR 200 PC1600, max. 4 GB	UDMA-100	AGP 4X	12/2000 FSB 200
<b>761</b>	Chipset AMD	Athlon à partir 1 Ghz (double data rate) Socket A	DDR 200 PC1600 et 266 PC210, max. 4 GB	UDMA-100	AGP 4X	FSB 266
<b>APPOLLO KX 133</b>	Chipset VIA	Athlon et Duron	SDRam PC 133 et PC100, max. 1,5 GB	UDMA-66	AGP 4X	Slot A et socket 370 (via adaptateur) bus EV6 200/266 Mhz
<b>761</b>	Chipset AMD	Athlon et Duron	DDR 266 (PC 2100)			Socket A
<b>APPOLO KT 133</b>	Chipset VIA	Athlon Tunderbird et Duron	SDRam PC 133 et PC100, max 1,5 GB	UDMA-66	Plus de bus ISA AGP 4X	Courant 2000 Socket 370 Identique au KX133
<b>KT 266</b>	Chipset VIA	Athlon et Duron	DDR 266			Slot A

<b>KT333</b>	Chipset VIA	Athlon et Duron FSB 200 et 266	200 - 266 DDR 333 (PC-2700)	UDMA-133	2 USB 1.1 6 c. audio port ACR modem AGP 4X pas de port ISA	06/2003 Socket A
KT400	Chipset VIA	Athlon XP et Duron, FSB 333	DDR 333 et DDR 400	UDMA-133	AGP 8X 6 USB 2.0 Audio	
KT600	Chipset VIA	Athlon XP		UDMA-133	AGP 8X 6 USB 2.0 Audio	10/2003
KT880	Chipset VIA	Athlon XP	Dual Channel	UDMA-133  2 X SATA	AGP 8X 6 USB 2.0 Audio	02/2004
<b>NFORCE 420</b>	Chipset nVidia	Athlon et Duron, socket A	Dual Channel PC2100 DDR 266	UDMA-100	AGP 4X 2 * USB 1.1	Carte graphique GeForce2 intégrée qui peut être remplacée par une carte externe
<b>Chipset Itanium</b>						
<b>460 GX</b>	Chipset Intel	Itanium				
<b>Athlons 64 bits</b>						
KT800	Chipset VIA	Opteron (single ou Dual processeur) et Athlon 64 bits	Dual Channel DDR 266- 333-400, max. 4 GB	2 X UDMA- 133, 2 SATA en Raid 0, 1 ou 0+1	AGP 8X 8 ports USB 2.0 audio carte réseau 10/100 (Giga en option)	06/2003

### L'horloge et la pile du CMOS

L'**horloge temps réel** (notée **RTC**, pour *Real Time Clock*) est un circuit chargé de la synchronisation des signaux du système. Elle est constituée d'un cristal qui, en vibrant, donne des impulsions (appelés *tops d'horloge*) afin de cadencer le système. On appelle *fréquence de l'horloge* (exprimée en *MHz*) le nombre de vibrations du cristal par seconde, c'est-à-dire le nombre de *tops d'horloge* émis par seconde. Plus la fréquence est élevée, plus le système peut traiter d'informations.

Lorsque l'ordinateur est mis hors tension, l'alimentation cesse de fournir du courant à la carte mère. Or, lorsque l'ordinateur est rebranché, le système est toujours à l'heure. Un circuit électronique, appelé *CMOS* (*Complementary Metal-Oxide Semiconductor*, parfois appelé *BIOS CMOS*), conserve en effet certaines informations sur le système, telles que l'heure, la date système et quelques paramètres essentiels du système.

Le CMOS est continuellement alimenté par une pile (au format *pile bouton*) ou une batterie située sur la carte mère. Ainsi, les informations sur le matériel installé dans l'ordinateur (comme par exemple le nombre de pistes, de secteurs de chaque disque dur) sont conservées dans le CMOS. Dans la mesure où le CMOS est une mémoire lente, certains systèmes recopient parfois le contenu du CMOS dans la RAM (mémoire rapide), le terme de « *memory shadow* » est employé pour décrire ce processus de copie en mémoire vive.

Le « *complementary metal-oxide semiconductor* », est une technologie de fabrication de transistors, précédée de bien d'autres, telles que la *TTL* (« *Transistor-transistor-logique* »), la *TTLs* (*TTL Schottky*) (plus rapide), ou encore le *NMOS* (canal négatif) et le *PMOS* (canal positif).

Le CMOS a permis de mettre des canaux complémentaires sur une même puce. Par rapport à la *TTL* ou *TTLs*, le CMOS est beaucoup moins rapide, mais a consommé en revanche infiniment moins d'énergie, d'où son emploi dans les horloges d'ordinateurs, qui sont alimentées par des piles. Le terme de CMOS est parfois utilisé à tort pour désigner l'horloge des ordinateurs.

Lorsque l'heure du système est régulièrement réinitialisée, ou que l'horloge prend du retard, il suffit généralement d'en changer la pile !

## Le BIOS

Le BIOS (*Basic Input/Output System*) est le programme basique servant d'interface entre le système d'exploitation et la carte mère. Le BIOS est stocké dans une *ROM* (mémoire morte, c'est-à-dire une mémoire en lecture seule), ainsi il utilise les données contenues dans le *CMOS* pour connaître la configuration matérielle du système.

Il est possible de configurer le BIOS grâce à une interface (nommée *BIOS setup*, traduisez *configuration du BIOS*) accessible au démarrage de l'ordinateur par simple pression d'une touche (généralement la touche *Suppr.* En réalité le setup du BIOS sert uniquement d'interface pour la configuration, les données sont stockées dans le *CMOS*. Pour plus d'informations n'hésitez pas à vous reporter au manuel de votre carte mère).

## Le support de processeur

Le processeur (aussi appelé *microprocesseur*) est le cerveau de l'ordinateur. Il exécute les instructions des programmes grâce à un jeu d'instructions. Le processeur est caractérisé par sa fréquence, c'est-à-dire la cadence à laquelle il exécute les instructions. Ainsi, un processeur cadencé à 800 MHz effectuera grossièrement 800 millions d'opérations par seconde.

La carte mère possède un emplacement (parfois plusieurs dans le cas de cartes mères multi-processeurs) pour accueillir le processeur, appelé **support de processeur**. On distingue deux catégories de supports :

- **Slot** (en français *fente*) : il s'agit d'un connecteur rectangulaire dans lequel on enfiche le processeur verticalement

- **Socket** (en français *embase*) : il s'agit d'un connecteur carré possédant un grand nombre de petits connecteurs sur lequel le processeur vient directement s'enficher

Au sein de ces deux grandes familles, il existe des version différentes du support, selon le type de processeur. Il est essentiel, quel que soit le support, de brancher délicatement le processeur afin de ne tordre aucune de ses broches (il en compte plusieurs centaines). Afin de faciliter son insertion, un support appelé **ZIF** (*Zero Insertion Force*, traduisez *force d'insertion nulle*) a été créé. Les supports ZIF possèdent une petite manette, qui, lorsqu'elle est levée, permet l'insertion du processeur sans aucune pression et, lorsqu'elle est rabaisée, maintient le processeur sur son support.

Le processeur possède généralement un détrompeur, matérialisé par un coin tronqué ou une marque de couleur, devant être aligné avec la marque correspondante sur le support.



Dans la mesure où le processeur rayonne thermiquement, il est nécessaire d'en dissiper la chaleur pour éviter que ses circuits ne fondent. C'est la raison pour laquelle il est généralement surmonté d'un **dissipateur thermique** (appelé parfois *refroidisseur* ou *radiateur*), composé d'un métal ayant une bonne conduction thermique (cuivre ou aluminium), chargé d'augmenter la surface d'échange thermique du microprocesseur. Le dissipateur thermique comporte une base en contact avec le processeur et des ailettes afin d'augmenter la surface d'échange thermique. Un ventilateur accompagne généralement le dissipateur pour améliorer la circulation de l'air autour du dissipateur et améliorer l'échange de chaleur. Le terme « **ventirad** » est ainsi parfois utilisé pour désigner l'ensemble *Ventilateur + Radiateur*. C'est le ventilateur du boîtier qui est chargé d'extraire l'air chaud du boîtier et permettre à l'air frais provenant de l'extérieur d'y entrer.

### Type de socket

Il existe de nombreux slots et sockets pour les UC (unités centrales), les références des slots correspondant aux types de processeurs et à leurs fabricants. Les sockets les plus récents, reconnaissables à leur numéro à trois chiffres, tirent leur nom de leur nombre de broches, tandis que les plus anciens sont nommés dans l'ordre de leur invention ou de leur génération et leur nom ne comporte en général qu'un seul chiffre.

### Socket prenant en charge les CPU d'intel

- Socket 1 - 80486SX, 80486DX, 80486DX2, 80486DX4
- Socket 2 - 80486SX, 80486DX, 80486DX2, 80486DX4
- Socket 3 - 80486SX, 80486DX, 80486DX2, 80486DX4
- Socket 4 et 5 - premiers processeurs Intel Pentium
- Socket 6 - 80486DX4
- Socket 7 - Intel Pentium et Pentium MMX
- Socket 8 - Intel Pentium Pro
- Slot 1 -processeurs Intel Pentium II, premiers Pentium III et Celeron (233 MHz - 1, 13 GHz)
- Slot 2 - processeurs Intel Xeon à base de noyaux Pentium II/III
- Socket 370 -processeurs Pentium III et Celeron plus récents (800 MHz - 1, 4 GHz)
- Socket 423 - processeurs Pentium 4 et Celeron (à base de noyau Willamette)
- Socket 478 - processeurs Intel Pentium 4 et Celeron
- Socket 603/604 - processeurs Intel Xeon
- Socket 755 - processeurs Intel Pentium 4 et Celeron

### Socketes prenant en charge les CPU d'AMD

- Socket 7 - processeurs AMD 80486, K5 et K6
- SuperSocket7 - processeurs AMD K6, K6-2 et K6-3
- Slot A - processeurs AMD Athlon et Duron plus anciens
- Socket 462 (ou Socket A) -processeurs AMD Athlon, Athlon XP et Duron plus récents
- Socket 754 - AMD Athlon 64 d'entrée de gamme avec support mémoire à canal unique
- Socket 939 - AMD Athlon 64 et AMD Athlon FX avec support mémoire à double canal
- Socket 940 - processeurs AMD Opteron et premiers AMD Athlon FX

### Les microprocesseurs pour PC

Ce chapitre est découpé en 4 parties:

- les microprocesseurs PC pour ordinateurs bureautiques: historique et processeurs actuels, leur technologie.
- les processeurs spécifiques pour portables
- les microprocesseurs serveurs (Pentium Pro, Xeon, Itanium et Opteron)
- le Dual-Core.

Le **processeur (microprocesseur)** est le composant hardware le plus connu d'un ordinateur. C'est l'unité de traitement des informations. De lui-même, il est incapable d'exécuter une action, quelle qu'elle soit. Son travail se limite à lire des programmes (des suites d'instruction en langage assembleur), à les décoder et à les exécuter. Il ne prend donc aucune décision, se contentant d'exécuter "bêtement" ce qu'on lui demande. C'est le programme, par des instructions conditionnelles, qui se charge de "l'intelligence" des ordinateurs. A son avantage, il exécute ces tâches sans erreur et très rapidement. Pour une introduction aux circuits électroniques à base de microprocesseur

Le premier microprocesseur est apparu en 1972 avec le 4004 d'Intel, destiné à un constructeur japonais de machine à calculer. Celui-ci le refuse pour une taille (dimension) excessive. INTEL le met alors sur le marché sans conviction mais avec le succès que l'on sait. Les années 80 voyaient l'émergence de ces circuits avec les Z80 (compatible au niveau instructions avec le 8080), 6800 de Motorola (dont les suivants sont utilisés encore par les MAC) et le 6500. Avec l'arrivée

des XT d'IBM et l'utilisation du 8088 (8086 pour les clones), INTEL devenait maître du marché fin des années 80.

Les processeurs de la famille INTEL compatible ont évolués. Avant d'examiner tous les tours de passe que les concepteurs de microprocesseurs utilisent actuellement pour améliorer les performances (structure des microprocesseurs), un petit historique des modèles utilisés dans les ordinateurs PC et de leurs caractéristiques.

### Le 8088 et 8086.

Premier microprocesseur de la famille PC (Personal computer), le 8088 (utilisé par IBM concepteur du PC) utilise un bus de donnée interne de 16 bits, mais externe de 8 bits. Le 8086 (utilisé par les copies du PC XT), totalement compatible au niveau des instructions, possédait un bus de données complètement sur 16 bits.

Le PC est sorti en 1980 sous la dénomination de XT. De part sa conception de départ (et malgré son prix élevé), le XT avait de multiples avantages sur la concurrence.

1. **Bus périphérique** (des connecteurs pour implanter les cartes) de type 8 bit ISA. C'est ce que l'on appelle un système ouvert.
2. **Capacité mémoire extensive** par socket. Un commodore 64 incluait d'office 64 MB de mémoires. Le XT pouvait en accueillir 640 MB, malgré qu'à la livraison, la capacité était inférieure.
3. Le **système d'exploitation** pas en mémoire ROM, mais bien implantée sur une disquette. Et voici les réels débuts de Microsoft au niveau systèmes d'exploitation avec le DOS que nous verrons en systèmes d'exploitation.
4. Une large partie des adresses et interruptions dans le PC sont normalisées, y compris pour des périphériques futurs. Avec l'implantation du bus ISA, une plage d'adresse est réservée pour des cartes périphériques, une autre pour la mémoire. Ceci permet de changer de carte écran sans remplacer de PC.

Comme le schéma de la carte mère était à base de circuits TTL très courants (74LS), les copies de cet ordinateur étaient faciles, c'est ce qu'on appelait les clones et actuellement: produits blancs, OEM, PC assemblés, ...

### Le 286.

IBM sort 2 ans plus tard l'AT à base du microprocesseur 286 d'INTEL. De performances supérieures (de 8 à 16 Mhz), le bus de données est totalement 16 bits. Apparaît le bus ISA 16 bits, celui que nous connaissons maintenant.

IBM implante quelques nouveautés comme une horloge (RTC - Real Time Clock), un BIOS sur EPROM et le Setup, sauvegardés par une batterie. Le bios est le firmware du PC et sert d'interface entre l'électronique du PC et le système d'exploitation. Sauf quelques systèmes d'exploitation UNIX du début des années 80, tous passent par ce BIOS pour récupérer les informations.

La mémoire est également modifiée. Elle n'est plus composée de petits composants que l'on rajoute, mais par un banc de mémoire couvrant l'ensemble de la plage mémoire basse de 1 MB, également les zones mémoires réservées dite mémoire haute.

## Le 386.

1991, c'est ici que commencent les vrais bricolages pour améliorer la vitesse de transfert et les premiers balbutiements d'INTEL dans le marketing, sous la concurrence d'un petit nouveau du nom d'AMD qui fabriquait les Intel sous licence.

INTEL sort tout d'abord le 386DX à 25 et 33 Mhz. Son bus de données est de 32 bits, en interne et en externe. Comme son prix est élevé et les mémoires chères, Intel sort une version réduite du processeur, le 386SX avec un bricolage identique au 8086, bus de donnée interne de 32 bits, mais externe de 16 bits. AMD sort son premier processeur 386DX avec un peu de retard, cadencé à une vitesse de 40 Mhz, totalement compatible avec le 386 d'INTEL.

Les cartes mères avec une mémoire **cache externe L2** font leur apparition. Cette mémoire tampon plus rapide que la mémoire normale stocke les lignes de programmes et les données les plus utilisées. Ceci augmente les performances du PC

Ce processeur utilise 3 **modes de fonctionnement**

- **Mode Réel** (Real): le processeur travaille comme un simple 8088, méthode identique au 286.
- **Mode protégé** (protected): le microprocesseur peut utiliser toutes les possibilités des 80286, plus les instructions spécifiques du 386 avec un adressage mémoire jusqu'à 4 GB
- **Mode virtuel** (virtual), émule plusieurs sessions de 8086, utilisé à partir de Windows 95.

Intel sort également un processeur spécialisé pour les 80386, le coprocesseur mathématique 80387. Ce composant est spécialisé pour le calcul réel (en virgule flottante). Ceci augmente les performances pour les jeux, dessin technique, ...

## Le 486.

INTEL sort le 486 de type DX avec un socket 1 le premier. Le coprocesseur mathématique est maintenant implanté dans le microprocesseur. Ce circuit est entièrement dédié aux calculs en virgule flottante (nombre réel). Le cache L2 externe plafonne à 256K. La vitesse débute à 33 Mhz et va atteindre 50 Mhz. Pour réduire les prix, INTEL sort le 486SX: identique au 486DX, mais sans co-processeur mathématique intégré.

Lorsque AMD arrive avec un 486 à 40 Mhz, INTEL sort le premier processeur multiplicateur, le 486DX2-66. Attention que ces processeurs utilisent un multiplicateur interne (à la différence de tous les processeurs suivants). Sauf pour les tensions d'alimentation éventuellement et types (marque), vous pouvez directement remplacer un 486DX33 par un 486DX2-66. Pour un 486DX2-66, la vitesse externe reste à 33 Mhz. Ceci sera différent pour les 486DX4-100 où l'on pouvait utiliser 33 X3, 40 X 2,5 ou 50 X 2. En DX4-100, le choix se fait obligatoirement sur la carte mère. Attention, un DX4-100 d'Intel ne peut pas être remplacé directement par un DX4-100 de Cyrix ou d'AMD.

Depuis les 486DX2, les processeurs utilisent donc 2 vitesses, une vitesse interne (sa vitesse de référence) et une vitesse externe pour les bus, appelée **FSB (Front Side Bus)**. La tension d'alimentation n'est plus de 5 V, mais passe à 3,3V.

Les modèles sont DX2-50, DX2-66, DX4-80 et DX4-100.

Cyrix sort son premier processeur. Avec le 486, INTEL prend le dessus au niveau du co-processeur mathématique et donc des performances

Avec l'apparition des **486DX2**, tous les microprocesseurs modernes doivent être montés avec ventilateur. Les ordinateurs de marque utilisaient néanmoins des simples radiateurs jusqu'aux Pentium à 120 Mhz.

Le bus VLB fait son apparition, c'est une extension de quelques broches du processeur 486 sur un connecteur ISA, incompatible avec les Pentium et autres.

## Les Pentium

INTEL sort le Pentium le premier (le nom vient d'une protection commerciale du terme). Sa principale caractéristique par rapport à ses prédécesseur est l'implantation d'une mémoire cache interne appelée **L1** de 8 KB pour les programmes et 8 KB pour les données directement implantée dans le processeur.

Suit directement en juin 1994 la sortie d'un microprocesseur révolutionnaire par son architecture interne, le NexGen 586. Ses instructions sont totalement compatibles avec celles du Pentium, ses performances plus rapides mais pas le socket (brochage). Les cartes mères sont difficiles à trouver et finalement la firme disparaît, rachetée par AMD qui appellera son premier "Pentium" **AM5X86**. Il est équivalent en performance mais utilise des cartes 486. A l'époque, les cartes 486 utilisaient des bus VLB, incompatible avec ce processeur. AMD poursuivra avec le K5 compatible broche à broche (et donc carte mère) avec le Pentium, suivi du K6 et Cyrix le 586. Le K5 n'était pas au niveau des performances des Pentium et sera vite remplacé par le K6, développé sur base du Nexgen 586 (firme rachetée par AMD).

Premièrement, INTEL sort les Pentium 60 et 66 Mhz. Ceux-ci utilisent un socket spécial, les suivant utiliseront le socket de type 7.

De nouveau, il va falloir utiliser des facteurs de multiplication entre la vitesse de travail interne et celle externe (FSB). Au contraire des 486DX2-66, le multiplicateur doit être signalé sur la carte mère.

La fréquence externe est liée à la vitesse des DIMM de l'époque (66 Mhz maximum). En cas de choix, plus la vitesse externe est rapide, plus le PC sera rapide. Par exemple, pour un Pentium 100: un PC configuré en 1,5 X 66 sera plus rapide qu'en 2 X 50.

## Les Pentium MMX, K-6, K6-2, K6- Cyrix 6X86

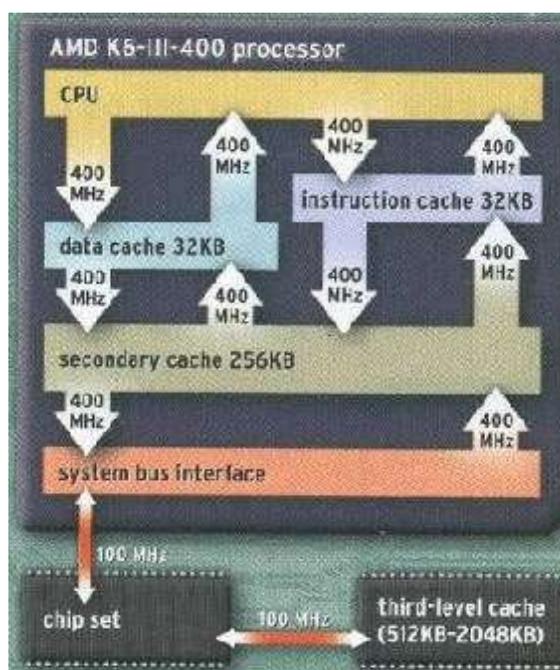
En mai 1997, INTEL rajoute des instructions supplémentaires dans son microprocesseur Pentium et l'appelle le Pentium MMX. Elles sont notamment dédiées à la compression, même si elles sont répertoriées dans les publicités comme "MULTIMEDIA", terme à la mode à l'époque. En même temps, le Pentium comportait un cache L1 de 16 K, le Pentium MMX de 32 K (moitié données, moitié programmes) AMD rajoute les mêmes instructions dans son K-6. Les fréquences des Pentium MMX sont 166 Mhz, 200 et 233 Mhz

AMD continue avec le K6-2 qui inclut 21 autres nouvelles instructions (3Dnow, implanté dans directX5.0 de Microsoft) pour concurrencer le Pentium II et, en août 1999, le K6-3 qui intègre un cache L3. Le K6-3 intègre un cache L1 de 64K et un cache L2 intégré de 256K à la

fréquence du processeur. Comme pour tous les processeurs sockets 7, le cache L2 est néanmoins toujours intégré sur la carte mère.

Vitesse	FSB	multiplicateur
233	66	3,5
266	66	4
300 (K6-2)	66	4,5
333 (K6-2)	66	5
350 (K6-2)	100	3,5
400 (K6-2)	100	4
450 (K6-2 et K6-3)	100	4,5

Les Pentium MMX, K6 (équivalents aux MMX), K6-2, K6-3 et MII de Cyrix utilisent toujours le socket 7.



### Au coeur du K6-III

Le K6-III d'AMD dispose d'un cache L2 de 256 KB intégré qui travaille à la vitesse du processeur (On Die), soit 400 ou 450 MHz. Dans les Pentium II et III de l'époque, le cache secondaire L2 ne fonctionnent qu'à la moitié de la vitesse du processeur. Le K6-III comportent également un cache **L3**, créé en enfichant ce processeur sur une carte mère pour K6 II (compatible broche à broche) disposant de 512, 1024 ou 2048 KB de cache L2

### Les Pentium II – CELERON et PENTIUM III.

Pendant qu'AMD prépare son K6-2, INTEL sort le PENTIUM II:

- le cache L2 n'est plus implanté sur la carte mère mais directement sur le boîtier du processeur et travaille à la moitié de la fréquence interne du processeur
- il utilise un nouveau connecteur pour s'insérer sur une carte mère, le **slot one**. Ce socket sera remplacé mi-2000 par les 370 (Celeron) et FC-PGA (Pentium III). Différents adaptateurs existent pour passer d'un slot one à un socket 370 PPGA / FC-PGA avec des incompatibilités si les marques sont différentes. Le FC-PGA est identique au 370, sauf quelques broches en plus.
- son architecture est totalement RISC

Viennent d'abord les Pentium II 233, 266 et 300 Mhz, FSB de 66 Mhz. Le chipset est le 440LX (au départ 440FX). Il gère les mémoires Dimm à 66 Mhz, le bus écran AGP, interface disque dur Ultra-ATA à 33 MB/s et bus externe USB.

Comme INTEL décide d'abandonner les PENTIUM MMX et que les Pentium II sont trop chères, INTEL sort le **CELERON** pour les machines de bas de gamme. Ce microprocesseur n'inclut au départ pas de mémoire cache L2. Dans les derniers sortis, il est de 128K, mais est cadencé à la même vitesse que le processeur. INTEL a sorti un chipset 440EX (fin98) et 440ZX qui gère moins de slot pour les bus PCI (3) et ISA (1). Le CELERON peut être couplé par 2 (même vitesse et si possible, même lot de fabrication).

L'étape suivante est le Pentium II à 333 (mars 1998), 350, 400, 450, 500 Mhz, ... avec un FSB de 100 Mhz. Ceux-ci utilisent le chipset 440BX et la mémoire Dimm 100Mhz (VIA sort l'Apollo P6 qui gère en plus la norme ultra DMA/66 pour 33 au 440BX).

Sorti début 1999, les PENTIUM III (Katmai) sont identiques au Pentium II mais intègrent des instructions multimédia supplémentaires, appelée **SSE**. Ils utilisent toujours le slot 1.

### Microprocesseur INTEL Pentium III Copermine

Avec la sortie fin 1999 du chipset CAMINO 820i, les PENTIUM III coppermine sont gravés en 0,18 microns, utilisent les mémoires Dimm 133 (via une interface) et les DRDRAM à 300 (PC600, 1,6GB/s) et 400 Mhz (PC800). Avec l'apollo Pro 133, VIA gère les SDRAM 133 Mhz, les performances augmentent de 4 à 7 % par rapport à un circuit INTEL 440 BX selon le type de mémoire (existe en plus 2 types de mémoires 133, les PC133 et les VCM133 qui VCM133 gèrent mieux les temps d'attente et sont plus rapides).

Pour parfaire la facilité de l'acheteur, INTEL sort en même temps 4 versions du même microprocesseur, selon la taille de la gravure et la vitesse du bus extérieur. Suivant la lettre accolée, vous trouverez les caractéristiques du processeur. Voici par exemple les caractéristiques d'un Pentium 600 Mhz.

Bus 100 Mhz, gravure 0,25 micron (actuel)	600
Bus 133 Mhz, gravure 0,25	600B
Bus 100 Mhz, gravure 0,18 micro	600E
Bus 133 Mhz, gravure 0,18 micron	600EB

Toutes les versions n'existent pas. Les PENTIUM III gravés en 0,18 microns (COPPERMINE) exploitent une mémoire L2 de 256K mais à la même vitesse que le processeur, pour 512k, mais à la moitié de la vitesse, pour ceux gravés en 0,25 microns (anciens). Ceci donne

un taux de transfert de **9,6 GB** par seconde pour un PE III E (ou EB) en 256 bits pour **2,4 GB/s** en Pentium III 600 normal qui fonctionne en 128 bits avec la moitié de la fréquence. Les copermine à 133 Mhz sont interfaçables par le i820 d'INTEL. Celui-ci n'accepte pas les Dimm 133 Mhz, mais les RamBus nettement plus chères dans les versions FSB 133.

Le dernier core est le Tualatin, développé en version station de travail (1 à 1,33 Ghz) avec cache L2 On died de 256 KB et la version Pentium III S (serveur, avec 512 KB de cache L2), seule version de Pentium III officiellement multi-processeur SMP.

### L'Athlon, Duron et Thunderbird d'AMD

Sorti en mai 1999, ce microprocesseur est le premier de la génération 7. L'Athlon prend pour la première fois la tête vis à vis des Pentium III en performance. Il intègre 22 millions de transistors contre 9,5 millions pour un Pentium III Katmai.

Quelles sont les caractéristiques de l'Athlon par rapport au Pentium III et au précédents AMD: tout **d'abord 3 unités de calcul** en virgule flottante, pour 2 pour les Pentium II et III (1 seule pour les anciens AMD). Si l'unité installée dans les Pentium a toujours été supérieure à celle des K6, les 3 unités des Athlons sont chacune du niveau des Pentium III. De plus, les 2 modules des Pentium III ne peuvent travailler simultanément, les 3 unités de l'Athlon, oui!

La deuxième différence vient des caches intégrés au processeur. Le cache L1 est de 32 k pour les Pentium III, contre 128K pour l'Athlon. Le cache L2 de 512k pour les Pentium II et III (128K pour les Celeron) débute à 512k pour aller jusqu'à 8 GB pour l'ATHLON.

Une autre différence est la vitesse du bus externe. Si les Pentium III classiques sont limités à 100 Mhz en bus externe (133 pour les coppermine), l'Athlon utilise la technologie Alpha EV6 de Compaq (utilisée par les processeurs Alpha) pour utiliser des bus externes à 200 Mhz (flancs montants et descendants sur 100 Mhz) , mais ne gère que des mémoires PC133 (mémoires Dimm). C'est le passage aux mémoires DDR (double data Rate) qui permet réellement aux Athlons de tourner à plein régime. Les premiers Athlon utilisent le Slot A, d'apparence similaire au Slot 1 INTEL mais les signaux (et donc les cartes mères) sont différents. Le **Slot A** (AMD) et le Slot One ne sont pas compatibles.

### Evolution

Mi-2000, AMD sort 2 versions de l'Athlon amélioré en socket A (socket 462): le Duron et le Thunderbird.

Le **Duron** intègre un cache L1 de 128 K et un cache L2 de 64 K à la même vitesse que le processeur. Le **Thunderbird** dispose d'un cache L1 de 128K, mais un cache L2 de 256K à la même vitesse que le processeur (on died). Ces processeurs utilisent un bus externe cadencé à 100 Mhz DDR (Double Data Rate, X2) qui correspond dans la pratique à 200 Mhz. Le modèle C de l'Athlon (vitesse de 1 à 1,4 Ghz) utilise un FSB de 133 Mhz

Les Athlon seront modifiés courant 2001. Le Thunderbird est remplacé par le XP(nom de Core Thoroughbred). L'architecture interne est différente, garantissant moins d'échauffement avec une protection contre la surchauffe. Le cache L2 reste à 256 KB avec 52 nouvelles instructions complémentaires appelées 3D Now Professional. En 2002, une version MP (version biprocesseur) voit également le jour. Les XP possèdent néanmoins une autre caractéristique comme un FSB de 133 Mhz. Nous verrons que la vitesse n'est plus la seule mesure

des performances. Depuis les XP, la vitesse effective des processeurs n'est pas celle sur la quelle ils sont vendus mais celle équivalente des processeurs Intel.

Au troisième trimestre 2002, la vitesse externe est une fois de plus augmentée pour passer à 166 MHz avec les Athlon XP 2,8 Ghz (cadencé en fait à 2250 Mhz). Cette montée provoque quelques problèmes de compatibilité avec les cartes mères "133" mais améliore les performances.

Janvier 2003, AMD décline l'Athlon sous le Core BARTON. Ce nouveau microprocesseur débutant à 2500 + (1833 Mhz). Le cache L2 passe de 256 à 512K. Les performances à vitesse équivalente sont forcément supérieure à celle des Thoroughbred.

Les indications de FSB reprennent la vitesse réelle du bus. Les Athlon utilisent les flancs montants et descendants de l'horloge. Par conséquent, un FSB de 133 tourne en fait à 266, un FSB de 166 à 333 et un FSB de 200 à 400. Les vitesses externes effectives sont gravées sur l'étiquette sous forme de lettre: C pour 266, D pour 333 et E pour 400.

Les Duron culminant à 1,3 Gh ont disparus début 2003. Le Sempron remplace les Athlon XP en janvier 2004. Pour les premiers modèles, seule la structure change, ils sont identiques aux Athlons XP.

### Microprocesseur Pentium IV (fin 2000)

Alors que les Pentium II, Pro, Celeron, Xenon et Pentium III utilisent la même architecture interne, INTEL sort un tout nouveau processeur appelé PENTIUM IV basée sur une architecture appelée NetBurst. L'architecture Netburst inclut différents particularités comme:



- un pipeline sur 20 niveaux contre 10 pour les Pentium III et Athlon de l'époque. Les derniers modèles montent jusque 31.
- la possibilité d'exécuter des instructions dans le désordre, notamment si les données ne sont pas dans le cache
- Le cache L1 d'instruction reçoit dorénavant jusque 12.000 instructions prédécodées en RISC contrairement aux modèles standards qui conservent des instructions CISC compatibles 8088.

### Caractéristiques.

- 42 millions de transistors, avec une architecture qui reste en 32 bits.
- Nouveau socket, de type PGA 423 (remplacé mi-2001 par le  $\mu$ PGA 478).
- Nouvelles instructions **SSE2** exploitées par DirectX 8.0, 144 instructions
- Unité de calcul modifiée (2 ALU tournant au double de la vitesse interne du processeur et une PGU "Virgule flottante").
- La mémoire **cache L2** reste à 256 K (portée à 512K début 2002) mais est amélioré. En effet, on passe là encore d'une bande passante de 14.9 GB /s pour un PIII 1 GHz à 41.7 GB /s pour un P4 1.4 GHz.
- Le **cache L1** ne contient plus qu'un cache données de 8 K et une "Instruction Trace Cache" qui stocke les instructions après leur décodage en RISC. Ce cache programme peut contenir jusque 12.000 instructions, ce qui n'est pas sans conséquences en cas de mauvaise prédiction de branchement.
- La fréquence de bus (externe) est de 200 Mhz, mais passera à 400 début 2002.
- La gravure est de 0,18  $\mu$ , passera à 0,15  $\mu$  début 2002.

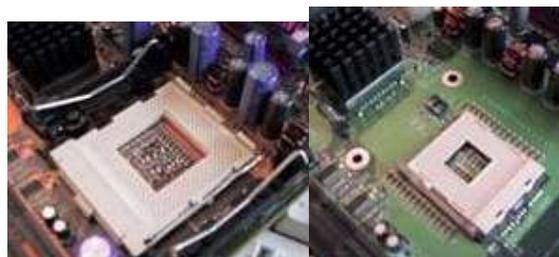
Le SSE, ou Streaming Simd Extension, est introduit en mars 1999 avec le Pentium III. C'est un jeu de 70 instructions utilisant la technologie SIMD (Single Instruction, Multiple Data), tout comme le MMX ou le 3D Now!. Le principe du SIMD est simple : traiter d'une traite plusieurs données avec une seule instruction. Ainsi, alors qu'avec une instruction SISD (Single Instruction, Single Data) de type x86 par exemple une instruction équivaut à un résultat, on peut obtenir jusqu'à 4 résultats avec une instruction SSE. Ainsi, en un cycle d'horloge on ne pourra faire que  $x + y$  en x86, mais on pourra faire  $x_1 + y_1$ ,  $x_2 + y_2$ ,  $x_3 + y_3$  et  $x_4 + y_4$  en SSE.

Le SSE 2 apporte des améliorations à deux niveaux. Tout d'abord, il rajoute 144 nouvelles instructions à celles existantes, ces dernières portant principalement sur la gestion de la mémoire et du cache. De plus, le SSE 2 permet de manipuler de nouveaux types de données, tel que les nombres entiers de 128 bits (1 par cycle) et les flottants double précision de 64 bits (2 par cycle). Tout comme le MMX et SSE, le SSE 2 nécessite que les applications soient programmées pour en tirer partie, ce qui n'est pas le cas pour une majorité d'entre elles.

L'architecture NetBurst est également capable d'exécuter les instructions dans le désordre (out of order execution). Ainsi, si par exemple la première ALU traite le calcul  $A = 5 \times 6$  et que le calcul suivant est  $B = A + 2$ ; la seconde ALU ne peut rien faire dans le cas d'une exécution in order (dans l'ordre), sauf attendre le résultat du calcul A. Avec le système out of order, la deuxième ALU peut sauter ce calcul pour passer à un suivant, tous les résultats étant bien sûr remis dans l'ordre à la fin du traitement.



Remarquez le support sur les cotés du processeurs pour le ventilateur de "forte taille" puisqu'il pèse dans les 450 gr.



A gauche le Socket 423, à droite le Socket 478

## Evolution

- **Mi 2001**, INTEL modifie le brochage des Pentium IV, le **PGA 423 Pin** est remplacé par un nouveau socket, le **µPGA 478 Pin**. Ceci ne modifie en rien l'architecture interne.
- **Début 2002**, le **Northwood** est la seconde version du Pentium 4 (P4N). La première version portait comme nom de code Willamette (P4W). Ses caractéristiques sont les mêmes que le P4W, si ce n'est qu'elle utilise une gravure en  $0,13\mu\text{m}$  et possède 512Ko de

cache L2. Elle n'est disponible qu'au format  $\mu$ PGA socket 478 Cette version apparaît à partir des 2,2 Ghz.

- **Début novembre 2002**, Intel annonce un changement de tension d'alimentation pour ces processeurs basés sur le core C-1 (0.13 $\mu$ ), qu'ils soient de type Pentium 4 ou Celeron. Au lieu d'une unique tension 1,525V, ces processeurs pourront fonctionner en 1,475, 1,5 ou 1,525V. Du coup, ces processeurs ne disposeront d'aucun marquage concernant leur tension d'alimentation.
- **Début 2003**, le Pentium IV à 3,06 Ghz inclus l'hypertreading qui émule deux processeurs en partageant les instructions à exécuter de manière interne entre les différentes parties dédiées au traitement.
- **Mi-2004**, le socket 775, un format FPGA, remplace le 478 (un connecteur ZIF). C'est le socket qui inclut les pin dorénavant. Intel en profite également pour changer les notations de ses processeurs.
- **En 2005**, une version Dual-core, notés **Pentium IV D** est également sortie. Contrairement au Pentium IV EE, il ne gère pas l'hypertreading

### Notations des Pentium IV

- **Série 3xx** : Celeron D (bus externe de 133 Mhz -533)
- **Série 5xx** : Pentium 4 avec cache L2 de 1 MB (anciennes versions E), de 2,66 à 3,8 GHz
- **série 6xx** : Pentium 4 avec 2 Mo de cache de niveau 2, de 3,0 à 3,8 GHz
- **série 8xx** : Pentium D (Dual Core) avec 2 x 1 MB de cache de niveau 2, de 2,66 à 3,2 GHz
- **série 9xx** : Pentium EE dual core avec 2 X 2 MB de cache de niveau 2, de 2,8 à 3,73 GHz, FSB800 ou 1066.

En 2006, l'architecture est remplacée par l'Intel Core, moins gourmand en énergie, mais surtout permettant de monter en fréquence. La taille du pipeline est également diminuée, diminuant le temps de latence lors des transferts mémoires.

### Pentium IV EE

INTEL ne développe pas de processeurs bureautiques en 64 bits (comme l'Athlon 64 bits). Les Pentium IV, même utilisant l'hypertreading ne peuvent résister aux performances des Athlons 64 bits (même en Windows 32 bits). INTEL a donc sorti une version améliorée du Pentium IV fin 2003, le Pentium IV Extreme Edition (désigné sous le core Prescott). Ses performances sont néanmoins inférieures à celles des Athlons 64 bits d'AMD. Ce processeur utilisant un FSB de 800 Mhz est clairement dédié aux joueurs. Utilisant un socket 478, il est donc compatible avec les cartes mères Pentium IV normales FSB800 (même si des incompatibilités sont toujours possibles).

Pour améliorer les performances, INTEL a repris les vieilles recettes. Le cache L1 dédié aux données est doublé (16 K pour 8 K pour les Pentium IV). Le cache L2 est également amélioré puisqu'il passe de 512 kB à 1024. Le nombre de transistors passe à 125 millions pour une gravure en 0,09 $\mu$ .

Ce n'est pas la seule évolution de cette architecture. Le pipeline pour le calcul d'entiers passe de 20 à 31 étages. L'augmentation de la longueur du pipeline a quelques défauts. Comme nous le verrons dans l'architecture des processeurs, lors d'instructions conditionnelles, si le

pipeline est chargées d'instructions A, et que l'instruction conditionnelle du programme demande la suite d'instruction B, le pipeline doit être vidé avant de recharger la nouvelle suite d'instructions. INTEL s'est donc attelé à améliorer au sein du Prescott cette prédiction de branchement.

Pour compléter le tableau, le Pentium IV EE intègre le **SSE 3**, 13 nouvelles instructions. Comme toutes les instructions multimédia, ceci nécessite l'utilisation de programmes recompilés en fonction de ces nouvelles instructions.

Le Pentium IV EE (Extreme Edition) est le haut de gamme des processeurs Intel bureautique. Quasiment toutes les avancées technologiques des serveurs Intel (à part le 64 bits) sont reprises dans ce microprocesseur au gré des avancées technologiques. Actuellement (07/2006), deux versions sont disponibles:

- la version standard avec cache de 2 MB avec une vitesse externe de 1066 avec une vitesse maximum de 3,73 Ghz, avec Hyper-treading
- l'**Extreme Edition 965 Dual core** avec mémoire cache L2 de 2 X 2 MB et hypertreading. Les autres caractéristiques sont identiques à la version standard. Seul le chipset 975X accepte ce processeur.

## Sempron D'AMD

Sortis en juin 2004, les SEMPRON remplacent les processeurs Athlon XP sur le même socket A (462). Les caractéristiques sont quasiment identiques, seul le core a été modifié. La vitesse est un peu en dessous dans les premières versions. Sauf le 3100 +, ils utilisent le même socket A. Toutes les versions des Sempron passent en socket 754 au quatrième trimestre 2005. Début 2006, les Sempron sont tous remplacés par des processeurs 32 / 64 bits, remplaçant l'Athlon 64 comme processeur bureautique standard. Comme les Opteron et Athlon 64 FX, le Sempron 64 bits utilise l'Hyper-transport pour les communications inter-bridges.

Courant 2006, le Sempron passe à la mémoire DDR-2. Comme le gestionnaire mémoire est intégré au processeur, le socket est donc changé pour l'AM2.

## Les processeurs 64 bits.

Les instructions assembleur des processeurs X86 sont de 32 bits maximum. L'évolution est de faire passer le codage de ces instructions 32 bits à 64 bits. Le **principal avantage** est une utilisation plus faible de la mémoire, les instructions utilisent moins d'octets. Un programme 32 bits prendra 10 lignes de programmation, la même en 64 seulement 6 ou 7 dans les cas les plus courants. Ceci implique moins de données à transférer (utilisation des bus externes) et moins d'instructions à traiter (utilisation interne du microprocesseur). Le défaut, les instructions 64 bits ne sont pas compatibles avec les instructions 32 bits. Ceci nécessite au minimum de recompiler le programme, mais pour la majorité des logiciels standards de le racheter. Les systèmes d'exploitation Windows standard ne sont pas non plus compatibles 64 Bits. Microsoft a créé une version spécifique: Windows XP 64 bits (du moins pour les processeurs AMD).

Le deuxième avantage vient de la quantité de mémoire adressable. En effet, la taille maximum de mémoire RAM que peut utiliser un processeur est de **232** bits, soit 4 GB. Ceci est lié à la taille des registres.

INTEL et AMD travaillent chacun sur des processeurs 64 bits. La philosophie est néanmoins différente.

- INTEL a conçu un vrai microprocesseur 64 bits, l'ITANIUM. Ceci réduit l'utilisation des Itanium à celle de serveur ou de station de très haute gamme. Ces machines doivent concurrencer les systèmes UNIX de Sun notamment et pas les Sempron ou les Pentium IV. Seuls deux versions de 2003 serveur peuvent utiliser ce composant.
- Par contre, AMD développe des processeurs 64 bit pouvant tourner soit en 32 bits, soit en 64 bits. Le choix du mode de travail se fait au démarrage du système d'exploitation. Dans ce cas, l'architecture interne a été peu modifiée, les registres internes en 32 bits ont simplement été allongés vers le 64 bits. En mode standard, le processeur travaille comme un standard. S'il passe en mode 64 bits, il va simplement utiliser les registres "allongés", quelques nouvelles instructions spécifiques, voire ne plus utiliser des anciennes instructions en assembleur plus utilisées. L'avantage reste l'utilisation de systèmes d'exploitation standards 32 bits, voire une simple recompilation pour les versions 64.

AMD développe en parallèle 2 types de processeurs 64 bits, l'Athlon 64 bits et l'Opteron (spécifique serveur). Cette méthode est également employée depuis peu dans les XEON par Intel, pas dans les Pentium IV et Intel Core-2..

### AMD 64 bits

Connus anciennement sous le nom de code "HAMMER", AMD développe 2 versions de son processeur 64 bits: l'Opteron et l'Athlon 64 bits (la version FX est une amélioration du 64 au niveau cache). L'Opteron est la version serveur – station informatique de haute gamme, au même titre que l'Itanium et son successeur l'Itanium II. L'Athlon 64 bits pour stations est sorti en septembre 2003. Ces 2 processeurs acceptent les instructions usuelles 32 bits.

Les principales modifications par rapport à l'architecture K7 viennent du nombre et de la taille des registres (les mémoires de travail internes) qui doivent supporter à la fois les nouvelles instructions AMD64 d'AMD et SSE d'INTEL. Les Opteron et Athlons 64 bits sont gravés en 0,13  $\mu$ , tout comme les Athlon actuels, et utilisent un socket spécifique de type 940. Le cache L2 passe de 512K à 1 MB. La gestion mémoire n'est plus dévolue au chipset, mais bien directement au processeur qui gère 2 bancs (32 bits) de DDR333.

L'Athlon 64 bits reprend 95 % du core d'exécution d'un Athlon XP avec quelques modifications importantes:

1. **L'apport de registres 64 bits** : Le K8 se distingue principalement de l'architecture Athlon par l'ajout de 8 registres 128 bits (SSE/SSE2) ainsi que la possibilité d'accéder aux registres généraux en 64 bits, ce qui permet d'exécuter du code 64 bits, tout en restant compatible avec le 32 bits. C'est la technologie x86-64.
2. **Intégration du contrôleur mémoire** : La majorité des fonctions autrefois dédiées au Northbridge sont incluses dans le processeur, en particulier le contrôleur mémoire. Supportant la DDR333 et 266, le contrôleur mémoire intégré fonctionne à la fréquence du CPU et permet une latence particulièrement basse.
3. **Contrôleur HyperTransport intégré** : L'HyperTransport est un bus de communication informatique inter-bridge permettant aux différents "points" de la carte mère de communiquer ensemble, y compris dans les versions serveurs pour communiquer entre 2 microprocesseurs sur la carte mère. Il utilise un bus de données à bande passante évolutive et variable. Il contraste donc avec les autres normes de génération inférieure, comme le V-Link de VIA ou le MuTIOL de SiS qui fonctionnent en mode client/serveur.

- Le bus Hypertransport offre une bande passante maximale de 6.4 Go/s (version 1.1) par direction, la version 2.0 actuelle accepte jusqu'à 22,4 GB/s par direction.
4. **Un cache L2 amélioré** : Contrairement à l'Athlon XP qui embarque un maximum de 512 ko de cache L2, le K8 existera avec un cache de second niveau pouvant aller jusqu'à 1 Mo. Certaines versions de l'Athlon 64 seront équipées de seulement 256 ko.
  5. **L'ajout du SSE2** : Développé par Intel pour les Pentium 4, les instructions SSE2 font également partie de l'Athlon 64 et de l'Opteron et permettent la manipulation de registres 128 bits. Ces instructions multimédia offrent un gain de performance significatif sur les applications de calculs Audio/Vidéo/3D.
  6. **Le pipeline des calculs entiers** passe de 10 à 12 étages (pour 20 en Pentium IV et 31 pour le Pentium IV EE).

### Intel Core, Intel Code 2 Duo.

L'architecture Netburst, quoique prometteuse au départ, pose quelques problèmes, notamment un échauffement excessif dès que l'on monte en fréquence. Ceci est lié comme nous le verrons dans le chapitre sur l'architecture des processeurs à la taille du pipeline. En même temps que le Pentium IV, Intel développe le Pentium M, spécifique aux ordinateurs portables. Son architecture est basée sur le vieux Pentium III (remaniée) et lui n'a pas ces problèmes d'échauffement.

En 2006, Intel sort les **Intel Core Solo** et **Intel Core Duo** basés sur ce Pentium M. La vitesse chute automatiquement et Intel change les notations (comme pour les Pentium IV). Différentes séries sont développées pour les ordinateurs portables, mais également pour les ordinateurs bureautiques. Avec une mémoire cache L2 de 2 MB, ils utilisent le socket 479, identique au 478 mais avec ... une broche de plus. Les processeurs 479 acceptent de fonctionner sur certaines cartes mères au format 478 (en perdant quelques fonctionnalités d'économie d'énergie, parfois après une mise à jour du Bios). Les séries T1XXX (mono-core) et T2XXX (dual-core) sont les versions spécifiques pour ordinateurs de bureau. Le FSB est 667 Mhz, soit moindre que les Pentium IV équivalents de l'époque. Tous ces processeurs incluent la "**Virtualization Technology**" (rien de bien nouveau puisque c'est la possibilité d'exécuter plusieurs instructions simultanées) mais aussi le SpeedStep (réduction de la vitesse du processeur en fonction de la charge).

Même si ces séries ne sont pas très performantes par rapport aux Pentium IV de l'époque, l'architecture semble prometteuse et permet de sortir une architecture (un peu) modifiée basée sur le marketing, le **Core 2** et **Core 2 Duo**. Le socket est remplacé par le 775. Les notations changent également. La fréquence du bus externe est modifiée en 800, 1066 et même 1333 Mhz pour les plus performants actuellement (avec 4 MB de cache) en mode quadruple - la réelle fréquence est donc respectivement de 200, 266 et 333 Mhz comme pour les Pentium IV de la dernière génération. Comme d'habitude, Intel va développer l'outils marketing pour faire passer ces nouveaux modèles de processeur mais va également au niveau des performances reprendre le dessus sur les Athlon 64 d'AMD.

Fin 2007, pratiquement aucun simple core ne sont fabriqués dans la gamme Intel. Ceci sonne également la fin des Celeron, malgré les versions Celeron D (dual) qui ne sont plus finalement utilisés que pour les portables.

## Les microprocesseurs pour PC portables

Les premiers ordinateurs portables datent quasiment du début des PC. Malheureusement, la technologie du début des années 80 ne permettait pas de réduire ni l'encombrement, ni la consommation. Les premiers ordinateurs de ce type utilisaient un processeur standard, un écran cathodique, ... et rarement une batterie. Ils étaient tout simplement transportables. C'est seulement au début des années 90 que les premiers PC véritablement mobiles ont fait leur apparition sur le marché. Ceci est lié à la technologie des écrans plats d'une part, mais également à des améliorations dans la technologie des batteries.

Ces premiers modèles utilisaient des processeurs, mémoires, chipsets, ... standard: 386, 486, Pentium, ... Même le Pentium II avec son slot A était quasiment inutilisable pour cette fonction.

C'est Intel de nouveau qui lance des modèles spécifiques avec les Pentium III-M (M pour mobile). En même temps, la firme Transmeta sort en 1999 des processeurs spécifiques, les Crusoe.

Si les modèles actuels n'ont plus grand chose à envier en terme de vitesse avec les modèles standards, les grosses distinctions viennent de la consommation réduite en diminuant généralement la tension d'alimentation. D'autres mécanismes d'économie d'énergie réduisent la vitesse et la tension d'alimentation en fonction de l'utilisation. Comme nous le verrons dans le chapitre sur les chipset, les fabricants actuels développent également des modèles spécifiques.

### Les processeurs Crusoe.

Arrivés en 1999 – 2000 (mais développé depuis 1996), le Crusoe de la firme Transmeta est un cas atypique du monde PC. Ce processeur Risc inclut un interpréteur interne qui permet de faire tourner les applications X86, y compris les systèmes d'exploitation Microsoft. L'avantage de ce processeur est une faible intégration de transistors, le dégagement de chaleur est moindre et le processeur peut tourner à une fréquence supérieure. Il est dédié aux PC portables. Désavantage, le processus est ralenti avec l'interprétation des commandes compatibles X86. Ce processeur a été développé dans le plus grand secret, notamment avec l'aide software de Linus Torvalds, le boss de l'exploitation Linux

Le microprocesseur existe en deux versions (TM3120 et TM5400). Le TM3120 disponible depuis mars 2000 est cadencé en 333 et 400 Mhz. Il intègre un cache L1 de 96K, pas de cache L2. Il gère la SDRam de 66 à 133 Mhz. Le TM5400 disponible fin 2000 atteint une fréquence de 500 et 700 Mhz avec un cache L1 de 128K et un cache L2 de 256K. Il gère la mémoire DDR-SDRam de 100 à 166 Mhz.

Avec l'annonce par Sony qu'ils ne produirait pas de portables incluant ce processeur début 2002, il a quasiment disparu sauf pour des applications spécifiques.

### Pentium III-M

Sorti en 2001, c'est le premier processeur développé spécifiquement pour les PC mobiles. Il est identique au Pentium III sous bien des aspects (core Tualatin, notamment gravure en 0,13 micron, cache L2 de 512KB, ...).

La première spécificité est le **SpeedStep**. Cette technologie permet au processeur de commuter entre deux modes de travail: haute et basse fréquence. Cette fonctionnalité est réalisée en modifiant le facteur de multiplication.

La deuxième vient d'un chipset spécifique, le **I830** acceptant également la fonction SpeedStep: i830MP avec circuit graphique externe AGP 2/4X, I830M qui intègre un circuit graphique avec mémoire en option mais acceptant un circuit externe et le I830MG qui intègre un circuit graphique. Ce chipset gère jusqu'à 1 GB de mémoire Dimm SDRAM PC133 (donc mémoire standard), 6 ports USB 1.1 et une carte réseau intégrée.

### Pentium IV Mobile

Les processeurs pour portable suivent l'évolution des modèles bureautiques. Celui-ci découle du Pentium IV. Malheureusement, les défauts qui conduiront à l'abandon de l'architecture Netburst sont présente dans ce modèle, notamment au niveau consommation et pipeline trop long nécessitant une fréquence de travail élevée (ce qui n'est généralement pas intéressant pour un microprocesseur mobile). Intel l'abandonne rapidement pour le modèle suivant.

### Pentium M

Sorti en 2003, le **Pentium M** est une évolution du Pentium III-M mais récupère quelques caractéristiques des Pentium IV comme:

- bus externe Quad Pumped permettant de faire transiter quatre informations simultanément sur le bus externe (chipset).
- prédicteur de branchement
- Instructions SSE2

C'est surtout avec ce modèle qu'Intel lance la technologie **Centrino** pour certains modèles. Cette terminologie est plus commerciale que technique. Elle reprend un Pentium-M, un chipset spécifique mobile et une carte réseau sans fils intégrée. Différents coeurs ont été utilisés comme le Baniyas, le Dothan et le Sonama. Ce dernier se caractérise principalement par le bit de verrouillage, une fonction DEP intégrée qui vérifie si un programme essaye d'utiliser la zone mémoire réservée aux données.

Le **Celeron M** est une version bridée de ce processeur.

Ces processeurs seront remplacés en 2006 par les versions Intel Core.

### Intel Core

Comme nous l'avons vu, le core Intel utilisé par les processeurs bureautiques est dérivé du Pentium-M. Lorsque Intel reprend les Core 2 Solo et Duo, les microprocesseurs mobiles suivent forcément la même direction (marketing oblige) avec des versions Dual et quadri-core. La technologie **SpeedStep** est également améliorée.

Tous ces processeurs utilisent des chipset de la série 9 en version spécifique portable utilisant de la mémoire DDR-2 mobile.

Mi-2007, trois versions sont proposées (en plus du Celeron-M de l'ancienne génération):

## A. Intel Core2 Extreme

Le haut de gamme est constitué de:

- **QX6X50**: quadri-core, cache L2 partagée de 8 MB, bus principal (Quad-pumped) de 1333 Mhz.
- **X7900**: dual-core, cache L2 partagée de 4 MB, bus principal de 800 Mhz

## B. Intel Core 2 Duo

Ces processeurs de la série T7XXX, T5XXX et U7XXX intègrent deux processeurs dans le même boîtier, le cache L2 partagé varie de 2 à 4 MB avec un bus externe variant de 533 à 800 Mhz. Ils sont similaires aux processeurs bureautiques.

## C. Intel Core 2 Solo

Série U2XXX sont les équivalents en mono-core. Ces versions sont supprimées.



### Et AMD?

Longtemps non rentable, AMD ne s'est pas trop penché sur les problèmes des ordinateurs portables. Si une version K6-3+ était effectivement développée dans ce but, ce microprocesseur a peu été vendu (ou même proposé par la firme). L'arrivée des Duron et Athlon XP allait un peu changer la donne. Le Duron, ne pose pas trop de problèmes d'échauffement et va être utilisé en version standard. Par contre, l'Athlon XP n'a pas cette chance. AMD va modifier ce dernier en une version M. La première version avec une consommation de 35 W (Intel est à peine à 15 W) est peu utilisée par les constructeurs. C'est la version basse tension sortie en mi-2003 qui va un peu changer la donne. Comme les versions standards, ces processeurs utilisent la technologie **PowerNow!** Qui gère dynamiquement la fréquence et la tension d'alimentation.

Au remplacement de l'XP par le Sempron 32 bits, les modèles équivalents vont forcément être développés. Quelques modèles basés sur le socket 754 en Athlon-M, basés sur le Sempron sont compatibles 32-64.

Malgré de bonnes performances, ces processeurs gardent néanmoins le handicap d'une consommation élevée par rapport aux technologies Intel. Pentium-M

### Turion



En passant aux versions 64 bits pour ses microprocesseurs standards, AMD a changé de dénomination pour le Turion. L'architecture est strictement identique aux Athlon 64. Seuls des fonctionnalités d'économies d'énergie, couplée à des circuits d'interface spécifiquement développés le distinguent. Comme pour Intel, des versions simple et doubles cœurs sont développés.

Toutes ces versions utilisent un bus hypertransport à 800 Mhz (soit l'ancienne version, tout comme le Sempron 64). Ces processeurs sont donc d'office moins performants que les versions bureautiques puisque la version 1.1 permet une bande passante de 6,4 GB contre 22,4 GB/s pour les Athlon 64 (en Full Duplex).

Les premières versions utilisent le socket 754 avec mémoire DDR. Comme Intel, les dénominations sont données par un code: les ML-xx et MK-xx utilisent un cache L2 varie de 512 KB à 1 MB. Ils sont tous mono-core.

Les dernières versions utilisent de la mémoire So-Dimm DDR-2 et utilisent leur propre socket, le **S1**. Comme pour les modèles bureautiques, ce socket peut-être utilisé tant en simple core qu'en dual. La mémoire cache L2 varie de 512 KB à 2 X 256 KB ou 2 X 512 KB suivant les versions. Ce sont les versions TL-xx et TK-xx.

### Microprocesseur DUAL-Core

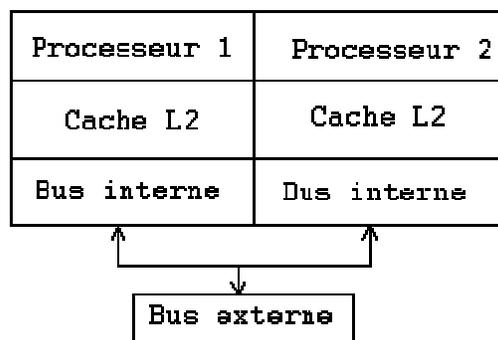
Depuis les premiers processeurs, l'augmentation de la vitesse du microprocesseur est à la base des gains de performances, ce n'est pas la seule comme nous le verrons dans le chapitre sur la structure des processeurs. Malheureusement, l'augmentation de la vitesse a ses limites. Intel a déjà implanté l'hypertreading dans quelques uns de ses microprocesseurs. L'Hypertreading permet au processeur, si le programme est compilé pour, de découper l'exécution en plusieurs parties, exécutées en même temps. Dans les serveurs, on utilise également plusieurs processeurs en parallèle pour augmenter les performances.

Ces 2 notions sont à la base du Dual-Core (coeur): implanter 2 processeurs qui travaillent en même temps mais dans **le même boîtier**.

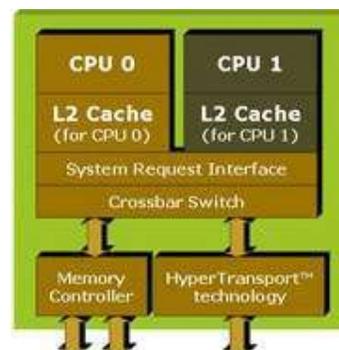
#### Différentes technologies du DUAL Core.

Pour fabriquer un microprocesseur bi-coeur, 3 techniques sont envisageables:

- fabriquer 2 processeurs distincts et les implanter dans le même boîtier, chacun gérant sa mémoire et ses entrées sorties. Cette méthode a peu d'intérêt, puisqu'il n'y pas de partage de mémoire et nécessiterait un socket avec le double de Pin de ceux actuels.
- Graver les 2 microprocesseurs dans le même core (puce électronique), c'est le choix actuel d'INTEL
- Intégrer 2 processeurs distincts dans le même boîtier mais chacun gérant sont propre contrôleur mémoire, technique développée actuellement par AMD, même si les Opteron et Athlon 64 actuels utilisent toujours la méthode 2.



Technologie Intel



Technologie AMD actuel dans ses processeurs 64 bits. Le contrôleur mémoire ne passe pas par le chipset

La première solution n'est par réellement envisageable puisque les 2 processeurs ne partageraient les données que via la mémoire RAM, ce qui entraînerait des conflits de lecture / écritures ou l'utilisation d'un composants chargé de départager les accès mémoire.

### Avantages.

A la différence de l'Hyper-Treading qui nécessite des programmes compilés compatibles pour pouvoir fonctionner, les processeurs DUALCORE fonctionnent avec des programmes standards. Pour le système (logiciel), les 2 processeurs ne sont vus que comme 1 seul. Par contre, le traitement va être en théorie doublée. En effet, l'utilisation des bus externes reste limitée. Seul la vitesse de traitement interne va augmenter les performances globales de l'ordinateur puisque 2 programmes vont pouvoir fonctionner en même temps.

Par contre, le processeur n'est pas l'unique goulot d'étranglement dans un système informatique. Les périphériques externes tels que le disque dur ou même la mémoire vont jouer sur les performances de la machines. Les performances des applications faisant de nombreux appels sur le disque dur ne seront donc pas améliorées par le Dual-Core. En même temps, l'utilisation de tels processeurs nécessite une quantité de mémoires supplémentaire.

### Processeurs Dual Core INTEL

Intel Implante cette technologie dans divers processeurs de haut de gamme comme dans:

1.les processeurs bureautiques Pentium IV Extreme Edition (dernières versions), Pentium IV **D**, Celeron **D**, Intel Core 2 Duo



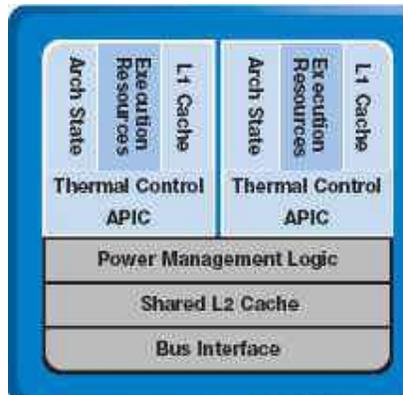
2.dans les serveurs, notamment dans les Xeon (également en Quadricode depuis novembre 2006) et Itanium II



3.dans les processeurs portables, le centrino Duo



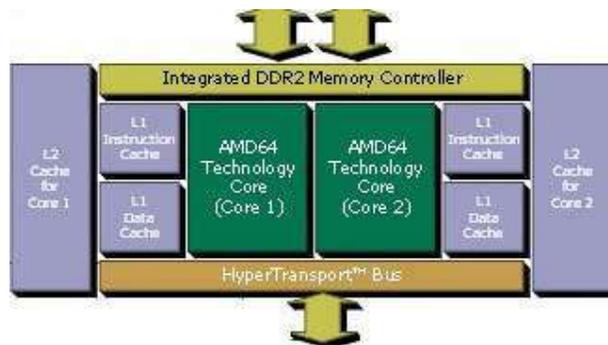
La technologie d'Intel utilise un cache L1 par processeur intégré, le cache L2 est lui partagé directement entre les 2 processeurs du Dual-core. L'utilisation de processeurs Dual-core nécessite un chipset spécifique (minimum l'i955X pour les P4 Extreme Edition).



## Dual Core AMD

Comme son concurrent, AMD développe également cette technologie. La méthode est différente du fait de l'architecture interne de ses processeurs 64 bits au niveau de la gestion de la mémoire, directement par le microprocesseur et plus par le northbridge du chipset. C'est d'ailleurs ce qui explique l'architecture NUMA des serveurs AMD alors qu'Intel utilise l'architecture SMP.

Dans les Athlon 64 et 64 FX, Sempron 64 bits (socket 754) et dans les Opteron, le contrôleur mémoire est directement implanté dans le processeur. La méthode utilisée par Intel dans le multiprocesseur standard n'est donc pas appliquée dans les serveurs AMD.



Dans le cas d'AMD, les caches L2 sont partagés directement via l'hypertransport (jusque 8 GB/s comme taux de transfert). Le bus mémoire lui est directement géré par les 2 microprocesseurs. La mémoire peut-être DDR (socket 939 ou 940) ou DDR-2 (AM-2). Le cache L1 reste à 128 KB mais pour chaque Core (64 pour les données et 64 pour les instructions). Suivant les versions (Turion pour ordinateur portable, Athlon 64 X2 ou Opteron), le cache L2 est de 512KB ou 1024 KB par coeur.

## Les systèmes d'exploitation.

En théorie, tous les systèmes d'exploitation standards permettent d'utiliser des processeurs Dual-Core. Au niveau de Microsoft, les ordinateurs Bi-processeurs nécessitent au minimum Windows 2000 ou Windows XP Pro. Par contre, l'utilisation de 4 processeurs ou plus passe par des versions dédiées serveur de Windows 2000 ou Windows 2003 (suivant également le nombre de processeurs). Par contre, les versions Bi-Coeur sont vues comme 1 seul processeur en XP (Home et pro), 2003 et VISTA (y compris avec l'hyperthreading toutes versions confondues). Les

versions basiques sont simple processeur (physique), les autres versions de VISTA acceptent 2 processeurs.

### Structure interne des processeurs

Les processeurs ont été analysés par type (famille, fabricants). Nous allons analyser l'architecture interne, la structure. Connaître la structure interne de chaque processeur en détail serait trop compliqué, mais ceci va permettre de clarifier, de comprendre, ces différences de performances des processeurs pour une vitesse équivalente, comme par exemple, pourquoi un Pentium IV est moins performant qu'un Intel Duo Core de vitesse nettement inférieure.

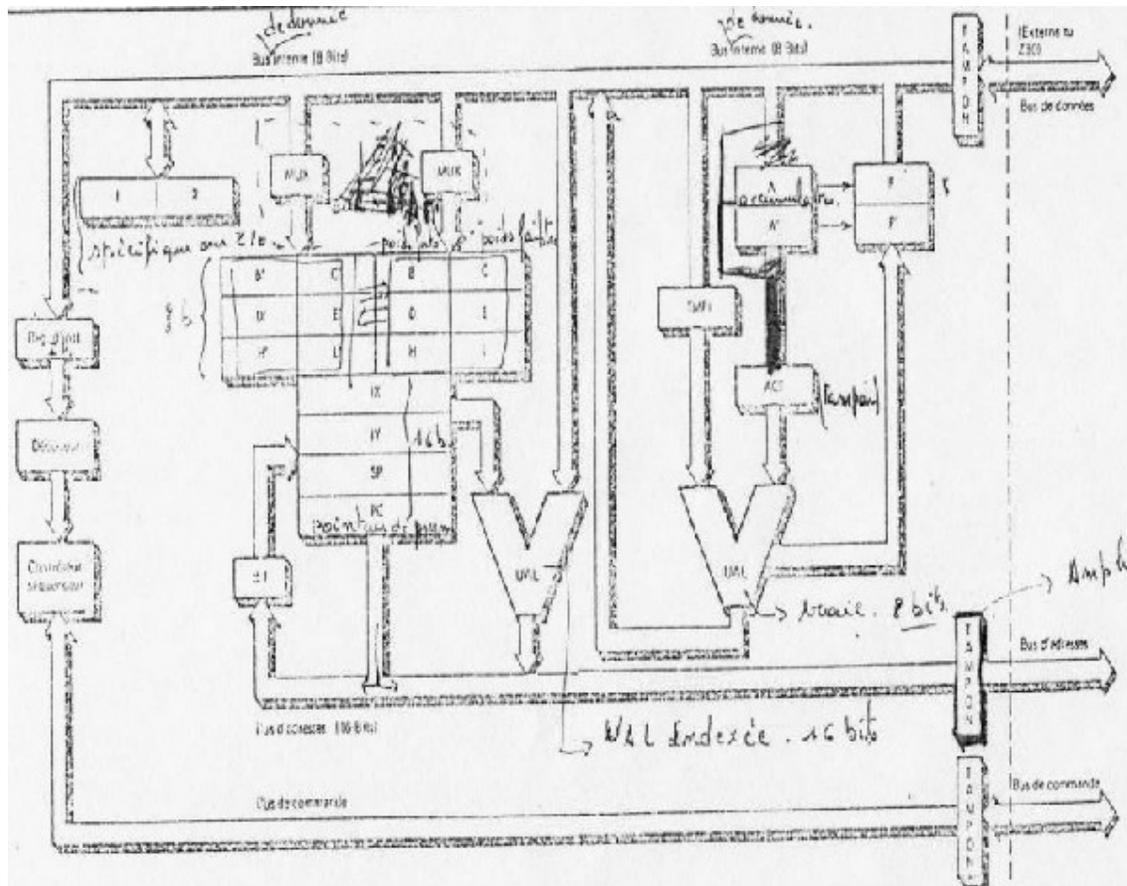
Tous les "processeurs PC" utilisent les mêmes instructions que les premiers processeurs 8086 d'INTEL (mais souvent le système d'exploitation n'est plus compatible avec ces anciens programmes). Les instructions qu'un processeur est capable de lire et exécuter sont programmées en assembleur. Tous les systèmes d'exploitation et programmes sont recodés à partir de langages évolués (C, pascal, Visual Basic, ...) vers cet assembleur. Si un nouveau processeur utilise les mêmes instructions de base qu'un processeur de la première génération, ceci n'est pas limitatif. En cours d'évolution, certaines instructions ont été ajoutées pour permettre des décodages d'adresses sur une plage supérieure (386) ou pour des instructions multimédia (MMX), SSE, SSE II et III pour les Pentium III, IV, 3DNow pour K6-2 et K6-3, ...

L'obligation de garder les instructions de base du 8086 pour les processeurs suivant pose de sérieux problèmes pour la conception de nouveaux microprocesseurs compatibles X86. Les instructions du 8088-8086 ne sont pas franchement faciles. Codées entre 1 et 5 bytes, elles ne permettent pas de prédire l'espace utilisés par l'instruction suivante. Suivant un codage à 8 bit des instructions, une instruction peut prendre plusieurs bytes.

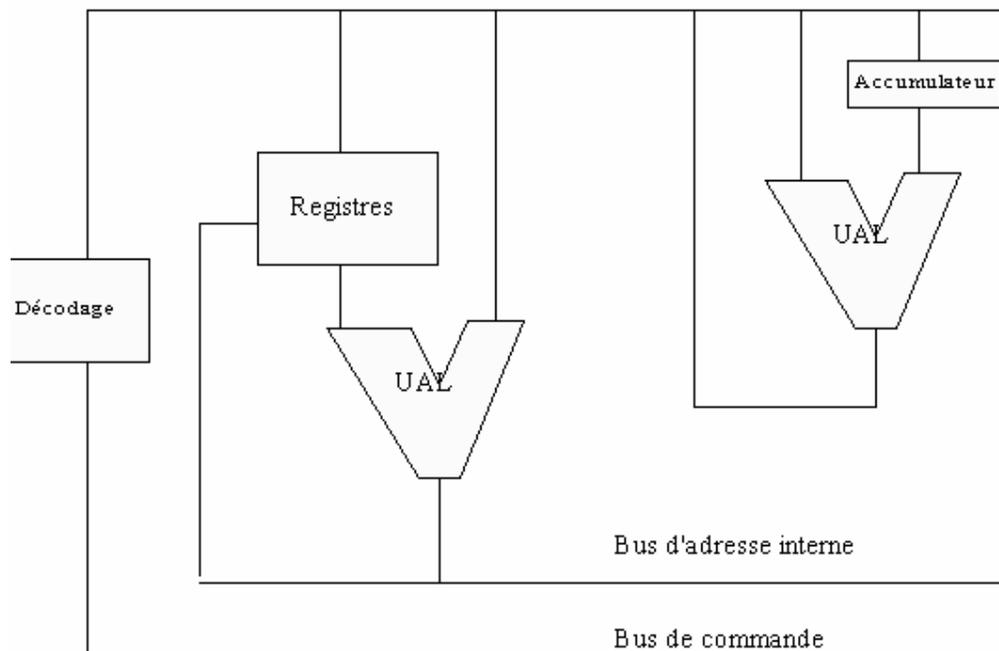
Les processeurs 8086-8088 sont de types **CISC** (Complex Instruction Set Computer). Les premiers concepteurs de processeurs rajoutaient le plus d'instructions possibles pour permettre à l'utilisateur de peaufiner ses programmes. Pourtant, ces multiples instructions ralentissent le fonctionnement du microprocesseur et sont peu utilisées en pratique. Actuellement, on utilise des processeurs **RISC** (Reduced Instruction Set Computer). Le nombre d'instructions est réduit, mais exécutées nettement plus rapidement. Chaque instruction complexe peut être programmée par plusieurs instructions simples. La technique actuelle est de charger des instructions CISC 8086-8088 pour les recoder en instructions RISC plus rapides.

### Structure interne d'un processeur de base

Reprenons notre Z80 et distinguons les différentes parties: au dessus, un bus interne de données, en dessous un bus interne de commande et un bus interne d'adresse. Connaissant le schéma des systèmes à microprocesseurs, ceci n'est pas étonnant. Dans la partie de droite, une ALU (Unit Arithmetic Logic) ou UAL (Unité arithmétique et logique en Français). Cette ALU est une unité de calcul, elle est connectée à une mémoire tampon notée l'accumulateur et au travers d'un bus à un ensemble de mémoire appelés registres. Dans notre cas, ces registres sont notés B,C, D, ... Les registres A', B', C', ... sont spécifiques au Z80, mais existent également dans les 8088 (compatibles au niveau des instructions). Ceci permet de travailler avec deux zones mémoires que l'on fait basculer. L'ALU de gauche permet de faire des calculs sur des adresses. Remarquez le petit incrémenteur +1 qui permet d'augmenter le pointeur programme (l'adresse où se trouve la prochaine instruction du programme).



Le schéma des processeurs de cette génération reprenaient le schéma suivant:  
 Bus de donnée interne



Cette structure charge un code d'instruction à la fois, soit pour une instruction de type LDA XX XX (chargement dans l'accumulateur de la donnée contenu à l'adresse mémoire XX XX), trois cycles d'instructions (une pour LDA et 2 pour l'adresse). Cette condition n'est de plus remplie que si le processeur charge une instruction par cycle d'horloge, mais c'est le cas dans un système INTEL. Comme les instructions des processeur actuels sont compatibles avec les 8088, les tailles des registres le sont aussi. Il est donc impossible d'augmenter la taille de ces mémoires pour augmenter la vitesse d'exécution.

**Quelles sont les possibilités d'augmenter la vitesse d'un processeur sans modifier la structure?**

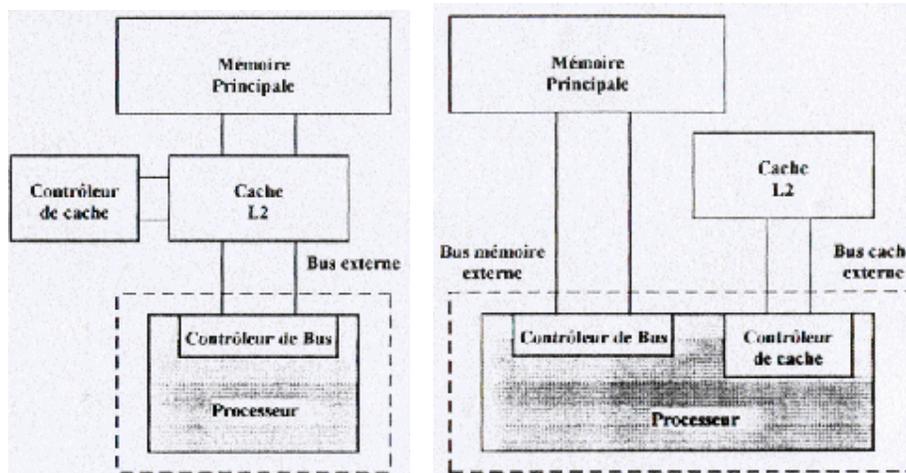
1. exécuter plusieurs instructions en même temps, mais cela est difficile puisque certaines instructions sont conditionnelles.
2. faire passer via un décodeur les instructions de type CISC en RISC plus rapide.
3. S'assurer que le processeur n'attend jamais d'instructions, qu'elles soient directement accessibles.
4. augmenter l'intégration en diminuant la taille des transistors. Ceci augmente la vitesse de transfert entre les différentes parties.

Tous ces trucs sont actuellement utilisés, tant par INTEL, que par AMD et Cyrix (VIA).

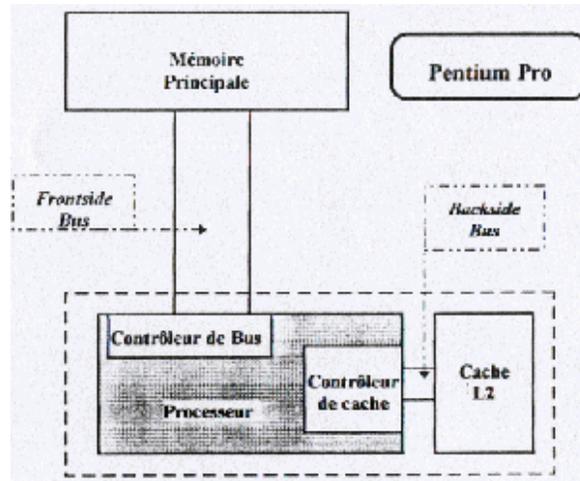
**Utilisation d'une mémoire cache**

En insérant une mémoire tampon rapide entre le processeur et la mémoire (plus lente que le processeur), on permet au processeur d'accéder rapidement à des données et instructions qu'il utilise souvent. On distingue un cache interne L1 (implanté dans le processeur) et un cache L2 (extérieur au processeur, même s'il est implanté dans le même boîtier, cas des Pentium II, K6-2 et K6-3 et suivants).

**Les différentes architectures de cache**



Architecture 486, Pentium, K6, 6X86 Architecture Nexen NX586



Selon le modèle de processeur, le cache L2 est géré différemment. Dans le cas des Pentium Pro et Pentium II, le cache L2 n'est plus en tampon direct entre la mémoire et le processeur, mais directement géré par le processeur. Le cache est implanté dans le même boîtier pour ces processeurs.

A titre d'exemple, un CELERON 300 sans cache L2 est 40 % inférieur en performance avec un Pentium II de même fréquence et de structure identique.

De plus, les processeurs actuels utilisent 2 caches L1 internes: un pour les données, un pour les instructions. Plus la taille du cache augmente, plus les difficultés à gérer ce cache augmentent, mais plus la vitesse augmente.

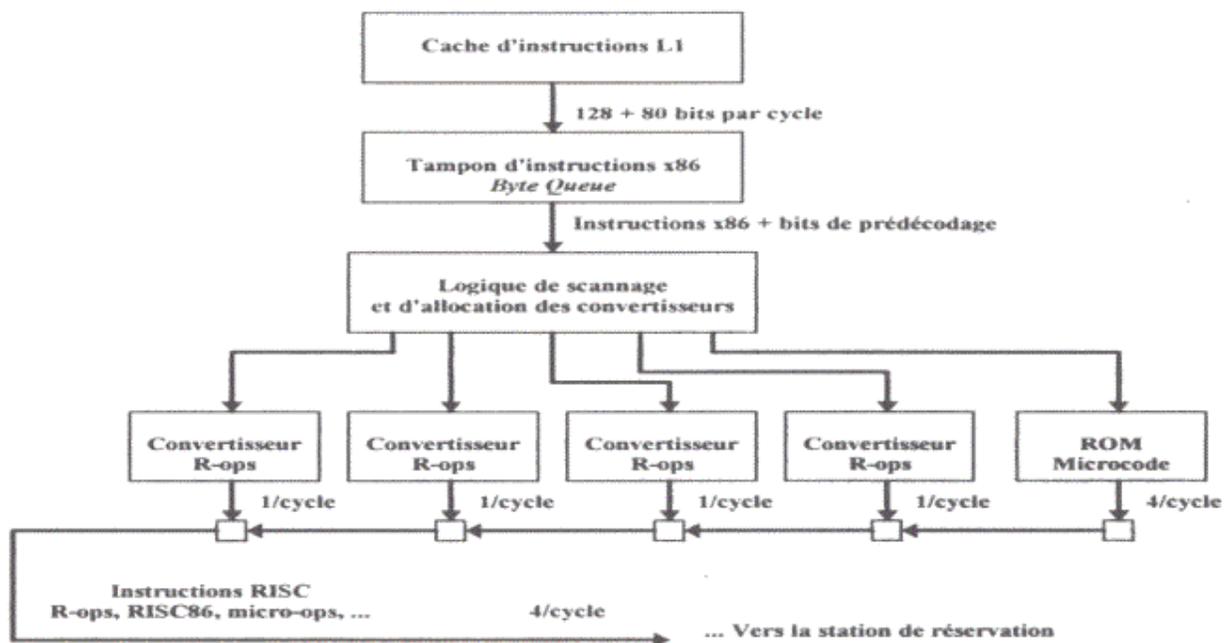
### Passage du CISC au RISC

Une autre solution pour augmenter les performances est de faire passer la structure des instructions du **CISC** (instruction 8088) en **RISC**. Chaque firme utilise en interne ses propres instructions, meilleur est ce transfert, meilleures sont les performances. Ceci explique que la vitesse effective d'un processeur n'est pas identique pour 2 processeurs de même fréquence, mais de structure différente.

Avantage des processeurs CISC:	Défauts des processeurs CISC
<ul style="list-style-type: none"> <li>• Instructions plus proches d'un langage de haut niveau</li> <li>• Programmation plus compact</li> <li>• Ecriture plus rapide et plus élégante des applications</li> <li>• Moins d'occupation mémoire des programmes</li> <li>• Exécution nécessite moins d'octets mémoire</li> </ul>	<ul style="list-style-type: none"> <li>• beaucoup trop de codes d'instruction différents.</li> <li>• taille des instructions élevée et variable (1 à 15 bytes – octets par instruction)</li> <li>• structure des instructions non standardisées: exécution complexe, peu performante</li> </ul>
Microprocesseur RISC	
<ul style="list-style-type: none"> <li>• Nette séparation entre les instructions d'accès mémoire et les autres</li> <li>• Instructions standardisées, en taille et en durée d'exécution</li> </ul>	

- Unité de décodage câblée, non microcodée architecture superpipeline, superscalaire
- Très nombreux registres à usages général
- Un ou plusieurs cache (s) internes(s) ainsi que des tampons internes et un jeu d'instruction réduit aux instructions simples.

Chaque processeur moderne va chercher des packets d'instruction X86, les décode sur plusieurs rangées de 8 bits pour les faire passer en instructions RISC. Ceci pose des problèmes lors d'instructions conditionnelles. Comment le processeur peut charger plusieurs instructions et les exécuter suivant le choix non prévu au départ? Chaque concepteur utilise soit un circuit de prédiction situé en amont de circuit de décodage CISC - RISC (cas d'INTEL), soit une table de branchement qui mémorise les derniers branchements (AMD). Comme les instructions sont chargées longtemps en avance, il se peut très bien que les données soient changées ...



Ce changement de type d'instruction a également permis d'implanter une (ou plusieurs) unités de calculs spécialisées en virgule flottante (nombres réels) au sein du processeur. Cette unité de calcul est le processeur mathématique. Ce coprocesseur est utilisé dans des calculs spécifiques, utilisés principalement en dessin technique et en jeux.

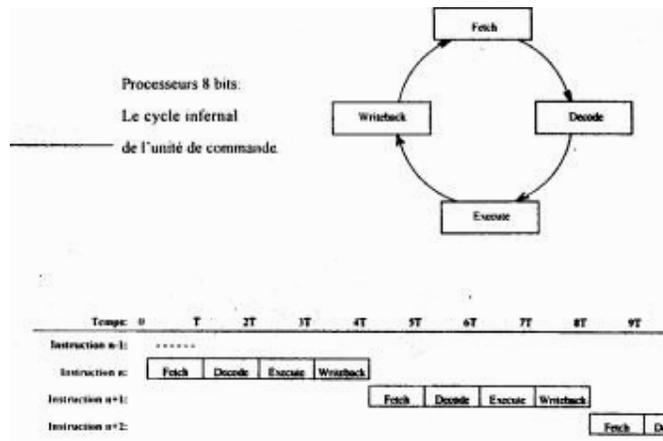
Le passage de processeurs RISC aux processeurs CISC a amené d'autres changements internes, nous allons les passer en revue.

### Architecture pipeline, superpipeline et superscalaire

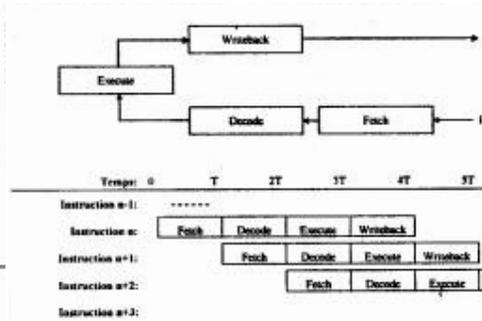
Dans les 8088-8086, le processeur ne commençait le traitement de l'instruction suivante que lorsqu'il avait terminé la précédente. L'utilisation d'une structure pipeline permet le traitement de plusieurs instructions en même temps. Dans le cas d'une structure pipeline, l'exécution d'une instruction est décomposée en:

1. chargement de l'instruction dans le processeur
2. décodage
3. exécution de l'instruction
4. renvoi du résultat

Comme seules la première et la dernière instruction utilisent les bus externes, le cycle de l'instruction suivante peut commencer en même temps que le décodage interne de l'instruction. Ceci permet d'exécuter plusieurs instructions en même l'Opteron temps.



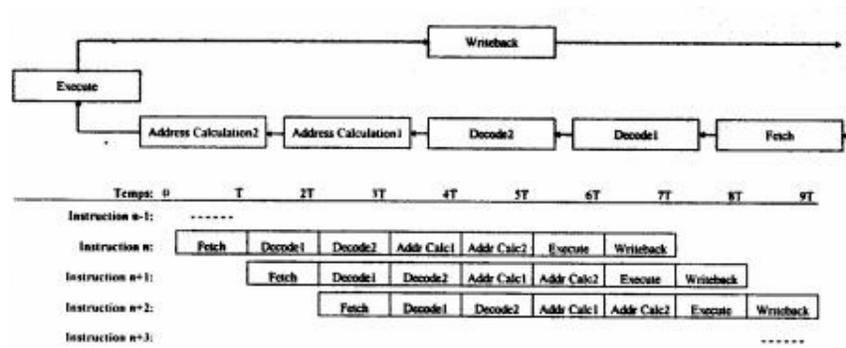
**Architecture standard**



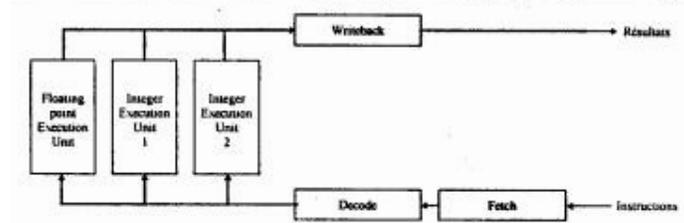
**Architecture PIPELINE**

Dans une architecture superpipeline, l'exécution de l'instruction est maintenant divisée par 6. Le nombre d'exécution simultanée n'est guère supérieure en super pipeline par rapport à une architecture pipeline dans les instruction habituelles. Par contre, comme de temps utilisé par chaque partie d'instruction est plus faible, chaque fois qu'une partie est omise (parce que pas nécessaire) l'exécution est encore plus rapide pour la suivante.

Super pipeline: le processeur ne traite pas les instructions d'un seul coup, mais en 6 étapes: pré extraction, décodage d'instruction (2), génération d'adresse et écriture différée (2) (cas d'un chargement vers ou de la mémoire), exécution, transfert des données.



Architecture superpipeline: le processeur M1 de CYRIX, l'architecture va jusqu'à 7 niveaux.



Architecture superscalaire

Dans la structure interne de processeurs superscalaires, le processeur inclut plusieurs unités d'exécution d'instruction qui travaillent en même temps. L'Athlon est superpipeline et super scalaire.

### Taille des transistors, alimentations.

Actuellement, les microprocesseurs utilisent des transistors de type CMOS ou BiCMOS. Plus la technologie progresse, plus la tension d'alimentation et la taille des transistors diminuent et plus le nombre de transistors augmentent. Une réduction de la taille des transistors diminue la distance de transfert entre 2 informations et donc moins de dissipations thermiques (puisque le signal suit une distance plus courte, moins d'échauffement dans le processeur). De plus, la complexité des processeurs actuels augmente le nombre de transistors.

En 1995, la taille d'un transistor était de 0,5 micron. En 2003, elle est de 0,13 micron avec les Pentium IV supérieurs 1,7Gh et derniers athlon.

Si depuis les Pentium MMX, l'alimentation est bi-tensions, la tension d'alimentation également puisque celle d'un K6-3 à 450 Mhz est actuellement de 2.2 V, les processeurs actuels descendent à 1,3V. Attention, vérifiez systématiquement la tension sur le processeur et non sur la carte mère ou dans les notes techniques de celle-ci. Pour les processeurs bi-tensions, prendre le V Core , même si la deuxième tension est affichée.

La modification des tensions des processeurs est arrivée avec les premiers 486DX2 de Cyrix (mais est devenue générale). Tous les modèles antérieurs utilisaient le 5V. Une erreur d'alimentation sur un processeur entraîne sa destruction à brève échéance.

### Architecture interne des processeurs

En additionnant quelques unes des avancées ci-dessus, on arrive par exemple à un K5 2,5 fois plus rapide qu'un 486 de même fréquence. Analysons brièvement quelques structures internes de processeurs. Commençons par le K5 d'AMD ci-dessus.

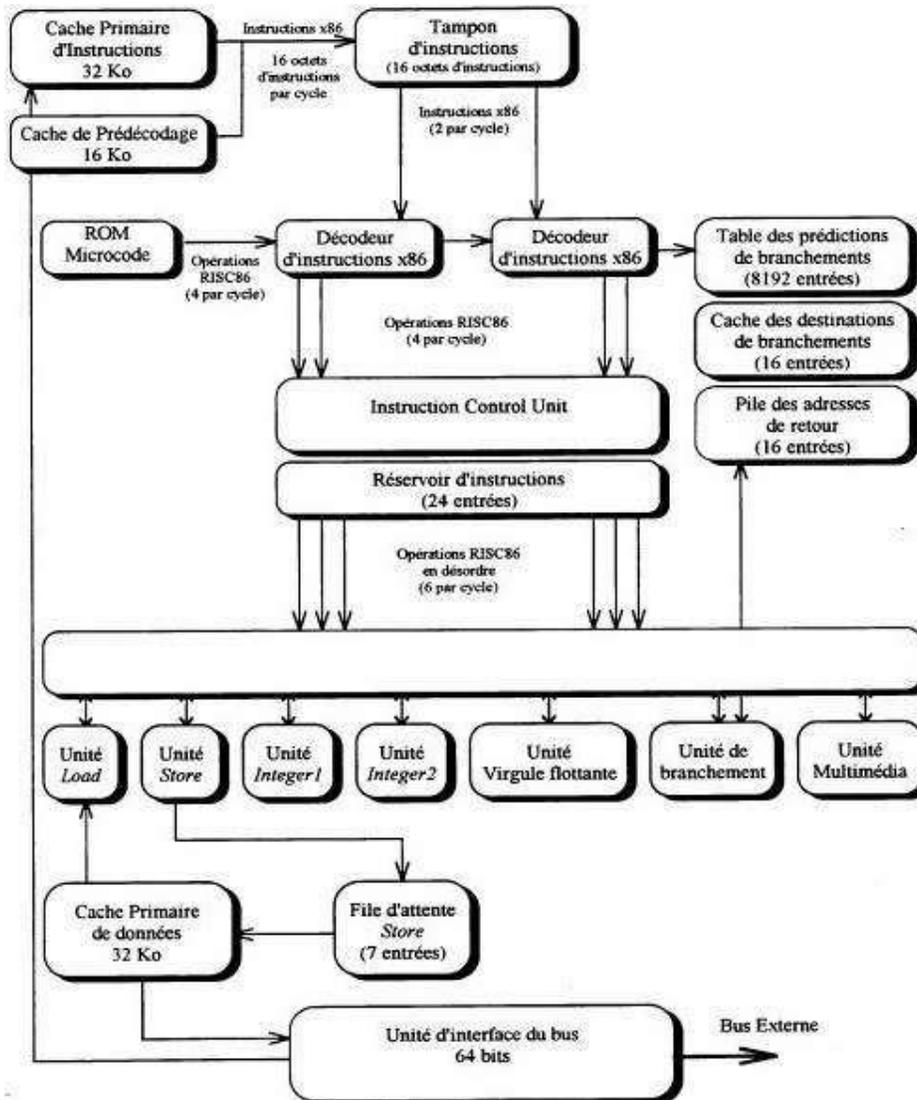
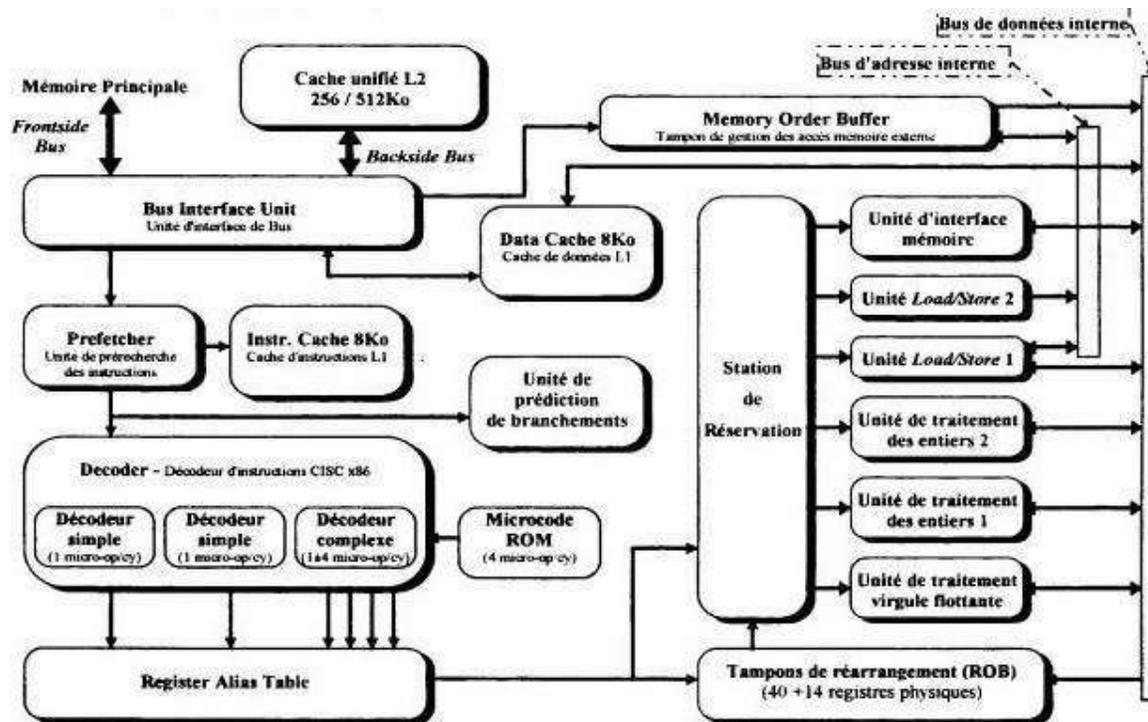


Schéma d'architecture générale d'un AMD K5

En haut à gauche, le cache L1 instruction (le cache de données se trouve en bas) par qui transite toutes les instructions en provenance de la mémoire (cas des AMD). Le système en-dessous de prédécodage est utilisé pour les vérifier le type d'instructions qui arrivent (mémoires ou exécution) et sert aux prédictions de branchements. Le tout est relié à un tampon d'instructions qui envoie 2 instructions X86 par cycle d'horloge vers 4 décodeurs d'instructions (passage du CISC au Risc) Ce décodage est exécuté par la ROM Microcode. Les instructions RISC sont transférées vers une table de prédilection de branchements, associé à un autre cache. D'autres part, les instructions sont traitées en RISC par l'instruction Control Unit qui envoie également les instructions à partir d'un réservoir de 24 commandes RISC86 vers les différentes parties en-dessous: Load et Store (gestion mémoires), deux unités d'entiers et une de virgule flottante, une unité de branchement qui gère également la table de prédiction et une unité multimédia liée aux instructions MMX. Le tout est relié au bus de donnée externe via le cache de donnée et l'Unité d'interface Bus qui gère également le cache d'instructions. Celle du Pentium Pro ci-dessous est identique, à part le cache L2 (externe) qui n'est pas relié directement à la mémoire. Cette structure est reprise dans les Pentium II et Pentium III.



Architecture générale d'un processeur Pentium Pro (identique au Pentium II et III).

## Conclusion.

Ceci n'est que les premiers balbutiements de la modification de la structure interne des processeurs. La modification de l'architecture rend les performances d'un processeur difficiles à vérifier en se basant uniquement sur la fréquence. Ceci explique les bricolages commerciaux de Cyrix qui désignait systématiquement les vitesses de ses processeurs, non pas par la vitesse effective, mais par des vitesses de comparaison par rapport aux équivalents INTEL en fonction des Benchmarks (quoique). AMD avec les Athlon XP et Sempron fait de même et Intel est revenu en arrière dans la course à la fréquence.. Si le nombre d'instructions décodées en même temps devrait encore augmenter, la direction actuelle des processeurs va vers la diminution de la taille des transistors, ce qui conduit à des possibilités d'augmentation de fréquence. En effet, plus le nombre de transistors augmente, plus la dissipation de chaleur est importante. Mais plus la taille d'un transistor diminue, moins cette dissipation est importante. D'autres matériaux sont en cours d'utilisation, IBM utilise déjà des conducteurs internes en cuivre, alors que la technologie actuelle est basée sur l'aluminium. Le cuivre est meilleur conducteur électrique (d'où moins de dissipations), mais nécessite une technologie plus élevée, avec des difficultés d'intégration plus importante. Ceci n'évite de toute façon pas le ventilateur sur le processeur.

## Structure interne des microprocesseurs Intel actuels

### L'architecture Netburst

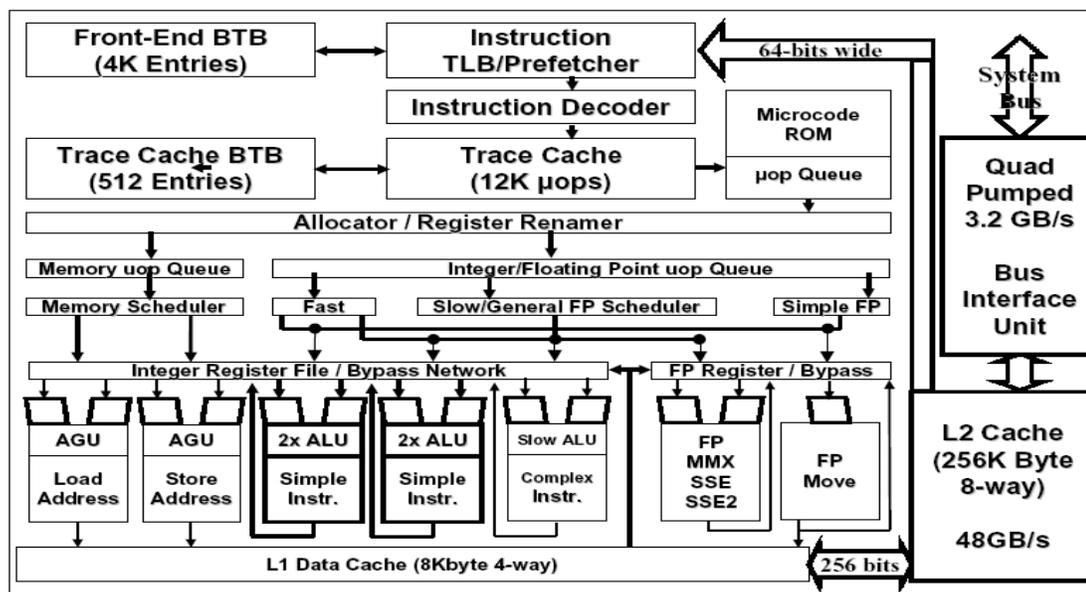
C'est en 2000 qu'Intel sort le Pentium IV (core Willamette). Ce processeur adopte l'architecture Netburst.

En analysant l'historique des processeurs, nous avons déjà en partie vu cette architecture. Un bref rappel de ses caractéristiques:

- Pipeline à 20 niveaux contre 10 pour les Pentium III. Ca signifie que la séquence d'exécution des instructions se fait maintenant sur 20 niveaux. Ceci sera augmenté à 21 dans la version suivante et même 31 niveaux pour les derniers processeurs utilisant cette structure
- Le **cache L1** ne contient plus qu'un cache données de 8 K et une "**Instruction Trace Cache**" qui stocke les instructions après leur décodage en RISC. Ce cache programme peut contenir jusque 12.000 instructions.
- Les instructions peuvent être exécutées en partie dans le désordre.

Le premier point est déjà un avantage par rapport aux concurrents de l'époque. Le processeur va traiter jusqu'à 20 instructions en même temps. L'Instruction Trace cache garde maintenant des instructions prédécodées alors que les modèles précédents les conservent en CISC (compatible X86) et doivent de nouveau les décoder pour les exécuter. La troisième caractéristique permet d'exécuter une instruction si la précédente n'est pas dans le cas. Cette solution améliore généralement les performances mais va parfois les réduire en utilisant l'hyperthreading.

Malheureusement, ce superpipeline à 20 niveau va vite trouver ses limites. Ceci est principalement du au temps de latente des mémoires. La bande passante mémoire demandée est trop élevée. L'autre problème vient de l'Instruction Trace Cache, trop élevé en cas d'erreur de prédiction. Tout le cache doit être vidé et rechargé en cas d'erreurs.



Analysons ce schéma. A droite, le bus d'interface externe, en mode QUAD (4 transferts par cycles). Il est relié au cache L2. Ce cache L2 est relié d'une part (en-dessous) au cache L1 donnée, et au préchargeur d'instruction. Ce dernier est relié au "Front End BTB" qui sert de prédicteur de branchement. L'ensemble est relié au décodeur CISC - RISC. Une fois le décodage effectué, on passe par le Trace cache, en fait le cache L1 programme.

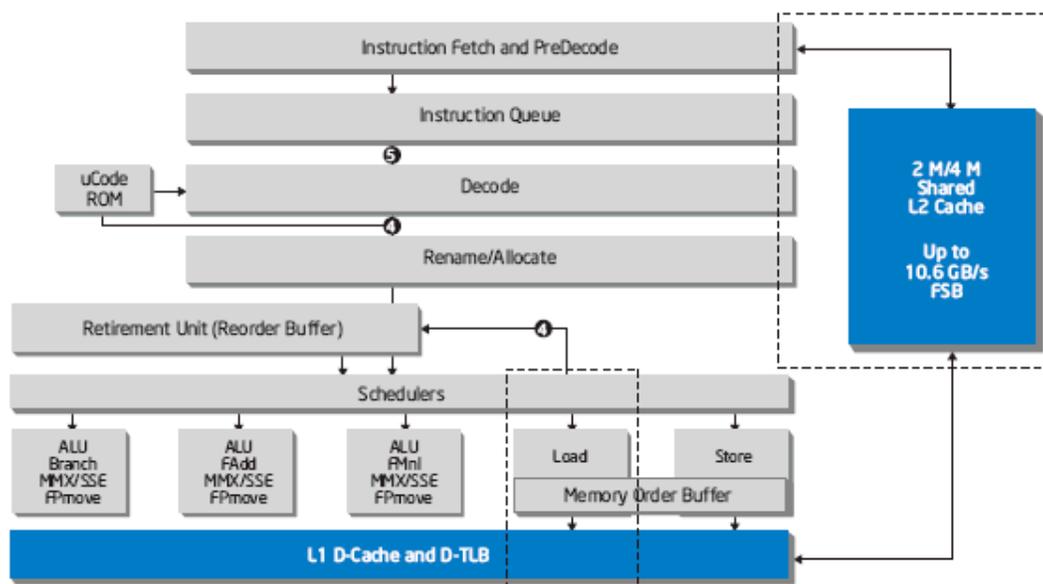
Le passage des instructions se fait via le μop Queue, éventuellement après un passage par la microcode ROM pour les instructions complexes. Suivant le type d'instructions, les données - programmes sont envoyés via un buffer FIFO mémoire, ou vers les unités de calcul.

L'AGU sert d'unité de calcul pour les adresses (saut du pointeur par exemple), les 2 ALU tournant au double de la vitesse du processeur s'occupant des données. A droite, on retrouve l'unité spécialisée pour les instructions MMX, SSE et SSE2. L'ensemble est contrôlé par le Bypass network qui permet l'exécution des instructions dans le désordre.

### Intel Core et Core 2 Duo

L'architecture Netburst était prometteuse. Néanmoins, pour permettre au pipeline de fonctionner optimalement, la fréquence de travail doit être élevée, notamment en cas d'erreur de branchement.

En même temps, INTEL développait une architecture M pour ses processeurs mobiles, basée sur les Pentium III. Finalement cette architecture permet de meilleures performances à vitesse moindre. C'est ce qui a conduit Intel à arrêter les Pentium IV pour se pencher sur l'architecture Intel Core, dérivée de celle des Pentium M, un processeur dérivé du Pentium III pour PC portables.



Les Dual-Core et les quadri-core d'Intel ne sont finalement qu'une architecture Intel Core reprenant deux ou quatre processeurs dans le même core (boîtier).

### Structure des microprocesseurs Athlon 64

Les athlon 64 dans le sens général utilisent une architecture interne mixte 32 - 64 bits. Ils sont donc capable d'utiliser des systèmes d'exploitation 32 bits (Windows 95/98/me/2000/NT/XP) et des système d'exploitation 64 bits (Linux, Windows 64 bits - XP ou les différentes versions 64 bits de Vista et 2003). Si Intel a créer de nouvelles instructions pour l'Itanium, AMD (suivi par INTEL avec le XEON) n'a pas fondamentalement changer l'architecture. Elle repose sur des registres au format 64 bits pour les instructions et données mais dans ce cas, elles ont été simplement allongées (ce qui explique la compatibilité avec les instructions 32 bits). La limitation précédente à 32 explique notamment la limitation mémoire des 4 GB ( $2^{32}$ ). En passant en 64, cette limitation n'existe plus. D'autres instructions passent également en 1 seul cycle d'horloge sur le bus au lieu de 2, d'où une augmentation de la vitesse

globale de l'ordinateur avec des versions de systèmes d'exploitation spécifiques pour Windows ou recompilées (Linux). Ces processeurs se basent sur le **jeu d'instructions MIPS64™** (l'Itanium utilise l'**IA64**). En mode 64, quelques instructions ont été rajoutées, alors que des anciennes instructions X86 peu utilisées ont simplement été supprimées. Divers registres multimédia ont également été ajoutés pour les instructions SSE.

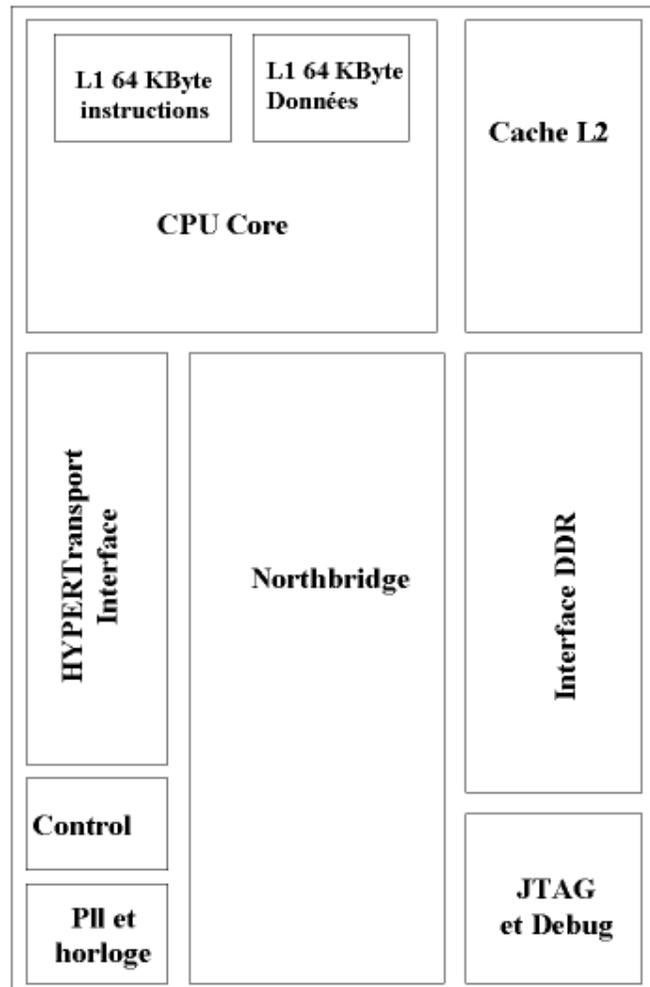
Le microprocesseur Itanium d'Intel utilise uniquement une architecture 64 bits spécifique, il ne peut donc utiliser que des systèmes d'exploitations 64 bits spécifiques, en gros les versions serveurs haut de gamme de Windows 2003 et 2008 à partir de la version entreprise.

Un processeur 64 bits utilisant un système d'exploitation 64 bits peut utiliser conjointement des applications 32 et 64 bits (quoique). Par contre, cas des AMD, un processeur 64 bits tournant sur un système d'exploitation 32 bits ne peut faire tourner que des applications 32 bits. A l'inverse, un processeur 32 bits ne peut utiliser que des systèmes d'exploitation et des applications 32 bits.

### Architecture standard des athlon 64, athlon 64FX et Opteron

Reprenons pour commencer l'architecture standard d'un microprocesseur athlon de la famille 64. Il est constitué de 2 parties: la partie processeur comme nous connaissons et la partie "Northbridge", anciennement inclus dans le chipset.

Le cache L2 varie suivant le processeur: 256, 512 ou 1 MB pour l'Athlon 64, 1 MB pour l'Athlon 64 FX51 et Athlon FX53 et Opteron. Par rapport à l'architecture K7, le pipeline passe de 10 à 12 niveaux.



Dans la partie "Northbridge", on va retrouver l'interface mémoire DDR ou DDR2. C'est la première différence entre les différents processeurs de la famille 64 bits d'AMD. Les Athlon 64 gèrent un bus de donnée externe de 64 bits (socket 754), ou comme les FX et Opteron, un bus de données de 128 Bits (en fait 2 bus 64 bits dual Channel) sur un socket 939 ou AM2. Comme ces processeurs gèrent les mémoires ECC, les Athlon 64 bits accepteront la mémoire ECC sur 8 bits de contrôle contre 16 bits-Ecc pour les autres.

La partie JTAG - Debug est liée à la difficulté de tester des circuits électroniques complexes. Cette méthode développée fin des années 80 permet de tester "de l'extérieur" le circuit électronique en mémorisant les valeurs d'entrées et sorties.. Elle n'est donc pas directement utilisée dans les PC.

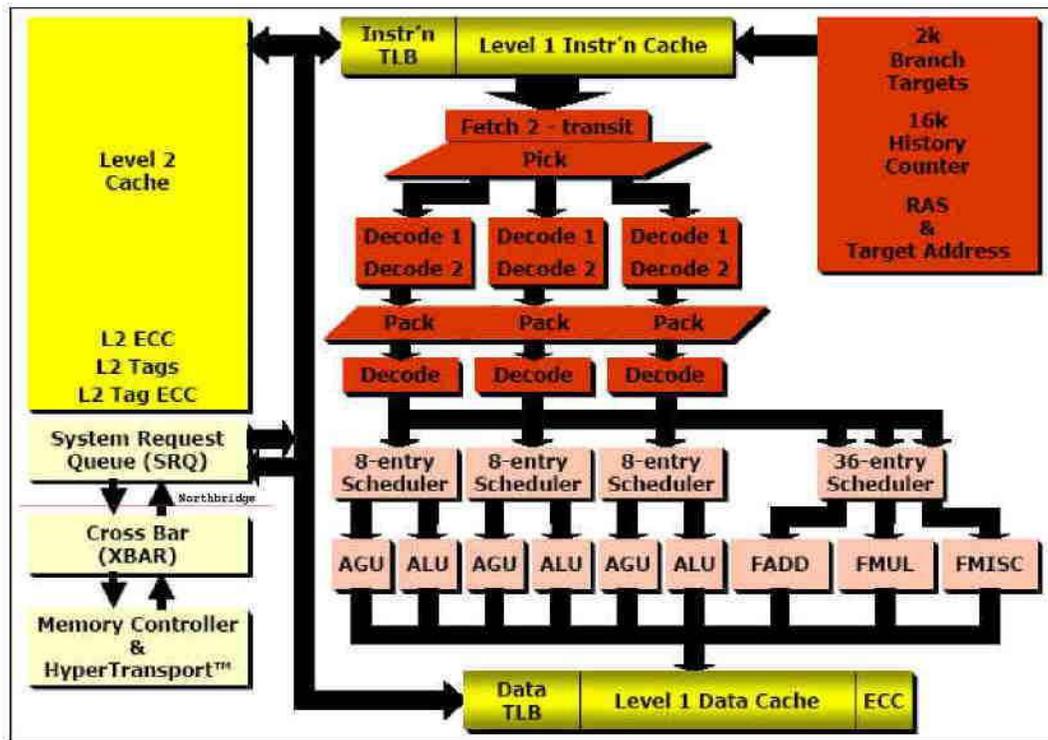
L'Hyper transport est le bus rapide permettant la connexion entre le northbridge et le southbridge. Le nombre de canaux est de 1 pour les Athlons 64 et Athlons 64FX. Pour l'opteron, il est de 3 canaux. Chaque canal est capable de transférer 3,2 GB/s dans chaque sens (6,4 GB/s au total).

Tout ceci explique notamment les différents socket pour la famille Athlon 64

- Athlon 64: **socket 754**, évolué mi- 2004 vers le socket 939 qui gère le Dual Channel

- Athlon 64 FX: **socket 940** (identique à l'opteron), évolué mi- 2004 vers le socket 939 (identique à l'Athlon64). Les premiers Athlon FX51 étaient donc des Opteron avec 2 des trois canaux hyper transports désactivés. L'athlon FX53 redevient un Athlon 64 (qu'il remplace) avec un cache plus important.
- Athlon Opteron: socket 940
- Le **socket AM2** est spécifique aux versions utilisant de la mémoire DDR-2. Début 2008, il est remplacé sur les processeurs de haute gamme par l'AM2+, totalement compatible, un processeur utilisant la version + accepte une carte mère en version standard et vis versa. La différence vient du bus hypertransport qui passe de la version 2 (1,6 Ghz maximum) à la version 3.0 (2,6 Ghz maximum), à condition que le FSB du microprocesseur l'accepte bien sûr.

## Architecture interne d'un Athlon 64

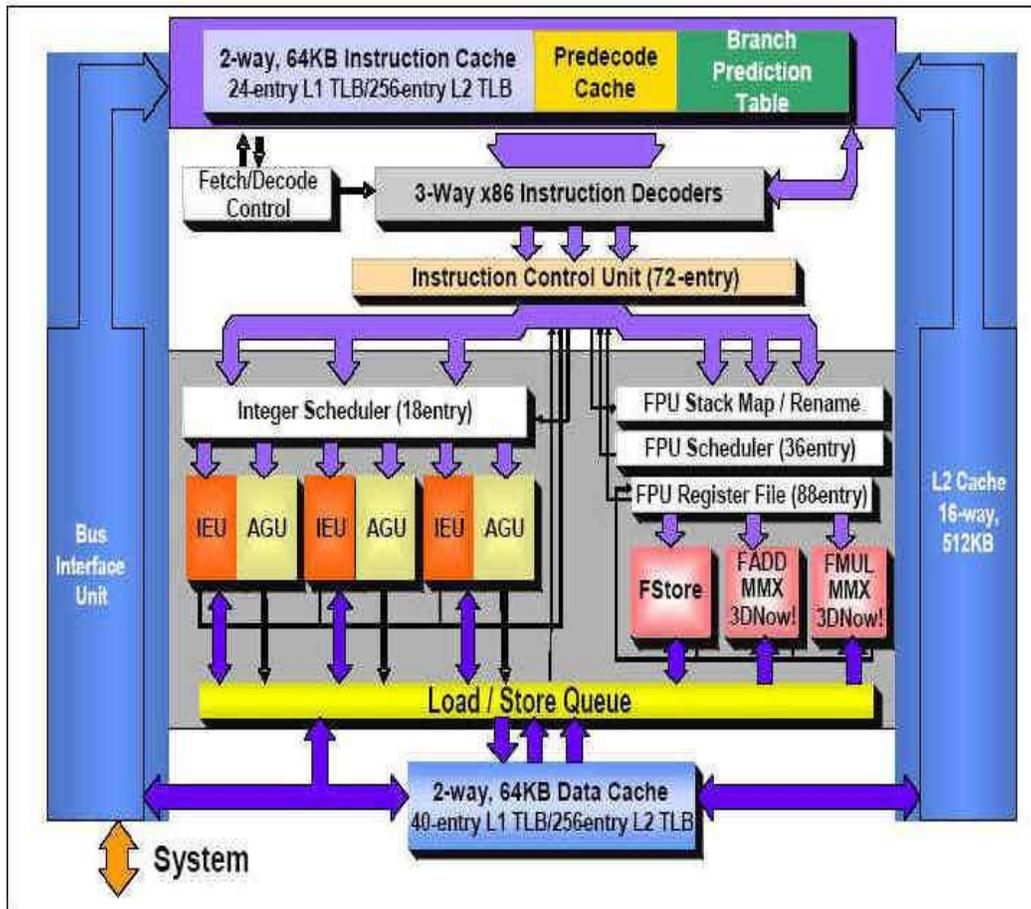


L'architecture d'un microprocesseur Athlon 64 n'est pas en soit révolutionnaire. Les données sont envoyées et reçues de la mémoire via le contrôleur Hyper-transport et contrôleur mémoire à gauche. Elles sont envoyées soit le réservoir d'instruction. Le cache d'instructions L1 raccordé au circuit de prédiction de branchement permet également de fournir des instructions au processeur.

La partie en rouge rassemble le décodage du CISC vers RISC avec 3 niveaux de décodage. Les diverses parties internes ne sont pas différentes des processeurs que nous connaissons: unités de calculs ALU, AGU (Adress Generation Unit) pour les calculs sur des adresses, FADD (unité de calcul en virgule flottante), ...

Les données sont ensuite renvoyées soit vers le cache L1 correspondant, soit vers la mémoire.

Dans un sens, l'architecture interne des Athlons 64 (en rose) n'est pas très différente de celle des Athlons XP ci-dessus (on s'en doutait un peu) mais ça devrait évoluer.



On retrouve de nouveaux d'un côté le bus de commande, données et adresses (Bus Interface Unit) et les données provenant du cache L2 (inclus dans le processeur). La partie au-dessus sert à prédécoder les instructions et inclut la table de prédiction de branchement. La partie suivante rassemble le passage des instructions CISC vers RISC et le transfert vers les différentes unités internes de l'Athlon XP.

### Les connecteurs de mémoire vive

La mémoire vive (RAM pour *Random Access Memory*) permet de stocker des informations pendant tout le temps de fonctionnement de l'ordinateur, son contenu est par contre détruit dès lors que l'ordinateur est éteint ou redémarré, contrairement à une mémoire de masse telle que le disque dur, capable de garder les informations même lorsqu'il est hors tension. On parle de « volatilité » pour désigner ce phénomène.

Pourquoi alors utiliser de la mémoire vive alors que les disques durs reviennent moins chers à capacité égale ? La réponse est que la mémoire vive est extrêmement rapide par comparaison aux périphériques de stockage de masse tels que le disque dur. Elle possède en effet un temps de réponse de l'ordre de quelques dizaines de nanosecondes (environ 70 pour la DRAM, 60 pour la RAM EDO, et 10 pour la SDRAM voire 6 ns sur les SDRAM DDR) contre quelques millisecondes pour le disque dur.

La mémoire vive se présente sous la forme de barrettes qui se branchent sur les connecteurs de la carte mère.

## Types de mémoires vives

On distingue généralement deux grandes catégories de mémoires vives :

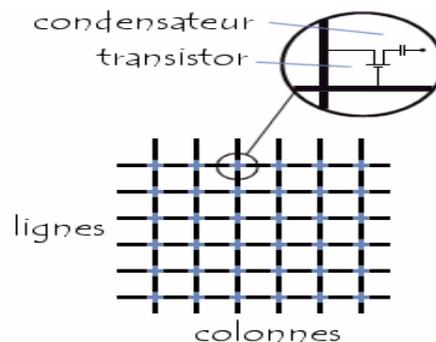
- Les **mémoires dynamiques (DRAM, *Dynamic Random Access Module*)**, peu coûteuses. Elles sont principalement utilisées pour la mémoire centrale de l'ordinateur ;
- Les **mémoires statiques (SRAM, *Static Random Access Module*)**, rapides et onéreuses. Les SRAM sont notamment utilisées pour les mémoires cache du processeur ;

## Fonctionnement de la mémoire vive

La mémoire vive est constituée de centaines de milliers de petits condensateurs emmagasinant des charges. Lorsqu'il est chargé, l'état logique du condensateur est égal à 1, dans le cas contraire il est à 0, ce qui signifie que chaque condensateur représente un bit de la mémoire.

Etant donné que les condensateurs se déchargent, il faut constamment les recharger (le terme exact est *rafraîchir*, en anglais *refresh*) à un intervalle de temps régulier appelé **cycle de rafraîchissement**. Les mémoires DRAM nécessitent par exemple des cycles de rafraîchissement est d'environ 15 nanosecondes (ns).

Chaque condensateur est couplé à un transistor (de type *MOS*) permettant de « récupérer » ou de modifier l'état du condensateur. Ces transistors sont rangés sous forme de tableau (matrice), c'est-à-dire que l'on accède à une *case mémoire* (aussi appelée *point mémoire*) par une ligne et une colonne.



Chaque point mémoire est donc caractérisé par une adresse, correspondant à un numéro de ligne (en anglais *row*) et un numéro de colonne (en anglais *column*). Or cet accès n'est pas instantané et s'effectue pendant un délai appelé **temps de latence**. Par conséquent l'accès à une donnée en mémoire dure un temps égal au temps de cycle auquel il faut ajouter le temps de latence.

Ainsi, pour une mémoire de type DRAM, le temps d'accès est de 60 nanosecondes (35ns de délai de cycle et 25 ns de temps de latence). Sur un ordinateur, le temps de cycle correspond à l'inverse de la fréquence de l'horloge, par exemple pour un ordinateur cadencé à 200 MHz, le temps de cycle est de 5 ns ( $1/(200 \cdot 10^6)$ ).

Par conséquent un ordinateur ayant une fréquence élevée et utilisant des mémoires dont le temps d'accès est beaucoup plus long que le temps de cycle du processeur doit effectuer des **cycles d'attente** (en anglais *wait state*) pour accéder à la mémoire. Dans le cas d'un ordinateur cadencé à 200 MHz utilisant des mémoires de types DRAM (dont le temps d'accès est de 60ns), il y a 11

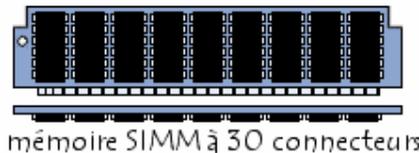
cycles d'attente pour un cycle de transfert. Les performances de l'ordinateur sont d'autant diminuées qu'il y a de cycles d'attentes, il est donc conseillé d'utiliser des mémoires plus rapides.

### Formats de barrettes de mémoire vive

Il existe de nombreux types de mémoires vives. Celles-ci se présentent toutes sous la forme de barrettes de mémoire enfichables sur la carte-mère.

Les premières mémoires se présentaient sous la forme de puces appelées *DIP (Dual Inline Package)*. Désormais les mémoires se trouvent généralement sous la forme de barrettes, c'est-à-dire des cartes enfichables dans des connecteurs prévus à cet effet. On distingue habituellement trois types de barrettes de RAM :

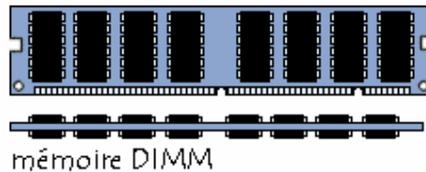
- les barrettes au format **SIMM** (*Single Inline Memory Module*) : il s'agit de circuits imprimés dont une des faces possède des puces de mémoire. Il existe deux types de barrettes SIMM, selon le nombre de connecteurs :
  - Les barrettes SIMM à 30 connecteurs (dont les dimensions sont 89x13mm) sont des mémoires 8 bits qui équipaient les premières générations de PC (286, 386).



- Les barrettes SIMM à 72 connecteurs (dont les dimensions sont 108x25mm) sont des mémoires capables de gérer 32 bits de données simultanément. Ces mémoires équipent des PC allant du 386DX aux premiers Pentium. Sur ces derniers le processeur travaille avec un bus de données d'une largeur de 64 bits, c'est la raison pour laquelle il faut absolument équiper ces ordinateurs de deux barrettes SIMM. Il n'est pas possible d'installer des barrettes 30 broches sur des emplacements à 72 connecteurs dans la mesure où un détrompeur (encoche au centre des connecteurs) en empêche l'enfichage.



- les barrettes au format **DIMM** (*Dual Inline Memory Module*) sont des mémoires 64 bits, ce qui explique pourquoi il n'est pas nécessaire de les appairer. Les barrettes DIMM possèdent des puces de mémoire de part et d'autre du circuit imprimé et ont également 84 connecteurs de chaque côté, ce qui les dote d'un total de 168 broches. En plus de leurs dimensions plus grandes que les barrettes SIMM (130x25mm) ces barrettes possèdent un second détrompeur pour éviter la confusion.



Il peut être intéressant de noter que les connecteurs DIMM ont été améliorés afin de faciliter leur insertion grâce à des leviers situés de part et d'autre du connecteur.

Il existe en outre des modules de plus petite taille, appelés **SO DIMM** (*Small Outline DIMM*), destinés aux ordinateurs portables. Les barrettes *SO DIMM* comportent uniquement 144 broches pour les mémoires 64 bits et 77 pour les mémoires 32 bits.

- les barrettes au format **RIMM** (*Rambus Inline Memory Module*, appelées également *RD-RAM* ou *DRD-RAM*) sont des mémoires 64 bits développée par la société Rambus. Elles possèdent 184 broches. Ces barrettes possèdent deux encoches de repérage (détrompeurs), évitant tout risque de confusion avec les modules précédents.

Compte tenu de leur vitesse de transfert élevée, les barrettes RIMM possèdent un film thermique chargé d'améliorer la dissipation de la chaleur.

Comme dans le cas des DIMM, il existe des modules de plus petite taille, appelés **SO RIMM** (*Small Outline RIMM*), destinés aux ordinateurs portables. Les barrettes *SO RIMM* comportent uniquement 160 broches.

## DRAM PM

La **DRAM** (*Dynamic RAM*, RAM dynamique) est le type de mémoire le plus répandu au début du millénaire. Il s'agit d'une mémoire dont les transistors sont rangés dans une matrice selon des lignes et des colonnes. Un transistor, couplé à un condensateur donne l'information d'un bit. 1 octet comprenant 8 bits, une barrette de mémoire DRAM de 256 Mo contiendra donc  $256 * 2^{10} * 2^{10} = 256 * 1024 * 1024 = 268\,435\,456$  octets =  $268\,435\,456 * 8 = 2\,147\,483\,648$  bits = 2 147 483 648 transistors. Une barrette de 256 Mo possède ainsi en réalité une capacité de 268 435 456 octets, soit 268 Mo ! Ce sont des mémoires dont le temps d'accès est de 60 ns.

D'autre part, les accès mémoire se font généralement sur des données rangées consécutivement en mémoire. Ainsi le mode d'accès en *rafale* (**burst mode**) permet d'accéder aux trois données consécutives à la première sans temps de latence supplémentaire. Dans ce mode en rafale, le temps d'accès à la première donnée est égal au temps de cycle auquel il faut ajouter le temps de latence, et le temps d'accès aux trois autres données est uniquement égal aux temps de cycle, on note donc sous la forme X-Y-Y-Y les quatre temps d'accès, par exemple la notation 5-3-3-3 indique une mémoire pour laquelle 5 cycles d'horloge sont nécessaires pour accéder à la première donnée et 3 pour les suivantes.

## DRAM FPM

Pour accélérer les accès à la DRAM, il existe une technique, appelée **pagination** consistant à accéder à des données situées sur une même colonne en modifiant uniquement l'adresse de la ligne, ce qui permet d'éviter la répétition du numéro de colonne entre la lecture de chacune des lignes. On parle alors de **DRAM FPM** (*Fast Page Mode*). La FPM permet d'obtenir des temps

d'accès de l'ordre de 70 à 80 nanosecondes pour une fréquence de fonctionnement pouvant aller de 25 à 33 Mhz.

### **DRAM EDO**

La **DRAM EDO** (*Extended Data Out*, soit *Sortie des données améliorée* parfois également appelé "*hyper-page*") est apparue en 1995. La technique utilisée avec ce type de mémoire consiste à adresser la colonne suivante pendant la lecture des données d'une colonne. Cela crée un chevauchement des accès permettant de gagner du temps sur chaque cycle. Le temps d'accès à la mémoire EDO est donc d'environ 50 à 60 nanosecondes pour une fréquence de fonctionnement allant de 33 à 66 Mhz.

Ainsi, la RAM EDO, lorsqu'elle est utilisée en mode rafale permet d'obtenir des cycles de la forme 5-2-2-2, soit un gain de 4 cycles sur l'accès à 4 données. Dans la mesure où la mémoire EDO n'acceptait pas des fréquences supérieures à 66 Mhz, elle a disparu au bénéfice de la SDRAM.

### **SDRAM**

La **SDRAM** (*Synchronous DRAM*, traduisez *RAM synchrone*), apparue en 1997, permet une lecture des données synchronisée avec le bus de la carte-mère, contrairement aux mémoires EDO et FPM (qualifiées d'*asynchrones*) possédant leur propre horloge. La SDRAM permet donc de s'affranchir des temps d'attente dus à la synchronisation avec la carte-mère. Celle-ci permet d'obtenir un cycle en mode rafale de la forme 5-1-1-1, c'est-à-dire un gain de 3 cycles par rapport à la RAM EDO. De cette façon la SDRAM est capable de fonctionner avec une cadence allant jusqu'à 150 Mhz, lui permettant d'obtenir des temps d'accès d'environ 10 ns.

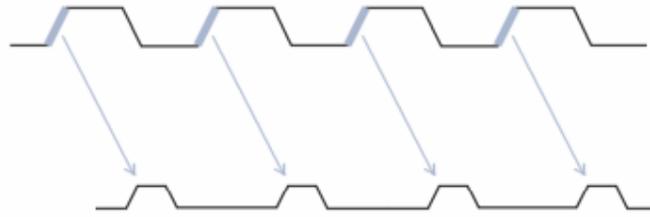
### **DR-SDRAM (Rambus DRAM)**

La **DR-SDRAM** (*Direct Rambus DRAM* ou encore *RDRAM*) est un type de mémoire permettant de transférer les données sur un bus de 16 bits de largeur à une cadence de 800Mhz, ce qui lui confère une bande passante de 1,6 Go/s. Comme la SDRAM, ce type de mémoire est synchronisé avec l'horloge du bus pour améliorer les échanges de données. En contrepartie, la mémoire RAMBUS est une technologie propriétaire, ce qui signifie que toute entreprise désirent construire des barrettes de RAM selon cette technologie doit reverser des droits (royalties) aux sociétés RAMBUS et Intel.

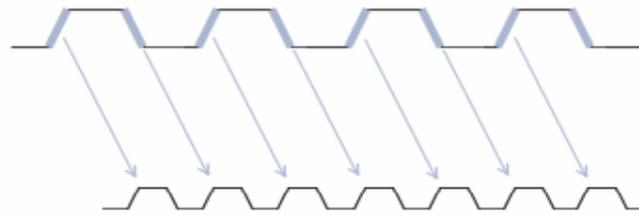
### **DDR-SDRAM**

La **DDR-SDRAM** (*Double Data Rate SDRAM*) est une mémoire basée sur la technologie SDRAM, permettant de doubler le taux de transfert de la SDRAM à fréquence égale.

La lecture ou l'écriture de données en mémoire est réalisé sur la base d'une horloge. Les mémoires DRAM standard utilisent une méthode appelé **SDR** (*Single Data Rate*) consistant à lire ou à écrire une donnée à chaque front montant.



La DDR permet de doubler la fréquence des lectures/écritures, avec une horloge cadencée à la même fréquence, en envoyant les données à chaque front montant, ainsi qu'à chaque front descendant.

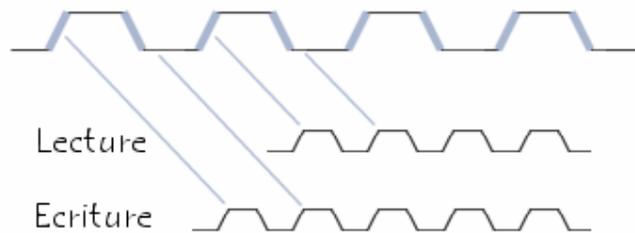


Les mémoires DDR possèdent généralement une appellation commerciale du type PCXXXX où «XXXX» représente le débit en Mo/s.

### DDR2-SDRAM

La mémoire DDR2 (ou DDR-II) permet d'atteindre des débits deux fois plus élevés que la DDR à fréquence externe égale.

On parle de QDR (*Quadruple Data Rate* ou *quad-pumped*) pour désigner la méthode de lecture et d'écriture utilisée. La mémoire DDR2 utilise en effet deux canaux séparés pour la lecture et pour l'écriture, si bien qu'elle est capable d'envoyer ou de recevoir deux fois plus de données que la DDR.



La DDR2 possède également un plus grand nombre de connecteurs que la DDR classique (240 pour la DDR2 contre 184 pour la DDR).

### Tableau récapitulatif

Le tableau ci-dessous donne la correspondance entre la fréquence de la carte-mère (FSB), celle de la mémoire (RAM) et son débit :

Mémoire	Appellation	Fréquence (RAM)	Fréquence (FSB)	Débit
---------	-------------	-----------------	-----------------	-------

DDR200	PC1600	200 MHz	100 MHz	1,6 Go/s
DDR266	PC2100	266 MHz	133 MHz	2,1 Go/s
DDR333	PC2700	333 MHz	166 MHz	2,7 Go/s
DDR400	PC3200	400 MHz	200 MHz	3,2 Go/s
DDR433	PC3500	433 MHz	217 MHz	3,5 Go/s
DDR466	PC3700	466 MHz	233 MHz	3,7 Go/s
DDR500	PC4000	500 MHz	250 MHz	4 Go/s
DDR533	PC4200	533 MHz	266 MHz	4,2 Go/s
DDR538	PC4300	538 MHz	269 MHz	4,3 Go/s
DDR550	PC4400	550 MHz	275 MHz	4,4 Go/s
DDR2-400	PC2-3200	400 MHz	100 MHz	3,2 Go/s
DDR2-533	PC2-4300	533 MHz	133 MHz	4,3 Go/s
DDR2-667	PC2-5300	667 MHz	167 MHz	5,3 Go/s
DDR2-675	PC2-5400	675 MHz	172,5 MHz	5,4 Go/s
DDR2-800	PC2-6400	800 MHz	200 MHz	6,4 Go/s

### Synchronisation (timings)

Il n'est pas rare de voir des notations du type 3-2-2-2 ou 2-3-3-2 pour décrire le paramétrage de la mémoire vive. Cette suite de quatre chiffres décrit la synchronisation de la mémoire (en anglais *timing*), c'est-à-dire la succession de cycles d'horloge nécessaires pour accéder à une donnée stockée en mémoire vive. Ces quatre chiffres correspondent généralement, dans l'ordre, aux valeurs suivantes :

- **CAS delay** ou **CAS latency** (CAS signifiant *Column Address Strobe*) : il s'agit du nombre de cycles d'horloge s'écoulant entre l'envoi de la commande de lecture et l'arrivée effective de la donnée. Autrement dit, il s'agit du temps d'accès à une colonne.
- **RAS Precharge Time** (noté *tRP*, RAS signifiant *Row Address Strobe*) : il s'agit du nombre de cycles d'horloge entre deux instructions RAS, c'est-à-dire entre deux accès à une ligne. opération.
- **RAS to CAS delay** (noté parfois *tRCD*) : il s'agit du nombre de cycles d'horloge correspondant au temps d'accès d'une ligne à une colonne.
- **RAS active time** (noté parfois *tRAS*) : il s'agit du nombre de cycles d'horloge correspondant au temps d'accès à une ligne.

Les cartes mémoires sont équipées d'un dispositif appelé SPD (Serial Presence Detect), permettant au BIOS de connaître les valeurs nominales de réglage définies par le fabricant. Il s'agit d'une EEPROM dont les données seront chargées par le BIOS si l'utilisateur choisi le réglage « auto ».

### La correction d'erreurs

Certaines mémoires possèdent des mécanismes permettant de pallier les erreurs afin de garantir l'intégrité des données qu'elles contiennent. Ce type de mémoire est généralement utilisé sur des systèmes travaillant sur des données critiques, c'est la raison pour laquelle on trouve ce type de mémoire dans les serveurs.

## Bit de parité

Les barrettes avec bit de parité permettent de s'assurer que les données contenues dans la mémoire sont bien celles que l'on désire. Pour ce faire, un des bits de chaque octet stocké en mémoire sert à conserver la somme des bits de données.

Le bit de parité vaut 1 lorsque la somme des bits de données est impaire et 0 dans le cas contraire.

De cette façon les barrettes avec bit de parité permettent de vérifier l'intégrité des données mais ne permettent pas de corriger les erreurs. De plus pour 9 Mo de mémoire, seulement 8 serviront à stocker des données, dans la mesure où le dernier mégaoctet conservera les bits de parité.

## Barrettes ECC

Les barrettes de mémoire ECC (*Error Correction Coding*) sont des mémoires possédant plusieurs bits dédiés à la correction d'erreur (on les appelle ainsi *bits de contrôle*). Ces barrettes, utilisées principalement dans les serveurs, permettent de détecter les erreurs et de les corriger.

## Dual Channel

Certains contrôleurs mémoire proposent un double canal (en anglais *Dual Channel*) pour la mémoire. Il s'agit d'exploiter les modules de mémoire par paire afin de cumuler la bande passante et ainsi exploiter au maximum les capacités du système. Il est essentiel, lors de l'utilisation du Dual Channel, d'utiliser des barrettes identiques par paire (fréquence, capacité et préférentiellement de même marque).

## Récapitulatif

Type de mémoire vive	nombre de bits (+ parité éventuelle)	bande passante	Processeurs	installation
simm 32C	8 (9)		286, 386SX (16b)	par 2
			386DX, 486 (32b)	par 4
simm 70c	32 (36)		486 (32b)	par 1
			Pentium (64b)	par 2
FPM		176 MB/s		
EDO	32	264 MB/s	Pentium (64b)	par 2
SDram 66 Mhz	64	528 Mhz	Pentium	par 1
			Pentium II, celeron	par 1
SDRam 100 Mhz	64	800 MB/s	Pentium II 100 Mhz	par 1
SDRam 133 Mhz	64	1,06 GB/s	Pentium III (Via chipset VIA), Athlon	par 1
RamBus	64	1,6 GB (version 800)	Pentium III	par 1
DDR	64		Pentum IV, Athlon, Opteron	par 1

Remarque: dans ce tableau, Pentium, cyrix M6, K6-2 et K6-3 utilisent les mêmes mémoires

## Mémoire spéciale ordinateur portable.

Les deux types de mémoire ci-dessous sont utilisées uniquement dans un ordinateur portable. Certains ordinateurs portables utilisent également les différents types de mémoires standards.



La **So-Dimm 144** contact est équivalent à la SDRam



La **So-Dimm 200** contacts est équivalent à la DDR.

Un nouveau modèle est également utilisé, la So-Dimm DDR-2, équivalent à la mémoire DDR2.

### Mémoires avec correction d'erreurs

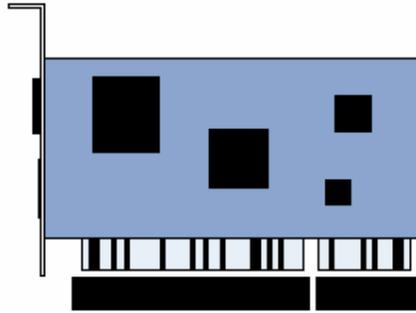
La première méthode (historique) de correction d'erreur mémoire est le contrôle de parité. Cette méthode est également utilisée dans les liaisons séries. Un neuvième bit mémoire est donc ajouté pour ce contrôle. Dans le cas des ordinateurs PC compatibles, une erreur de parité est associée à une interruption non masquable. Cette interruption provoque l'arrêt de la machine. Les mémoires avec bit de parité ne sont plus utilisées dans les PC bureautiques. Pour un contrôle de parité, la mémoire doit être sur 9 bits et que l'option doit être activée dans le BIOS.

Actuellement, le contrôle d'erreur mémoire est utilisé dans les serveurs avec des mémoires de type auto-correctives ECC (Error Checking and Correcting) ou AECC (Advanced ECC) relativement chères qui détectent et corrigent jusque 3 erreurs. La carte mère (ou plutôt le chipset) doit accepter cette fonction. Nous verrons celle-ci avec les spécificités serveurs PC en deuxième année.

### Les connecteurs d'extension

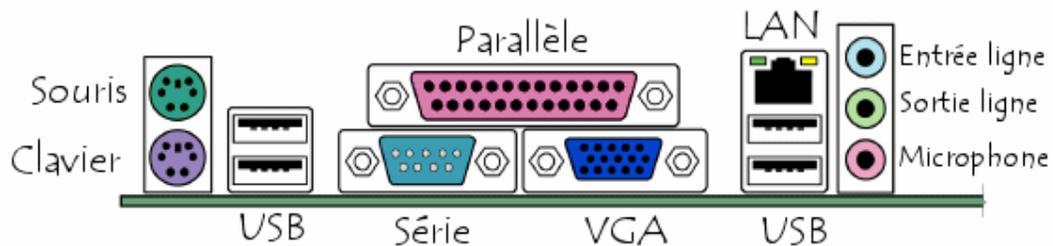
Les **connecteurs d'extension** (en anglais **slots**) sont des réceptacles dans lesquels il est possible d'insérer des cartes d'extension, c'est-à-dire des cartes offrant de nouvelles fonctionnalités ou de meilleures performances à l'ordinateur. Il existe plusieurs sortes de connecteurs :

- Connecteur ISA (*Industry Standard Architecture*) : permettant de connecter des cartes ISA, les plus lentes fonctionnant en 16-bit
- Connecteur VLB (*Vesa Local Bus*): Bus servant autrefois à connecter des cartes graphiques
- Connecteur PCI (*Peripheral Component InterConnect*) : permettant de connecter des cartes PCI, beaucoup plus rapides que les cartes ISA et fonctionnant en 32-bit
- Connecteur AGP (*Accelerated Graphic Port*): un connecteur rapide pour carte graphique.
- Connecteur PCI Express (*Peripheral Component InterConnect Express*) : architecture de bus plus rapide que les bus AGP et PCI.
- Connecteur AMR (*Audio Modem Riser*): ce type de connecteur permet de brancher des mini-cartes sur les PC en étant équipés



### Les connecteurs d'entrée-sortie

La carte mère possède un certain nombre de connecteurs d'entrées-sorties regroupés sur le « **panneau arrière** ».



La plupart des cartes mères proposent les connecteurs suivants :

- Port série, permettant de connecter de vieux périphériques ;
- Port parallèle, permettant notamment de connecter de vieilles imprimantes ;
- Ports USB (1.1, bas débit, ou 2.0, 2.2, haut débit), permettant de connecter des périphériques plus récents ;
- **Connecteur RJ45** (appelés *LAN* ou *port ethernet*) permettant de connecter l'ordinateur à un réseau. Il correspond à une carte réseau intégrée à la carte mère ;
- **Connecteur VGA** (appelé *SUB-D15*), permettant de connecter un écran. Ce connecteur correspond à la carte graphique intégrée ;
- **Prises audio** (*entrée Line-In, sortie Line-Out* et *microphone*), permettant de connecter des enceintes acoustiques ou une chaîne hi fi, ainsi qu'un microphone. Ce connecteur correspond à la carte son intégrée.

### Les bus internes des PC

Dans l'introduction aux systèmes à base de microprocesseurs, nous avons distingué les bus: bus d'adresse, bus de données et bus de commande. Avec le développement des ordinateurs de type PC, on a étendu la dénomination bus à tous les connecteurs internes et parfois externes reliés au système microprocesseur (y compris à ses circuits d'interfaçage) permettant d'insérer des cartes informatiques (carte graphique, séries, parallèle, SCSI, son, USB, firewire, réseau, ...).

Un bus interne est un connecteur permettant d'insérer des cartes électroniques périphériques. Nous les rencontrerons principalement lors des exercices pratiques. Chaque bus (connecteur) est normalisé en:

- **mécanique:** la forme et la taille, le nombre de broches du connecteur sont définis à l'avance
- **en vitesse.** La fréquence des signaux est définie.
- **en signaux.** Les timing des signaux de commandes, données et adresses transitant sur le bus sont clairement définies, y compris les temps d'attente, ...

### Le bus ISA

Le **bus ISA** est apparu avec les premiers PC XT d'IBM en 1981. A l'époque, le nombre de bits de données était de 8 bits (comme le processeur en externe) et la fréquence de travail était de 4,77 Mhz. Avec le 286, IBM a sorti le même, mais les caractéristiques changent puisque le bus passe en version 16 bits (bus de données) avec une vitesse de 8 Mhz. Cette vitesse est toujours celle utilisée actuellement, la bande passante d'un bus ISA 16 bits est de 16 MB / secondes. Une carte 8 bits peut s'insérer dans un bus ISA 16 bits. Fin 1999, seuls quelques modems, des cartes de rajoute port parallèle ou série, ... utilisent encore ce type de connecteur.

Les cartes mères actuelles n'incluent plus de bus ISA, remplacé par le bus PCI plus rapide et Plug&Play.

### Le bus MCA d'IBM

Le **Bus MCA** est apparu avec les PC 386 d'IBM. Celui-ci n'avait prévu aucun brevet lors de la sortie des XT et les firmes concurrentes utilisaient leurs données techniques. Avec l'introduction des 386, IBM implantait un nouveau bus de 32 bits de données. Son débit maximum est de 40 MB / s. Comme IBM avait breveté l'ensemble des normes mécaniques et électriques, seule cette firme l'a réellement utilisé.

### Le bus VLB

A partir du 486DX2-66 est apparu le **bus VLB ou VESA Local bus**. Celui-ci comportait maximum 3 connecteurs, rajoute sur un bus ISA 16 bits. Connecté directement sur le processeur et reprenant directement les signaux, celui-ci ne pouvait que disparaître avec les 486. Attention, installer 3 cartes VLB pose généralement des problèmes. Le bus VLB se décline en 2 versions: le A et le B. Le B est relié au processeur via un circuit d'interface plus évolué qu'un simple buffer. Les performances sont identiques.

On trouvait deux types de cartes: les cartes écrans et les cartes IO (2 connecteurs disques durs E-IDE, port parallèle, 2 port série et contrôleur disquette).

### Le bus PCI

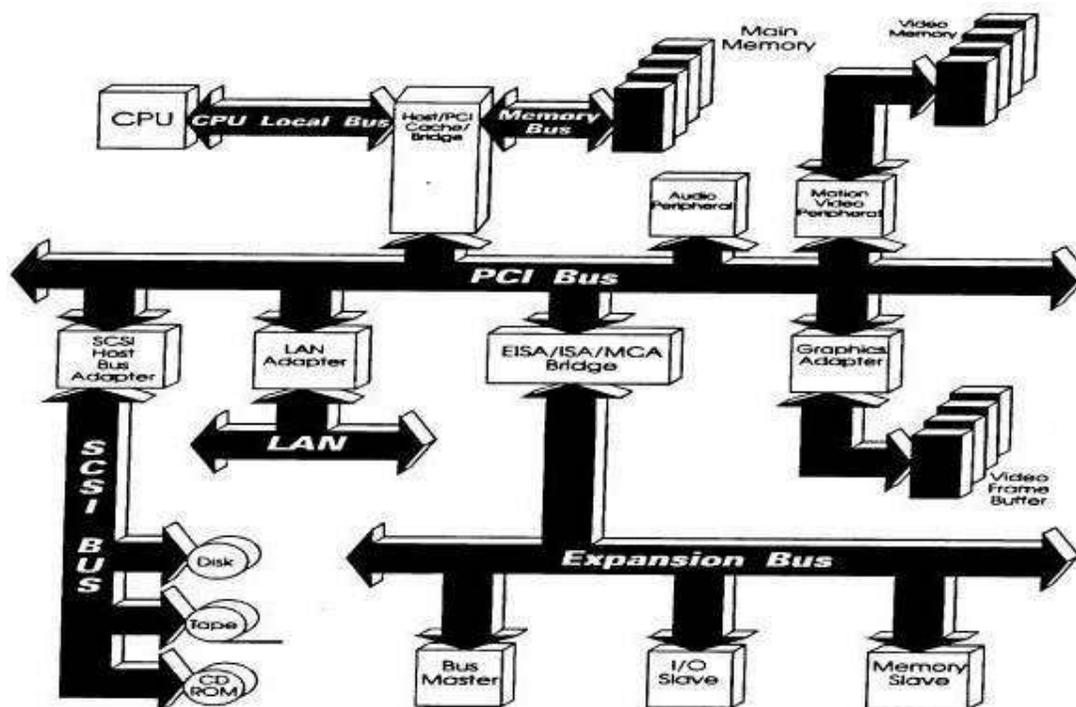
Avec les premiers PENTIUM, le bus d'extension est devenu clairement un goulot d'étranglement avec sa vitesse limitée à 8 Mhz sur un bus de donnée de 16 bits. Son successeur est le **bus PCI**, cadencé à 33 Mhz sur une largeur de 32 bits (soit 4 byte - octet). Celui-ci était au début utilisé par la carte graphique, mais les modems, carte son, cartes réseaux, ...utilisent ce bus au détriment du bus ISA obsolète. Sa bande passante est de 132 MB/s, soit 8 fois supérieure à celle du bus ISA. Une version supérieure, le PCI-X, est développée pour les serveurs.

Les cartes PCI peuvent être alimentées en 5 V (à l'origine) ou en 3,3 V. La tension d'alimentation est réglée directement lors du démarrage du PC. Les cartes fonctionnant en 3.3 V ont 2 encoches. Ceci ne pose pas de problèmes dans ce sens. Par contre, certains slots de cartes

mères n'acceptent pas les cartes 5 V, uniquement 3,3V. Dans ce cas, le connecteur de la carte mère inclut les 2 encoches, vous ne pouvez pas insérer une carte 5V. De même, certaines anciennes cartes mères (Pentium) acceptent les cartes 5V mais ne détectent pas correctement les cartes 3,3V. Généralement, cela se passe néanmoins sans problèmes.

Le bus PCI a la particularité de permettre à un périphérique connecté sur ce bus de prendre le contrôle et de transférer directement des données vers un autre périphérique PCI, la mémoire principale ou même vers le bus AGP. Ceci s'appelle le DMA (Direct Memory Access). Cette dernière possibilité est parfois limitée par le circuit d'interface (chipset).

Les schémas d'un système à bus PCI (ce bus n'est pas limité aux PC, également en Mac) ci-dessous apporte quelques renseignements intéressants. La fonction représentée entre le processeur et le bus est ce que nous appellerons le chipset. Il gère le bus PCI et la mémoire. Les autres bus sont liés au bus PCI par un circuit qui sert de pont. Toutes les transactions actuelles entre le bus ISA et le processeur passent par le processeur.



Architecture du bus PCI

### Bus PCI-X

Développé pour les serveurs conjointement par IBM, Compaq, HP et Intel, le bus PCI-X est une évolution du bus PCI standard. Si un bus PCI tourne à 33 Mhz avec une largeur de bus de donnée de 32 bits (soit une bande passante maximum de 132 MB/s), le bus PCI-X tourne à une fréquence d'horloge de 66 à 533 Mhz suivant les versions en 32 ou 64 bits. Les taux de transferts associés à ces vitesses sont nécessaires pour les cartes réseaux Gigabits, mais surtout pour les connections SCSI Ultra 160 et Ultra 320 (respectivement débit maximum de 160 MB/s et 320 MB/s)

La version PCI-X 1.0 est sortie en 1999, elle est alimentée à une tension de 3,3 V. 6 versions sont reprises dans la norme

Fréquence du bus PCI-X 1.0	Tension	largeur du bus de données	bande passante
66 Mhz	3,3 V	32 bits	264 MB/s
	3,3 V	64 bits	528 MB/s
100 Mhz	3,3 V	32 bits	400 MB/s
		64 bits	800 MB/s
133 Mhz	3,3 V	32 bits	532 MB/s
		64 bits	1064 Mb/s

La version PCX-2.0, sortie en 2002, est également alimentée en 1,5 V suivant les versions. Les cartes sont hot Plug (déconnectables à chaud).

Fréquence du bus	Tension	largeur du bus de données	bande passante
66 Mhz	3,3 V	32 bits	264 MB/s
	3,3 V	64 bits	528 MB/s
100 Mhz	3,3 V	32 bits	400 MB/s
		64 bits	800 MB/s
133 Mhz	3,3 V	32 bits	532 MB/s
		64 bits	1064 MB/s
266 Mhz	3,3 V / 1,5 V	32 bits	1064 MB/s
		64 bits	2128 MB/s
533 Mhz	3,3 V / 1,5 V	32 bits	2128 MB/s
		64 bits	4256 MB/s

De par la conception du connecteur, les cartes PCI-X 32 bits peuvent être insérées dans 1 bus 64 bits. Les bus PCI-X sont directement connectés sur le Southbridge du chipset, ce qui nécessite des cartes mères (chipset) spécifiques.

### Le bus AGP

Malgré ses progrès, le bus PCI est vite devenu trop lent pour les cartes graphiques. INTEL a développé le **bus AGP** spécialement pour celles-ci. La vitesse de ce bus est de 66 Mhz dans les premières versions. Il est capable par DMA (direct Memory Access) de prendre le contrôle de la mémoire pour le transfert direct des informations. Cette caractéristique est également implantée dans le bus PCI. Les spécifications d'INTEL proposent 4 modes de fonctionnement:

L'**AGP 1X**, avec un débit théorique de 266 MB/s sur 32 bits (vitesse de 66 Mhz), tension des signaux: 3,3 V ou 1,5 V

L'**AGP 2X**: celui-ci permet de charger les informations 1 fois sur la partie montante de l'horloge, une fois sur la partie descendante (doublement du débit, 533 MB/s) sur une fréquence de base de 66 Mhz. Tension des signaux sur le bus: 1,5 V Ce bus est géré par les circuits d'interface Intel 440 LX, EX, ZX et BX.

L'**AGP 4X** utilise le même bus - connecteur (mais à 133 Mhz) que ses prédécesseurs, comme en AGP 2X, on double les informations chargées sur les flancs montants et descendants de l'horloge. Ce type de bus implique que la mémoire vive soit suffisamment rapide. Il existe une spécification supérieure, extension de l'AGP 4X, dénommée AGP Pro qui est destinée aux futures stations de travail. La spécificité est liée à l'alimentation électrique de la carte (50W pour l'AGP Pro 50 et 110 W pour l'AGP 110 contre 25 W pour l'AGP 4X). Le taux de transfert atteint 1,07 GB/s avec une tension d'alimentation de la carte de 1,5 V.

Sorti réellement début 2003, l'**AGP 8X** reste en 32 bits mais à 266 Mhz. Seule réelle avancée, le transfert atteint en théorie jusqu'à 2,13 GB / s. En pratique, le gain de performance est inférieur à 5% par rapport à l'AGP 4X. L'AGP est en mesure de détecter le mode de fonctionnement AGP de la carte graphique. La tension des signaux de la carte graphique passe à 0,8V. Cette faible tension explique qu'une carte mère gérant l'AGP 8X peut accepter des cartes AGP 2X et 4X, mais pas une carte graphique AGP 1X de la première génération.

Par comparaison, le bus ISA ne faisait que 16 MB/s et le bus PCI 132 MB/s

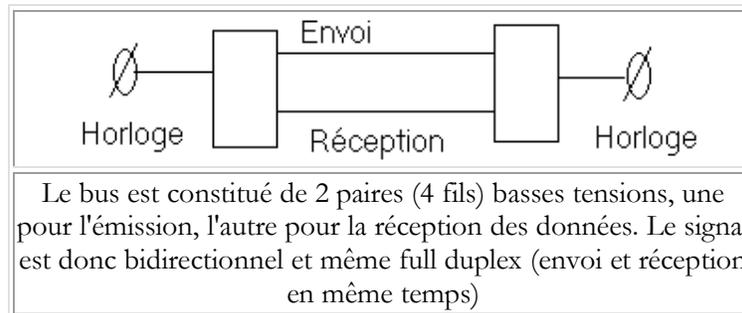
### PCI-Express 1.1 et 2.0

Sorti en 2004 sous l'impulsion d'INTEL, le bus **PCI-Express 1.1** est destiné à remplacer les bus AGP (cartes graphiques) et PCI pour les autres périphériques. Il utilise 32 canaux de type série. Chaque canal peut en théorie être utilisé indépendamment des autres. Le PCI-express 1.1 permet un débit maximum de 250 MB/s et par canal, la **version 2.0**, ratifiée en janvier 2007, double ce taux de transfert à 500 MB/seconde



Les connecteurs implantés sur les cartes mères permettent d'utiliser **1** ou **16** canaux. La version 16 canaux est dédiée aux cartes graphiques (connecteur noir sur la photo) et permet un taux de transfert maximum de 4 GB/s en 1.1 et 8 GB/s en 2.0. Le débit maximum d'un AGP 8 X reste limité à 2 GB/s. Le connecteur 1 X (les deux blancs sur la photo) est destiné à des cartes d'extensions comme des tuners TV ou cartes réseaux (même si ces dernières sont actuellement implantées directement dans le chipset). Elles sont finalement peu répandues. Deux versions, utilisant 4 et 8 canaux sont réservés pour les serveurs réseaux.

Une voie est composée de 2 paires de connexion, une pour la transmission, une pour la réception des données. Chaque paire (communication + masse) assure un débit unidirectionnel de 250 MB/s en émission ou en réception. En couplant 16 voies de communication (carte graphique), on obtient donc 4 GB/s.



Contrairement aux autres bus utilisés dans les PC, il se raccorde directement sur le northbridge du chipset (le composant qui gère les transactions entre le processeur et la mémoire), ce qui assure une meilleure connexion avec les mémoires du PC.

Comme tous bus séries, le contrôle des données est assuré directement par le contrôleur. Contrairement aux autres bus utilisés dans les PC bureautiques (pas les serveurs réseaux), les cartes sont déconnectables à chaud (hot plug)

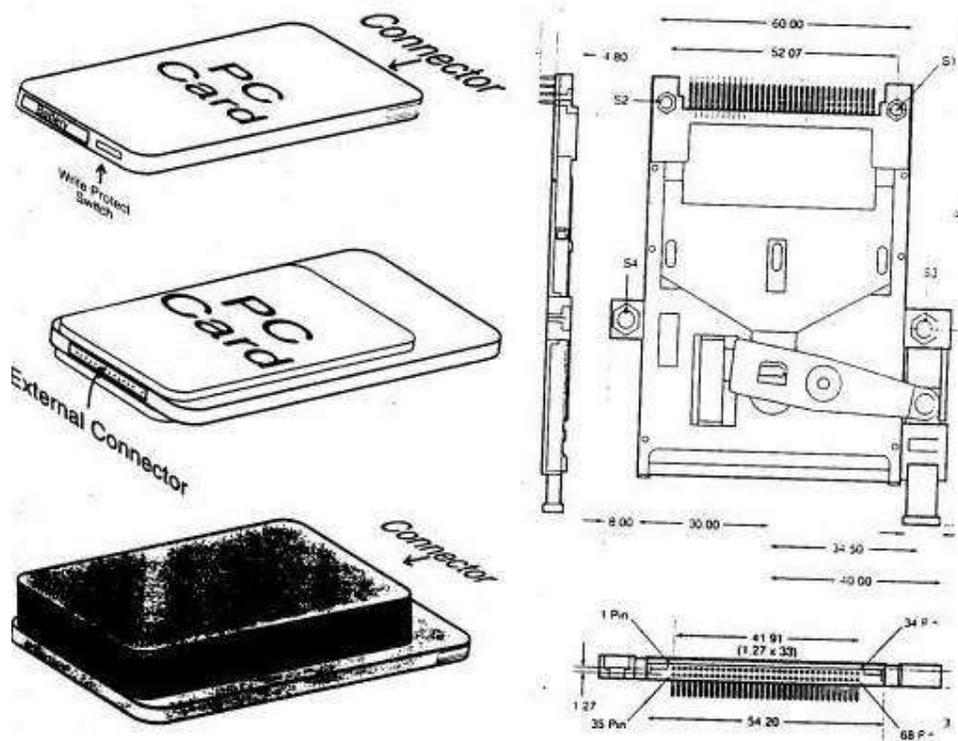
Une version pour ordinateur portable est prévue, appelée mini PCI Express. La dimension est de 3 cm de large (pour 8,5 cm dans le cas d'un bus PCMCIA)

Son concurrent chez AMD se nomme hypertransport, notamment avec la future version 3.0. Il accepte un débit théorique allant jusqu'à 22 GB/s (version actuelle 2.0). Il est actuellement implanté dans les chipsets AMD pour microprocesseur 64 bits comme inter-bridge mais sera disponible comme bus interne avec la future version 3.0. Nous en avons déjà parlé dans le chapitre sur les processeurs.

Une dernière remarque, le PCI-express est un bus interne, il ne remplace ni le serial ATA pour les disques durs ni les ports de communication externes comme l'USB ou IEEE 1394 Firewire. Quelques essais de connexion de périphériques externes sont en cours d'élaboration, mais sans réelles normes.

## Le bus PCMCIA

Le bus **PCMCIA (Personal Computer Memory Card International Association)** est utilisé dans un ordinateur portable comme bus interne. Quelques cartes pour PC bureautiques permettent d'insérer de telles cartes dans le PC. Ceci est utilisé notamment pour des "cartes de crédit" de reconnaissance personnelles anti-piratage.



Existent 3 versions PCMCIA I, II et III. 2 connecteurs PCMCIA I donnent un PCMCIA II, et 2 connecteurs II superposés peuvent également donner un bus PCMCIA III pour des disques durs par exemple. Le connecteur interne est donc le même, la seule différence vient de la hauteur disponible pour insérer le périphérique. Un emplacement PCMCIA II prend la place de 2 PCMCIA I. Un portable comportant un slot PCMCIA II intégrera automatiquement 2 slots PCMCIA I. On trouve sous ce format des modems, cartes réseaux, mémoires Flash, ... Les cartes PCMCIA peuvent être intégrées "à chaud", sous tension grâce à certaines précautions très particulières de la mécanique du bus. En effet, les bornes d'alimentations sont d'abord connectées lors de l'insertion de la carte, ensuite la quasi-totalité des signaux et pour terminer 2 broches destinées à la détection d'insertion.

Le PCMCIA I est réservé uniquement à l'usage de carte mémoires dans les notebook. Le type II, plus épais de 1 mm, autorise plus de choix: carte réseaux, modem, ... Le PCMCIA III, effectif début 2000, autorise l'usage de disques durs de 1,8 pouces.

Le bus de connexion est équipé d'un connecteur (broches femelles). Le réceptacle est solidaire du boîtier du système. L'extrémité des cartes est pourvue d'un socket (broches mâles). Le connecteur et le socket comportent 2 rangées de 34 contacts espacés de 1 mm. Les broches du réceptacle présentent 3 longueurs différentes. Les cartes sont alimentées par 2 broches Vcc de 5 V, 2 broches Vpp de 12 V et 4 broches de masse. Ces broches sont insérées **avant** les autres broches pour éviter des tensions sur les bus lors de l'insertion ou après lors de l'enlèvement de la carte.

**Dimensions physiques du bus PCMCIA:** la longueur de la carte est toujours de 85,6 mm et 54 mm de largeur.

Certaines cartes actuelles fonctionnent en 3,3V dans la release 2.0. Dans ce cas, le portable lit une mémoire interne à la carte avant d'alimenter les broches de la carte et transforme le 5 V en 3.3 V puis alimente les différentes broches de la carte.

Type	Epaisseur	Défini par
I	3,3 mm	release 1.0
II	5 mm	release 1.0
III	10,5 mm	release 2.0

Alors que les versions précédentes étaient sur 16 bits, la release 5.0 de la norme autorise des cartes 32 bits.

### Calculer la bande passante d'un bus informatique.

Le calcul du taux de transfert maximum d'un bus se calcule par la formule suivante:

Fréquence du bus (hz) \* largeur du bus (bit) / 8 **ou** Fréquence du bus \* largeur de bus en Byte

Soit par exemple un bus PCI 64 bits tournant à 64 Mhz. Le taux maximum (théorique) de transfert est:  $64\ 000\ 000 * 64\ \text{bits}/8 = 512.000.000\ \text{Byte/s}$ .

En informatique hardware, on utilise (à tort) des kilo de 1000. Les kilos doivent être de 1024,  $2^{10}$ . Pour exprimer le chiffre ci-dessus en MB/s (MO/s), il faut donc diviser par 1024 \* 1024. Le calcul ci-dessus donne un résultat de 488 MB/s.

### Le bus PCI Express

Le **bus PCI Express** (*Peripheral Component Interconnect Express*, noté *PCI-E* ou **3GIO** pour «*Third Generation I/O*»), est un bus d'interconnexion permettant l'ajout de cartes d'extension dans l'ordinateur. Le bus PCI Express a été mis au point en juillet 2002. Contrairement au bus PCI, qui fonctionne en interface parallèle, le bus PCI Express fonctionne en interface série, ce qui lui permet d'obtenir une bande passante beaucoup plus élevée que ce dernier.



### Caractéristiques du bus PCI Express

Le bus PCI Express se décline en plusieurs versions, 1X, 2X, 4X, 8X, 12X, 16X et 32X, permettant d'obtenir des débits compris entre 250 Mo/s et 8 Go/s, soit près de 4 fois le débit maximal des ports AGP 8X. Ainsi, avec un coût de fabrication similaire à celui du port AGP, le bus PCI Express est amené à le remplacer progressivement.

### Connecteurs PCI Express

Les connecteurs PCI Express sont incompatibles avec les anciens connecteurs PCI et possèdent des tailles variables et une plus faible consommation électrique. Une des caractéristiques intéressantes du bus PCI Express est la possibilité de brancher ou débrancher des composants à

chaud, c'est-à-dire sans éteindre ou redémarrer la machine. Les connecteurs PCI Express sont reconnaissables grâce à leur petite taille et leur couleur anthracite :

- Le connecteur PCI Express 1X possède 36 connecteurs et est destiné à un usage d'entrées-sorties à haut débit :



- Le connecteur PCI Express 4X possède 64 connecteurs et est destiné à un usage sur serveurs :



- Le connecteur PCI Express 8X possède 98 connecteurs et est destiné à un usage sur serveurs :



- Le connecteur PCI Express 16X possède 164 connecteurs, et mesure 89 mm de long et a vocation à servir de port graphique :



Le standard PCI Express a également vocation à supplanter la technologie PC Card sous la forme de connecteurs « PCI Express Mini Card ». De plus, contrairement aux connecteurs PCI dont l'usage était limité à la connectique interne, le standard PCI Express permet de connecter des périphériques externes à l'aide de câbles. Pour autant il ne se positionne pas en concurrence des ports USB ou FireWire.

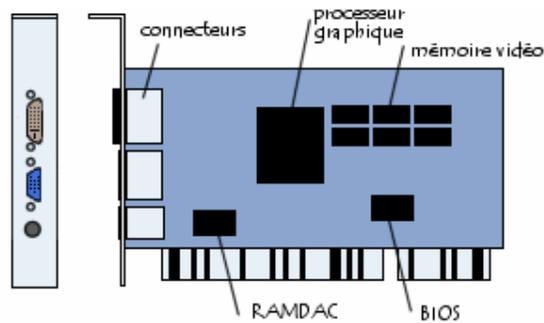
### Les cartes graphiques accélératrices 2D

La carte graphique (en anglais graphic adapter), parfois appelée carte vidéo ou **accélérateur graphique**, est l'élément de l'ordinateur chargé de convertir les données numériques à afficher en données graphiques exploitables par un périphérique d'affichage.

Le rôle de la carte graphique était initialement l'envoi de pixels graphique à un écran, ainsi qu'un ensemble de manipulation graphiques simples :

- déplacement des blocs (curseur de la souris par exemple) ;
- tracé de lignes ;
- tracé de polygones ;
- etc.

Les cartes graphiques récentes sont désormais équipées de processeurs spécialisés dans le calcul de scènes graphiques complexes en 3D !



Les principaux composants d'une carte vidéo sont :

- Un **processeur graphique** (appelé **GPU**, pour *Graphical Processing Unit*), constituant le cœur de la carte graphique et chargé de traiter les images en fonction de la résolution et de la profondeur de codage sélectionnée. Le **GPU** est ainsi un processeur spécialisé possédant des instructions évoluées de traitement de l'image, notamment de la 3D. En raison de la température que peut atteindre le processeur graphique, il est parfois surmonté d'un radiateur et d'un ventilateur.
- La **mémoire vidéo** chargée de conserver les images traitées par le processeur graphique avant l'affichage. Plus la quantité de mémoire vidéo est importante, plus la carte graphique pourra gérer de textures lors de l'affichage de scènes en 3D. On parle généralement de *frame buffer* pour désigner la partie de la mémoire vidéo servant à stocker les images avant affichage. Les cartes graphiques sont tributaires du type de mémoire utilisée sur la carte, car leur temps de réponse est déterminant pour la vitesse d'affichage des images, ainsi que de la quantité de mémoire, jouant sur le nombre et la résolution des images pouvant être stockées dans le frame buffer.
- Le **RAMDAC** (*random access memory digital-analog converter*) permet de convertir les images numériques stockées dans le *frame buffer* en signaux analogiques à envoyer au moniteur. La fréquence du **RAMDAC** détermine les taux de rafraîchissement (nombre d'images par seconde, exprimé en Hertz - Hz) que la carte graphique peut supporter.
- Le **BIOS vidéo** contient les paramètres de la carte graphique, notamment les modes graphiques que celle-ci supporte.
- L'**interface** : Il s'agit du type de bus utilisé pour connecter la carte graphique à la carte mère. Le bus AGP est ainsi spécialement prévu pour accepter des débits importants de données, nécessaire pour l'affichage de séquences vidéo ou 3D. Le bus PCI Express possède de meilleures performances que le bus AGP et est amené à le remplacer.
- La **connectique** :
  - L'interface VGA standard : Les cartes graphiques sont la plupart du temps équipées d'un connecteur VGA 15 broches (Mini Sub-D, composé de 3 séries de 5 broches), généralement de couleur bleue, permettant notamment la connexion d'un écran CRT. Ce type d'interface permet d'envoyer à l'écran 3 signaux analogiques correspondant aux composantes rouges, bleues et vertes de l'image.



- L'interface DVI (*Digital Video Interface*), présente sur certaines cartes graphiques, permet d'envoyer, aux écrans le supportant, des données numériques. Ceci permet d'éviter des conversions numérique-analogique, puis analogique numérique, inutiles.



- L'interface S-Video : De plus en plus de cartes sont équipées d'une prise S-Video permettant d'afficher sur une télévision, c'est la raison pour laquelle elle est souvent appelée *prise télé* (notée « **TV-out** »).



## Fonctionnement d'une carte graphique

### Introduction.

La carte graphique (écran) est la carte d'extension d'un ordinateur qui transforme un flux de données numériques en un signal analogique compréhensible par un écran. Les cartes graphiques Hercules furent les premières cartes dédiées au PC, elles étaient en noir et blanc (ou plutôt selon le moniteur ambré). La couleur est réellement apparue avec les premiers 286. A cette époque, la norme était le CGA ou EGA, le moniteur devait être compatible avec la carte écran. Utilisant une résolution et un nombre de couleurs restreint, cette norme fut remplacée par la norme VGA. La norme VGA permettait d'afficher une résolution de 640 \* 480 points en 16 couleurs. Une carte graphique VGA inclut 256 k de mémoire.

Comme chaque constructeur souhaite toujours améliorer, chacun a sorti des cartes graphiques S-VGA (pour Super VGA). Ces cartes ne répondent pas à une norme, chaque constructeur développe un pilote (un programme) qui permet au système d'exploitation de gérer ses possibilités. Chaque pilote est spécifique au modèle de carte graphique mais aussi au système d'exploitation. Le mode VGA est reconnu par tous les systèmes d'exploitation actuels, c'est le mode par défaut. Pour passer en mode supérieur, un pilote-driver (le programme spécifique au matériel) doit être installé. Il est normalement spécifique au système d'exploitation (XP et 2000 acceptent généralement les mêmes pilotes). Vista en nécessite des nouveaux.

Une zone mémoire est réservée dans la mémoire haute depuis les premiers 8088. Elle est actuellement largement inférieure à celles utilisées par une carte graphique (jusque 700 MB) et peut être récupérée par les commandes DOS de mémoires hautes. Le processeur graphique garde les zones non modifiées dans sa propre mémoire. Ceci permet de ne pas trop utiliser le processeur.

### Caractéristiques d'une carte graphique (écran).

Une carte écran se caractérise par:

Bus utilisé (ISA, PCI, AGP, PCI-Express). Seul le bus AGP est spécifique aux cartes écrans

- le circuit intégré qui gère l'affichage, le GPU
- Mémoire
- Convertisseur digital – analogique (DAC)
- Taux de rafraîchissement maximum.
- Possibilités 3D

- BIOS Vidéo
- Connecteur de sortie

Quelques unes de ces caractéristiques sont identiques à celle d'un système à microprocesseur: processeur, mémoire et Bios. En effet, le sous système vidéo est un système à microprocesseur tout à fait conforme dont le seul but est de traiter des informations digitales provenant du PC pour les transférer vers le convertisseur.

### Le bus utilisé.

L'installation d'une carte graphique dépend forcément du bus utilisé (ISA, PCI ou pour les cartes récentes AGP). Les performances en dépendent. Rien ne sert de rouler en Ferrari sur des chemins de terre. Le bus ISA est limité à 16 MB/s, le bus PCI à 132 MB/s, l'AGP 2X à 528 MB/s et l'AGP 4X culmine à plus de 1,7GB/s (en théorie), PCI-Express, 4 GB/s en 1.1 et 8 GB/s en 2.0.

### Le circuit intégré.

Le chipset graphique transforme les signaux digitaux en signaux digitaux organisés et lisibles par le convertisseur digital - analogique. Généralement, le circuit est conçu pour utiliser une certaine capacité mémoire. Par exemple, une S3 TRIO 3D gère une mémoire de 4 MB, mais une extension de 4 MB supplémentaire est possible. Ceci n'est pas à conseiller. Même si le bus d'adresse est conçu pour gérer l'extension, l'intérieur du processeur est conçu pour la mémoire de base. Vous autorisez de meilleures résolutions, mais n'améliorez pas les performances globales de la carte graphique.

Parler des circuits pourrait prendre de nombreux chapitres, et vu la vitesse où les chipsets vidéo apparaissent et disparaissent, une mise à jour hebdomadaire serait nécessaire, surtout au niveau 3D. Les informations suivantes sont donc données à titre indicatif.

Les principaux constructeurs de cartes graphiques sont

- **S3.** Cette firme très active fabrique généralement en bas de gamme, mais peut parfois surprendre, comme la SAVAGE 4 (16MB de base, 3D). Racheté par VIA.
- **ATI:** d'excellentes cartes écrans de bas milieux de gamme. Racheté par AMD en 2006
- **INTEL:** s'est intéressé un peu aux cartes écrans avec le i740. Ce processeur est également intégré dans le chipset 1810, gérant en base 8 MB, il est 3D. Ceci explique que les cartes à base de I810 n'incluent pas de bus AGP. Actuellement, INTEL est le principal fabricant de cartes graphiques, intégrées dans le chipset.
- **Diamond.** Généralement haut de gamme, mais spécialisé plutôt vers les jeux. Racheté par S3 en 1999.
- **Nvidia:** premier fabricant mondial, toute la gamme avec la série Gforce
- Matrox, autrefois célèbre ne produit plus que des solutions pour des affichages multiples sur grands écrans réservés à l'industrie.

### La mémoire.

La mémoire de la carte graphique permet d'augmenter la résolution tout en gardant un nombre de couleurs respectables. Par exemple, un écran de 1024 \* 768 pixels en 2 couleurs (noir et blanc) nécessite 786.432 pixels \* 1 bit /8 (pour codage en octets)=98 K. Pour 4 couleurs (nb.

Octets =2), 196,6K et en 16 couleurs (4 bits):384 K. Ceci donnerait pour 1600 \* 1200 en 16 millions de couleurs (14 bits)=30.000 MB

Résolution	Nombre de couleurs	Mémoire
640 * 480	256	512K
	32 K	1 MB
	64 K	1 MB
	16 Millions ( 24 bits)	1 MB
800 * 600	16	512 K
	256	512K
	32 K	1 MB
	64	1 MB
	16 M	2 MB
1024 * 768	16	512K
	256	1 MB
	32 K	2 MB
	64 K	2 MB
	16 M	4 MB
1280 * 1024	16	1 MB
	256	2 MB
1600 * 1200	256	4 MB
	64 K	4 MB

Dans le cas des cartes 3D, une large partie de la mémoire est utilisée pour enregistrer des textures. Une carte 3D 32 MB ne permet donc pas d'utiliser des résolutions très élevées. De toute façon, aucun écran actuel n'est capable d'afficher les résolutions maximum des cartes écrans. Pour rappel, l'œil humain est limité à 16 Millions de couleurs.

Le processeur est relié à la mémoire à travers un large bus compris entre 64 et 128 bits.

Les types de mémoires utilisées ont également leur importance.

- les premières cartes écran utilisaient des mémoires **Ram** peu coûteuses mais lentes.
- La mémoire **VRAM** était-elle y a peu couramment utilisées. Le processeur et le RAMDAC (convertisseur digital – analogique) peuvent y accéder simultanément.
- La mémoire **WRAM** développée par SAMSUNG offre des débits 25% supérieurs à la VRAM et dispose de fonctions pré décodées tel que le dessin de texte ou le remplissage de blocs. Elle est apparue avec la Matrox Millennium.
- La **DRAM** a été développée par MoSYS et constituée d'une multitude de petits bancs mémoires (32K). Ceci permet d'adapter la mémoire à celle requise par blocs de 32K. Par exemple, une résolution de 1024\*768 en true color (24bits) utilise 2.5 MB et non 4MB. Le débit est ici aussi largement supérieur à celui des précédents.
- La **SGRAM** (Synchronous Graphic RAM) a la capacité de fonctionner à la même vitesse que le bus et peut atteindre des fréquences supérieures à 100 Mhz. Elle est jusqu'à 4 X plus rapide que les DRAM, mais est nettement plus chère.
- La **SDRam** permet des vitesses de 150 Mhz. **Déjà utilisée dans les PC**

- La **DDR SDRam**, sortie en 1999 permet de faire passer la mémoire de 150 à 300 Mhz, faisant progresser la vitesse des cartes entre 25 et 30 % par rapport à une SDRam. Elle utilise le flanc montant et descendant du signal d'horloge. Sa première utilisation date de mai 2000 avec le GeForce 256 de Nvidia 64 MB. Cette mémoire est également utilisée dans les PC.
- **DDR2** et **DDR3**: les cartes vidéo utilisent finalement maintenant les mêmes mémoires que les ordinateurs, juste avec un peu d'avance.

### Le convertisseur digital – analogique

Le convertisseur digital analogique permet la conversion. Il est souvent appelé **RAMDAC**. Plus sa vitesse est élevée, plus la fréquence de rafraîchissement verticale de l'écran est élevée (nous en reparlerons). Des cartes graphiques de haut de gamme actuels dépassent les 200 Mhz.

### Les possibilités 3D

Les cartes graphiques actuelles offrent des possibilités 3D: la carte graphique est capable de créer directement des polygones et remplissages (mapping), sans interventions directes du processeur du PC. Ces possibilités sont utilisées principalement pour les jeux. Sans cartes 3D (éventuellement Ad-on), peu de possibilités de s'exprimer sur les nouveaux jeux. C'est à ce niveau que se distinguent les nouvelles cartes graphiques par rapport à celles dites 2D.

Le fonctionnement de ces cartes nécessitent 2 circuits électroniques (en plus de la mémoire). Au tout début des cartes 3D, le CPU du PC créait les images et un circuit électronique (la carte écran) se contentait des les afficher. Avec l'arrivée de la première puce Voodoo de 3Dfx, la carte incluait un Triangle Setup Engine et un pixel Rendering engine directement sur la carte graphique. Ceci permettait de réduire l'utilisation du CPU du PC pour la création de polygones et des effets de pixellisation dans les jeux. Avec l'arrivée de la puce GeForce 256 en 1999, le circuit intégré intégrait en plus Transformation & Lighting (T&L). Le gros avantage est que l'on pouvait travailler sur des images constituées de plus de triangles. Pour rappel, une image 3D est entièrement constituée de triangles et la qualité de l'image est directement liée au nombre de triangles. En 2000, la GeForce2 peut élaborer jusque 25 millions de triangles par seconde, la GeForce 256, modèle précédant, n'en gérait que 15 millions.

Mise à part la technologie T&L, les circuits actuels gèrent l'éclairage par un ombrage très élaborés des pixels: chaque pixel est pourvu individuellement de données d'éclairage. La GeForce2, par exemple, peut gérer simultanément 4 pixels.

Les cartes graphiques futures devraient inclure un anti-alias qui permet de lisser les angles, un peu comme pour les imprimantes laser.

Implanté par NVidia depuis le processeur graphique GeForce 6600 et dans Direct X 9.0, le **shader** est un mécanisme utilisant des morceaux de programmes directement exécutés par le GPU et non plus par le microprocesseur de l'ordinateur.

### Le taux de rafraîchissement.

Cette caractéristique est liée au **RAMDAC**. Le signal n'est pas envoyé d'un bloc vers l'écran, mais points par points suivant une ligne, ensuite celle en dessous, ... Les anciens systèmes utilisaient le mode entrelacés pour camoufler leurs faibles performances (par opposition au mode non entrelacé NI). Le principe entrelacé, utilisé dans les TV, ne rafraîchit d'abord que les lignes

impaires, et ensuite reprend les lignes paires. Ceci provoque des scintillements de l'écran, tout d'abord désagréable, mais vite fatigant. Un taux de rafraîchissement correct débute à 70 hz. Plus il est important, meilleure la qualité de l'affichage est. Néanmoins, les anciens écrans n'acceptent pas de fréquences trop élevées, supérieures à celles minimum des cartes écrans actuelles. Ceci explique qu'il faut parfois changer d'écran lorsque l'on change de carte graphique. Nous en reparlerons.

### Le Bios Vidéo.

Au même titre qu'un PC, une carte graphique possède son propre BIOS. Celui-ci est spécifique au fabricant de la carte. Ceci explique des différences de performances entre 2 cartes de Chipset identiques. Sur certain modèles, le bios est flashable.

### Le connecteur de sortie.

Le connecteur de sortie est identique en mode S-VGA (toutes les cartes actuelles) à celui de la norme VGA. Certaines cartes acceptent le RGB. Dans ce cas, le signal est séparé en 3 couleurs (rouge - vert - bleu) avant d'être transmis à l'écran via 3 connecteurs BNC. Ceci nécessite un écran spécial. Ce type de connexion ne se trouve que sur les stations de travail graphiques de haute gamme, rarement pour les PC.

### Double affichage et sortie TV

L'affichage sur plusieurs écrans est utilisé dans les stations de travail dédiées au graphisme et dessin technique depuis plusieurs années. Depuis Win98, vous pouvez utiliser 2 cartes écran simultanément (avec une carte écran PCI et une carte écran AGP quand les pilotes le permettent) pour afficher soit l'écran complet sur 2 écrans, soit le même affichage simultanément mais la solution actuelle utilise une seule carte écran munie de deux sorties (généralement une VGA et une DVI). Pour le même affichage simultané, vous pouvez aussi utiliser des "splitter", boîtiers amplificateurs qui permettent de raccorder jusque 32 écrans simultanément avec la même image.

Certaines cartes graphiques sont munies d'une sortie TV qui permet d'afficher votre écran informatique sur une télévision (sur les deux en même temps ou l'un ou l'autre suivant les modèles). Ce modèle de carte graphique n'est pas réellement fait pour "travailler" sur un téléviseur, mais bien pour les DVD, DIVX, ... La résolution d'une télévision est trop faible et la qualité de l'affichage de la partie informatique s'en ressent. Par contre, l'affichage des vidéos est identique à celle de la TV.

Les signaux TV sont repris suivant 3 normes: PAL, SECAM et NTSC. Ces 3 normes de télévision ne sont pas compatibles entre elles. L'émission en PAL (Belgique, Allemagne, ...) vers une TV SECAM (France) et vis versa produit un signal en noir et blanc. Le système NTSC est utilisé aux Etats-Unis, Canada, ... Le problème avec les cartes graphiques avec sortie TV est identique. Les 3 normes sont déclinées suivant différentes versions qui dépendent du pays.

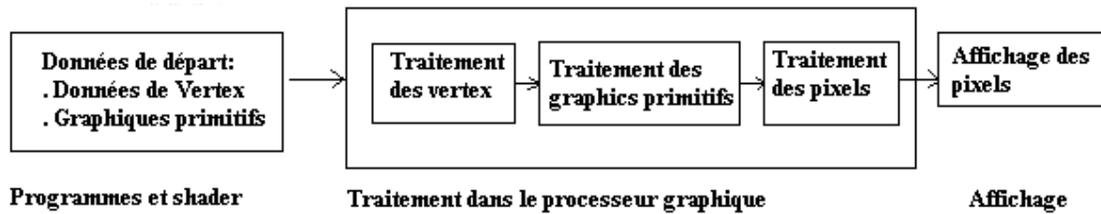
### Technologies diverses, SLI, Shader...

NVIDIA a sorti en 2006 la **technologie SLI**. Cette technologie permet de coupler deux cartes graphiques de même type et de même fabricant en PCI-Express via un connecteur spécial livré avec la carte même. Le chipset (obligatoirement un NForce de NVIDIA) de la carte mère doit également être compatible avec cette technologie. Un seul connecteur de sortie est utilisé,

mais cette fonction double la vitesse de traitement pour les jeux compatibles. Une version Quad utilise 4 cartes raccordées entre-elles, seule deux sont connectées dans un bus.

Les **shaders** sont des petits morceaux de programmes exécutés directement par le chipset graphique et non plus par le microprocesseur. Ils sont implantés dans DirectX 9.0 et depuis les séries GeForce 6600 par NViDIA. La version actuelle est la 3.0. Sans cette solution, certains jeux actuels ne fonctionnent pas sur les anciens PC.

Ces programmes se rangent dans deux groupes: les **Vertex Shader** et les **Pixel Shader**. Les parties constituées de suites de lignes et de triangles sont traités par les Vertex Shader comme des suites de graphiques primitifs. Une fois ces graphiques de base décomposés, le chipset graphique va les utiliser pour créer le rendu de couleurs des différents points au niveau des Pixel.



## S-Ata - Serial ATA

S-ATA est une norme de connexion de disque dur et dérive des connexions ATA (E-IDE). La technologie interne de ces disques durs est identique. Par contre, la connexion ne se fait plus par un câble en nappe (connexion parallèle de 40 ou 80 fils) mais bien par un câble de type série reprenant 7 fils. Le débit maximum



(théorique) est de 150 MB/s (contre 133 Mega-byte pour les ATA les plus rapides). Les premiers périphériques datent de début 2003.

Si les connexions séries avaient la réputation d'être lente par rapport aux liaisons parallèles, le débit actuel est limité par les interférences électromagnétiques sur le signal. Cette constatation est également mise en pratique dans l'hyper-transport (bus inter-processeurs AMD). Le connecteur reprend 7 fils, 3 fils de masse et 4 pour la communication (bi-directionnelle avec 2 fils différentiels par canal - signal effectif et signal inversé). Il est muni d'un détrompeur.

L'alimentation des périphériques se fait soit par un connecteur 4 broches standard, soit par un connecteur 15 broches reprenant les tensions 3,3V, 5V et 12 V. Un adaptateur permet de raccorder une alimentation ATX sur ces connecteurs. Les alimentations S-ATA reprennent directement 2 connecteurs de ce type.

Contrairement à l'E-IDE, Un seul disque SATA est connecté par contrôleur. La bande passante n'est donc pas partagée. Les cartes mères actuelles incluent généralement 2 connecteurs (4 sur les plus performantes) et permettent la configuration en RAID 0 (déconseillé) ou en RAID 1 (sécurité). Certaines cartes acceptent jusqu'au RAID 10 (3 ou 4 disques durs). Il n'y a pas de pontage à paramétrer comme sur les IDE (master - slave). Le Serial ATA peut-être vu comme un SCSI "low Cost". Il reprend d'ailleurs quelques caractéristiques comme le hot plug (à condition que le système d'exploitation accepte, soit 2000, XP et 2003) ou le contrôle d'erreur. Il utilise la technologie LVDs (Low Voltage Differential signaling), similaire au LVD.

La **connexion sur la carte mère** ne pose donc pas trop de problèmes. Reste la **détection du périphérique par Windows**. Les versions actuelles de Windows ne gèrent pas nativement le SATA. Ceci nécessite un pilote spécifique. Même si certains drivers acceptent Win98 (disque supplémentaire), la connexion ne peut se faire en pratique qu'à partir de Windows 2000 ou XP. Quelques cartes mères (généralement de bas de gamme) nécessitent une disquette spécifique à créer à l'aide du CD-Rom fourni sur le CD d'installation de la carte mère. Au démarrage de l'installation de Windows, lorsqu'il propose de charger des pilotes supplémentaires, appuyez sur la touche F6. Après le chargement des pilotes standards, Windows vous demandera la disquette préalablement créée. Les cartes mères actuelles permettent généralement la reconnaissance directe

des périphériques en émulant un port ATA standard. Des limitations de capacités peuvent également se poser.

Même si la technologie semble intéressante, la vitesse de transfert est limitée par la construction interne des disques durs. Le temps d'accès restant limité au niveau des plateaux, la vitesse de transfert interne plafonne pour les disques les plus courants vers les 80 GB/s. De plus, les premiers contrôleurs utilisaient le bus PCI (limité à 133 MB/s pour l'ensemble des périphériques). Actuellement, le contrôleur SATA est intégré dans le Southbridge du chipset. La norme suivante (S-ATA2) sorti en 2005 double le taux de transfert théorique.

Le disque dur (Hard Disk ou HD en anglais) est le composant principal de sauvegarde de masse. Au cours hardware de première année, nous ne voyons que les disques durs de type IDE et SATA. Les disques durs et contrôleurs SCSI seront vus en deuxième année. Les disques durs MFM et RLL utilisables uniquement avec les 8088 sont eux-aussi passés, obsolètes. Les disques durs IDE sont utilisés depuis les ordinateurs à base de 80286, même si la norme a déjà été améliorée plusieurs fois depuis.

Le disque dur est géré par un contrôleur, les deux doivent être de même type. Par **contrôleur IDE**, on peut installer 2 disques, un maître (master) et un esclave (slave) via un câble en nappe, à condition de configurer des petits pontages sur le disque dur. Pour les **Sérial ATA**, un seul disque dur par contrôleur et pas de pontage. Les disques durs se dénomment C:, D:, E: , ... Le(s) lecteur(s) CD-ROM prend (prennent) les dernières lettres. Ceci oblige souvent à reconfigurer les programmes qui utilisent les CD-ROM en permanence lorsque l'on installe un deuxième disque dur. Ceux-ci utilisant lors de l'installation la lettre désignant le CD par défaut. Windows 2000, Windows XP ou Vista permettent de modifier les lettres de chaque disque - partition..

### Caractéristiques d'un disque dur.

Un disque dur est constitué de:



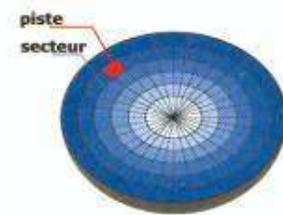
1. **plateaux** en aluminium superposés sur les quels une couche d'élément magnétique est déposée (généralement sous forme d'oxyde de fer). Quelques modèles actuels les plus performants utilisent des plateaux en verre, moins sensibles aux vibrations acoustiques et aux déformations dues à la chaleur.
2. le **moteur de rotation** unique pour l'ensemble des plateaux, il tourne en permanence.
3. Les têtes de lecture / écriture (une par plateau) lisent et écrivent les informations, elle survole les plateaux à une distance infime. Si la tête touche le plateau, celui-ci est rayé et donc inutilisable.
4. Le moteur pour les **déplacements** des têtes (généralement un seul)
5. **Composants électroniques** de contrôleur pour le transfert des données.

Les disques durs sont caractérisés par le nombre de plateau (et donc de têtes), la vitesse de rotation, le temps d'accès, ...

Le **Temps d'accès moyen** fait référence au temps moyen pour que la tête se déplace d'un point à l'autre sur un même plateau. Si les disques durs de 1996 tournaient aux alentours de 13 ms, les standards actuels arrivent à moins de 8 ms.

La **Vitesse de rotation** est la vitesse du moteur d'entraînement des plateaux. Elle est exprimée en tour par minutes. La vitesse actuelle sont de 5400 (généralement pour PC portables) et 7200 tours selon les modèles. Certains atteignent même 10.000 tours par minutes. Quelques modèles SCSI Ultra 160 atteignent 15.000 tours par minutes, mais avec une ventilation du disque.

Chaque plateau est divisé en **pistes**, zones circulaires. Plus la piste se rapproche du centre du plateau, plus elle se réduit. Chaque piste est divisée en secteurs par le formatage de bas niveau, découpages de piste. La taille est fixe et généralement de 512 Byte (octet). Les clusters sont créés au formatage de haut niveau (la commande Format), la taille varie suivant le type et la taille de la partition. Le cluster est la plus petite capacité minimum utilisée par un fichier, il est indivisible. Le Boot secteur est le secteur de démarrage unique, même en multi-boot.



La capacité d'un disque dur se déduit donc de la formule ci-dessous:

**Capacité nette** = secteurs \* capacité de secteurs \* cylindre \* nb. Têtes.

Par exemple, un Seagate 10232 possède 1245 cylindres, 255 têtes (plateaux) et 63 secteurs. La formule devient capacité nette =  $63 \times 512 \times 1245 \times 255 = 10322727645$ . Divisons par  $1000 \times 1000$

\* 1000 (le Kilo hardware) pour obtenir les GB et nous obtenons: 10,322 GB, soit un peu plus que la taille détectée de 10,241 GB par la carte mère. Pour rappel, l'hardware utilise des kilo de 1000 octets (ce qui est une erreur). Par contre, le software (par exemple Windows) utilise des kilos de 1024 ( $2^{10}$ ).

### Types de partitions Windows.

Pour être utilisé, un disque dur doit au préalable être préparé à recevoir des informations. La première commande DOS FDISK permet de partitionner le disque dur. Suit ensuite le **formatage** des différents disques durs. Attention, sur les anciens PC, une commande du BIOS permettait le **formatage de bas niveau**. Avec les disques actuels (à partir du 486), cette commande n'est plus nécessaire. Son utilisation fait même perdre une partie de la capacité dans certains cas. Elle est encore utilisée pour réparer un disque avec des clusters défectueux via un programme spécifique à la marque de disque dur. Renseignez-vous sur le site du constructeur.

La **Fat** est la manière de gérer les informations sur les disques durs. Elle représente la table de matière qui inclut les informations de saut. DOS 3.3 et inférieur gérait cette FAT de manière personnelle.

La **FAT 16** permet des partitions de 2 GB maximum (à partir de DOS 4.0 début 1990) comme disque principal ou logique (nous en parlerons en atelier). La taille du cluster (plus petite information possible) varie suivant la taille de la partition, mais également suivant le type de FAT. Elle a été modifiée en VFAT (Virtual FAT) pour accepter les noms longs et accepter une compatibilité avec les partitions DOS de Win95

La **FAT 32** est apparue avec Win95 OSR2 (Win95B) n'est pas compatible avec la FAT 16. Si Win95B peuvent lire les FAT16, DOS ne peut lire les informations sur les disques partitionnés en FAT32. La partition maximum théorique en FAT 32 est de 2 Téra Bytes (2000 GB)

Une **partition NTFS** n'est plus basée sur les clusters mais sur une table de fichier maître qui se base sur un descriptif de fichiers (emplacement, attributs, droits d'accès utilisateurs, quota d'espace pour les utilisateurs, ...) et permet de meilleures protections des fichiers au niveau accès utilisateurs. La taille des clusters est également fixe: 4 KB pour les partitions 2000 et XP (limitée à 2 TeraByte en théorie) et 256 Tera-byte pour 2003.

D'autres types de partitions existent suivant le système d'exploitation. Elles ne sont pas toujours lisibles par tous les systèmes d'exploitation. OS2 utilisait une table d'allocation en 32 uniquement lisible par des machines OS2, mais lit les partitions en FAT 16. Windows NT 4.0 lit les partitions en FAT 16, mais pas les FAT32, et inclut des partitions en NTFS, pas lisibles par Win95-98 et limitée à 2,1 GB. Windows 2000 et XP accepte FAT 16, FAT 32 et NTFS sans limitations. Pour complément sur les types de partitions, référez-vous au cours sur les systèmes d'exploitation.

### La norme IDE, E-IDE ou Ultra-IDE.

Avec l'arrivée des 80286, le standard des disques durs est l'IDE. Installé sur un contrôleur indépendant (une carte dans un connecteur ISA), le taux de transfert maximum ne dépasse pas les 4,7 MB/s.

La norme IDE permet de connecter 2 disques durs (un maître - master et un esclave – slave) sur le même contrôleur via un câble adapté. Ce câble est toujours le même. La taille maximum

d'un disque dur est limitée à 540 MB (donc pas de lecteur CD-ROM). Des programmes spécifiques permettent des capacités supérieures.

La norme E-IDE est sortie avec les 486DX-4 et les premiers Pentium. Elle dépasse la limite des 540 MB de la norme IDE. La taille maximum des disques durs supportés est de 8.4 GB. Des disques durs supérieurs sont supportés et même parfois détectés par le BIOS (suivant la carte mère), mais le formatage (ou FDISK) n'autorise jamais plus que 8,4 GB (même avec Fdisk). Les évolutions suivantes vont supprimer cette limitation.

La vitesse de transfert sur ces disques est limitée à 10 MB/s. Pour que les disques de capacités supérieures à 540 MB soient pris en compte, il faut que le disque soit détecté dans le bios comme LBA (logical Block Addressing).

Depuis l'E-IDE, les lecteurs CD-ROM sont reconnus comme périphériques IDE standards.

Avec la technologie des disques durs E-IDE (et suivante), vous pouvez installer 4 périphériques. En effet, les contrôleurs E-IDE incluent 2 ports (primaire - primary et secondaire - secondary) sur les quels vous pouvez connecter chaque fois un master (maître) et un slave (esclave). Le paramétrage des masters - slaves se fait par des pontages à l'arrière du périphérique. Attention tout de même, les contrôleurs UDMA 100-133 sont uniquement sur le premier contrôleur. Vous ne pouvez donc généralement pas connecter ces disques sur le port secondaire.

### Modes de transfert.

Apparu au niveau des Pentium, le **mode PIO** désigne la vitesse de l'interface. Elle est gérée par le CPU et correspond à la manière dont sont traitées les instructions chargées de transférer les données au disque dur. Si le procédé offre des débits intéressants, il présente l'inconvénient de mobiliser les ressources du processeur. Chaque disque supporte un mode PIO de par sa conception. Le plus simple est de laisser le mode de détection automatique dans le BIOS, ce qui garantit le mode le plus élevé possible.

Version ATA	Mode	Vitesse maximum (MB/s)
<b>ATA-0</b>	PIO Mode 0	3,3
<b>ATA-1</b>	PIO Mode 1	5,2
<b>ATA-1</b>	PIO Mode 2	8,3
<b>ATA-2</b>	PIO Mode 3	11,1
<b>ATA-3</b>	PIO Mode 4	16,7
<b>ATA-4, UDMA-33</b>	Ultra DMA	33,6

### Les disques UDMA-33 ou ATA-33 ou ATA 4

Si le mode PIO procure de bonnes performances (en mono-tâche), son principal défaut est l'utilisation abusive du processeur. Le mode DMA (Direct Memory Access) permet de transférer des données du disque dur vers la mémoire et vis versa sans passer par le processeur.

Apparue avec les Pentium II, la norme UDMA-33 peut charger les informations à la vitesse de 33 MB/s en mode rafale (tout d'un coup). Ce transfert sur 16 bits se fait sur les flancs montants et descendants de l'horloge à une fréquence de 8 Mhz.

### L'UDMA-66 ou Ultra ATA 66

Sortis en 1999, l'UDMA-66 permet un transfert en mode rafale de 66 MB/s. La fréquence a été doublée par rapport à l'UDMA-33 (16 Mhz) sur toujours une largeur de 16 bits. D'un point de vue technique, l'Ultra DMA-66 renforce l'intégrité des données en améliorant la forme du signal. Pour cela, la nappe (le câble) utilise 80 fils au lieu de 40 pour les anciens IDE. Le nombre de fils augmente, mais les connecteurs restent à 40 fils pour compatibilité.

Les contrôleurs UDMA-33 acceptent ces disques, mais forcément en mode UDMA-33. Les contrôleurs UDMA-66 gèrent également l'UDMA-33.

Seuls certains systèmes d'exploitation sont capables de le faire tourner correctement (Win95 OSR2, Win98, Win NT).

### ATA / 100 et ATA 133- ATA6

L'ATA 100 utilise les flancs montants et descendants du signal tout comme l'UDMA66, mais l'horloge est réduite à 40 ns. La norme 133 (sorti en septembre 2001) augmente encore la fréquence de l'horloge.

Une remarque, seuls les disques durs de marque MAXTOR utilisent le mode ATA-133. De plus, les chipsets INTEL n'acceptent que le mode ATA-100.

### Résumé des modes.

Pour rappel, le bus IDE travaille sur 16 bits, d'où 2 bytes sont transférés chaque fois.

MODE	Période d'horloge	Conteur. horloge	Temps d'horloge (ns)	Taux de transfert
PIO Mode 0	30 ns	20	600	(1/600 ns) X 2 byte = 3,3 MB /s
PIO Mode 1	30 ns	13	383	(1/383 ns) X 2 byte = 5,2 MB /s
PIO Mode 2	30 ns	8	240	(1/240 ns) X 2 byte = 8,3 MB /s
PIO Mode 3	30 ns	6	180	(1/180 ns) X 2 byte = 11,1 MB /s
PIO Mode 4	30 ns	4	120	(1/120 ns) X 2 byte = 16,6 MB /s
DMA Mode 0	30 ns	16	480	(1 / 480 ns) X 2 byte = 4,16 MB /s
DMA Mode 1	30 ns	5	150	(1/150 ns) X 2 byte = 13,3 MB /s
DMA Mode 2	30 ns	4	120	(1/120 ns) X 2 byte = 16,6 MB /s
UDMA 33	30 ns	4	120	(1 / 120 ns) X 2 byte X 2 = 33 MB /s
UDMA 66	30 ns	2	60	(1 / 60 ns) X 2 byte X 2 = 66 MB /s
UDMA 100	20 ns	2	40	(1 / 40 ns) X 2 byte X 2 = 100 MB /s
UDMA 133	20 ns	2	30	(1 / 30 ns) X 2 byte X 2 = 133 MB /s

## Limitations de capacité.

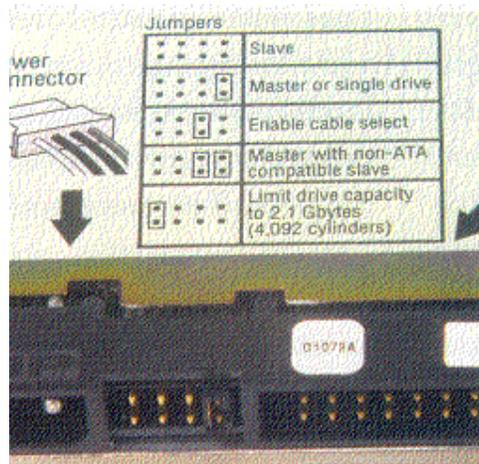
Peut-on mettre n'importe quel disque dur IDE dans un PC? Pas tout à fait. Le bios gère les contrôleurs et selon la date de ce BIOS, des limitations de capacité peuvent intervenir. Si vous placez un disque dur de 20 GB sur un 386, vous avez toutes les chances que le disque soit reconnu comme un ... 528 MB. En règle générale, les limites de capacité suivantes apparaissent suivant la date du Bios, système d'exploitation, ...

BIOS antérieur à	Limitation de capacité	
Août 1994	528 MB	IDE
Février 1996	2,1 GB	BIOS, partitions Win NT, DOS et FAT 16 (y compris Win 95 première édition), chipset 430FX
	3,27 GB	BIOS
Janvier 1998	8,4 GB	BIOS, FAT 16
Juin 1999	32 GB	UDMA
	64 GB	FDISK de Win 98 (pas le formatage si la partition est créée avec Millenium)
Fin 2001	120 GB (137GB)	Limite de Bios
	160 GB	Cette limitation peut être liée au système d'exploitation (Windows 2000 inférieur à SP4 ou XP SP2) ou au BIOS

En fait, il faut essayer. C'est la carte mère qui limite souvent la capacité de disque dur. Une limitation de Win98 aux disques supérieurs à 32 GB n'est pas une limitation de Windows par exemple. Par contre, la commande FDISK de Win98 n'accepte pas les disques supérieurs à 64 GB et Millenium s'il les accepte n'autorise pas de créer plusieurs partitions sur le disque. Pas vraiment d'explications pour la limitation à 120 GB mais certains flashages de BIOS pour des PC de 2002 résolvent le problème. Il semble que les fabricants gèrent les disques supérieurs chacun à leur manière: certaines marques de disques durs supérieurs à 120 GB soient reconnus par une carte mère, d'autres non. Une nouvelle norme adoptée en juin 2001 (ATA / ATAPI-6 ou Big Drive) supporte les disques durs jusqu'à 144.000.000 GB. Une limitation à 160 GB vient de Windows 2000 (version inférieure à SP4) et XP (version inférieure à SP2) liée à la carte mère (toutes n'ont pas ce problème). Un patch de correction est téléchargeable chez les fabricants de disques durs à partir de ces versions.

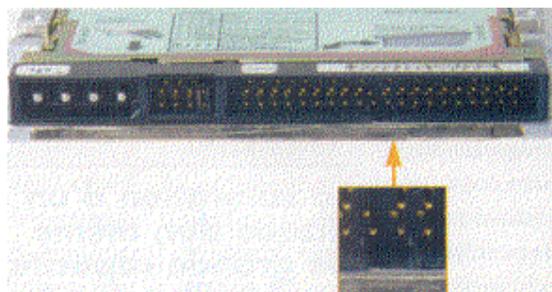
## Installer un disque dur IDE (ou CD-ROM)

Paramétrez les pontages (jumpers) du nouveau disque dur (et de l'autre si vous utilisez 2 périphériques sur le même contrôleur) en maître ou esclave ou single (seul) selon l'exemple ci-dessous (attention, le système est différent pour les disques durs de marque Western digital). Généralement, on utilise le nouveau comme disque primaire (premier) puisqu'il est plus rapide, mais vous devez réinstaller Windows ou à utiliser un programme tel que Ghost de Symantec.



Commencez par le fixer dans le boîtier. Repérez ensuite le connecteur IDE primaire ou secondaire sur la carte mère (contrôleur sur carte séparée pour les 486 et inférieurs). Chaque contrôleur peut accepter 2 disques durs (ou CD-ROM, graveur, Zip, ...). Sur chaque contrôleur, un disque peut être seul (master ou single), en maître (avec un deuxième disque en esclave) ou en esclave. Le disque dur principal (ou se trouve le système d'exploitation) doit se trouver en Maître sur le contrôleur primaire.

Connectez le câble IDE entre le contrôleur et le disque dur en prenant soin de repérer la ligne de couleur sur un côté du câble. En effet, les connecteurs sont munis parfois d'un détrompeur, mais c'est rarement le cas. Il faut que pour le contrôleur et les disques, la broche 1 soit chaque fois raccordée ensemble. Par principe, les techniciens relient le côté rouge du câble sur la borne 1 de chaque connecteur (vers le connecteur d'alimentation).



Vous pouvez maintenant démarrer le PC. Pour les anciens BIOS, il fallait faire une auto-détection (ou même plus anciens rentrer les paramètres manuellement). Actuellement, les paramètres des disques durs sont auto-détectés directement au démarrage.

En cas de problèmes, vérifiez la connexion du câble IDE (enfoncement, sens), notamment si le PC ne démarre pas avec la led rouge du disque dur qui reste allumée en permanence. Vérifiez également les pontages master – slaves.

### Disque dur Serial ATA (S-ATA)

Les disques durs de type Serial ATA ont été annoncés en mai 2001. Ils ne sortent véritablement qu'au deuxième trimestre 2003. Evolution du SATA, le **S-ATA 2** sorti début 2005 double en théorie le taux de transfert maximum: de 150 à 300 MB/s (pour 133 en ATA parallèle).

Les 2 versions sont compatibles entre-elles, un contrôleur S-ATA accepte les disques durs SATA II et vis versa.

Le serial ATA est un nouveau gestionnaire (contrôleur) de disques durs de type IDE. La technologie interne est identique à celle des disques durs PATA, seule la méthode de transfert des données est modifiée, passant en mode série. Le transfert parallèle commence nettement à poser des problèmes de synchronisation des signaux lorsque l'on augmente la vitesse de transmission. Quelques lecteurs CD - DVD et graveurs utilisent ce type de connexion, même s'ils n'ont pas besoin de performances puisque le taux de transfert est d'abord limité par la vitesse de lecture / écriture sur le CD.

La grosse différence vient donc de l'interface. Dans le cas des disques IDE, le contrôleur permet la connexion de 2 périphériques. La bande passante est donc partagée entre les 2. Ceci réduit les performances. Par contre, dans le serial ATA, chaque disque dur est relié à son propre contrôleur via son propre câble. Ceci implique qu'il n'y a plus de pontages à configurer pour les modes maîtres / esclaves. A la différence des précédentes normes, les disques durs S-ATA utilisent une connexion série sur 7 broches. Quatre fils sont utilisés pour le transfert des données (1 signal d'envoi, 1 signal de réception et 2 masses), d'où une réduction de la dimension du câble de raccordement. La longueur du câble est limitée à 1 mètre (contre 45 cm pour un ATA-133) Troisième différence, les disques serial ATA sont **hot plug**, ils peuvent être connectés (ou déconnectés) avec le PC allumé mais tous les systèmes d'exploitation n'autorisent pas cette fonction.

On trouve dans le commerce des adaptateurs permettant de passer d'un contrôleur ATA à un disque dur S-ATA. Cette solution limite néanmoins la vitesse à 133 MB/s (ATA-133) ou même 100 pour les chipsets INTEL d'où un intérêt limité. Inversement, des adaptateurs permettent de connecter des disques durs parallèles sur des contrôleur Sérial.

Certaines fonctions sont similaires aux disques durs SCSI comme le contrôle des erreurs lors des transferts. Le Serial-ATA se positionne comme un SCSI Low Cost (la norme SCSI la plus rapide permet des taux de transfert jusqu'à 320 MB/s).

Deux ou quatre contrôleurs S-ATA sont implantés sur les cartes mères, en plus d'un contrôleur ATA (deux pour les anciennes cartes mères). Généralement, le chipset accepte les modes RAID 0 et 1.

Le SATA n'est pas nativement reconnu par l'installation Windows dans la majorité des cas. Une disquette doit être créée au préalable. Au début de l'installation, vous devez appuyer sur la touche F6 pour installer le pilote supplémentaire.

### La commande DOS FDISK

Cette commande est reprise sous DOS pour les systèmes d'exploitation Win95/98 et Millenium. Dans le cas de XP, 2000 et NT, elle est accessible uniquement durant l'installation avec une interface différente ou dans les outils d'administration pour des disques supplémentaires. En pratique, vous pouvez créer des partitions avec Win98 et les réutiliser pour d'autres systèmes d'exploitation. **Elle est préalable à la commande FORMAT et efface complètement les données!**

Les partitions sont des découpages logiciels des disques durs. La première lettre accessible pour un disque dur est le C:, ensuite D: et ainsi de suite. Les lecteurs CD viennent ensuite (sauf Win2000, XP et Vista où vous pouvez modifier les lettres ensuite).

### Le block mode.

Ce procédé a pour but d'augmenter les performances générales du transfert. Dans la norme IDE, les données sont transférées 1 secteur à la fois (512 bytes). Chaque transfert nécessite une interruption. En block Mode, les données sont transférées par clusters. En Fat32, un clusters peut aller jusque 32 KB, une seule interruption est utilisée contre 64 dans le cas d'un IDE standard. Par contre, ce système peut-être à la source de quelques problèmes de stabilité du système d'exploitation. **A utiliser avec prudence.**

### Qu'entend-on par formatage de disque ?

Étant donné que le plus petit disque dur peut stocker des millions de bits, il doit y avoir un moyen de l'organiser pour faciliter la recherche d'une séquence particulière de bits. Cette forme d'organisation, qui est appelée formatage, prépare le disque dur de manière à ce que des fichiers puissent être écrits sur les plateaux et puissent au besoin être récupérés rapidement. Un disque dur doit donc être formaté physiquement et logiquement.

#### ❑ Formatage physique

Un disque dur doit être formaté physiquement avant de l'être logiquement. Le formatage d'un disque dur (que l'on appelle aussi formatage par niveaux) est normalement effectué par le fabricant.

Le formatage physique divise un disque dur en éléments physiques de base, c'est-à-dire en pistes, secteurs et cylindres. Ces éléments définissent la manière dont les données sont physiquement enregistrées et lues à partir du disque.

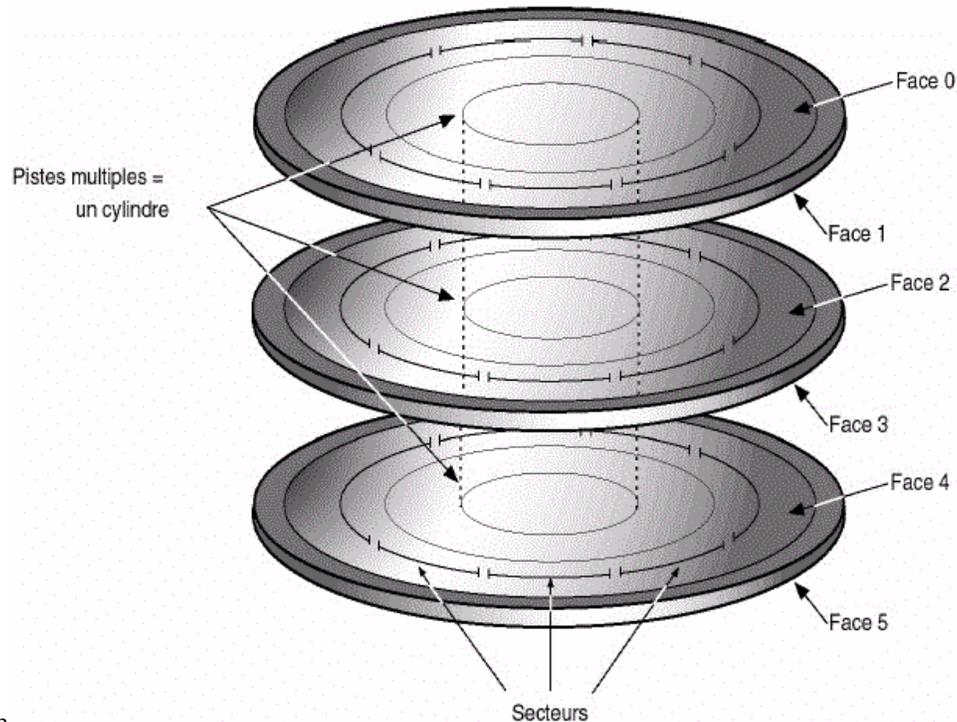
Les pistes sont les sillons circulaires concentriques gravés sur chacun des côtés du disque, comme ceux d'un microsillon ou d'un disque compact. Les pistes sont identifiées par un numéro, le bord extérieur étant la piste zéro.

L'ensemble des pistes qui se situe à la même distance du centre sur tous les côtés des disques s'appelle " cylindre ". Les éléments matériel et logiciel d'un ordinateur fonctionnent fréquemment au moyen de cylindres.

Les pistes sont divisées en zones ou "secteurs " utilisés pour stocker une quantité fixe de données. Les secteurs sont normalement formatés de manière à contenir 512 octets de données (chaque octet est composé de 8 bits).

Une fois qu'un disque dur a été physiquement formaté, les propriétés magnétiques du revêtement d'oxyde de fer sur certaines parties d'un disque risquent de se détériorer progressivement. De ce fait, il devient difficile pour les têtes de lecture/écriture de tracer sur le disque un schéma de bits pouvant être lu ultérieurement. Lorsque cela se produit, les secteurs dans lesquels les données ne tiennent pas bien sont appelés " secteurs défectueux ", mais ceci devient de plus en plus rare grâce aux disques actuels qui sont maintenant de meilleure qualité. De plus, les ordinateurs de fabrication récente peuvent généralement identifier un secteur défectueux, le marquer (de manière à ne pas l'utiliser) et en choisir un autre.

La figure ci-dessous indique le format physique



d'un  
disque dur standard.

#### □ **Formatage logique**

Une fois qu'un disque dur a été physiquement formaté, il faut effectuer un formatage logique. Le formatage logique place un système de fichiers sur le disque. Un système de fichiers permet à un système d'exploitation, tel que DOS, OS/2, Windows 95 ou Windows NT d'utiliser l'espace disponible pour stocker et récupérer des fichiers. Le formatage logique peut être fait au moyen des utilitaires de formatage qui sont fournis avec les systèmes d'exploitation.

Avant qu'un disque soit logiquement formaté, il peut être divisé en partitions. Un système de fichiers (format logique) différent peut être appliqué à chaque partition.

Les partitions feront l'objet d'une description plus détaillée dans la rubrique " Compréhension des partitions " et les systèmes de fichiers sont décrits en détail dans la section suivante.

Lorsqu'une partition de disque a été logiquement formatée, on s'y réfère en tant que volume. Au cours de l'opération de formatage, l'utilitaire vous demandera de nommer la partition, c'est-à-dire de choisir un " nom de volume " qui vous permettra d'identifier par la suite le volume (partition).

#### **Systèmes de fichiers**

Tous les systèmes de fichiers consistent en une structure nécessaire au stockage et à la gestion de données. Ces structures comportent généralement un enregistrement d'amorçage du système d'exploitation, des fichiers et des répertoires. Un système de fichiers effectue trois fonctions principales : 1) le suivi de l'espace alloué et libre, 2) la gestion des répertoires et noms de fichiers

et 3) le suivi de l'emplacement dans lequel les différentes parties de chaque fichier sont physiquement stockées sur le disque.

Il existe actuellement un grand nombre de systèmes de fichiers. Différents systèmes de fichiers peuvent être utilisés (reconnus) par des systèmes d'exploitation distincts. Certains systèmes d'exploitation ne reconnaissent qu'un seul système de fichiers, alors que d'autres peuvent en reconnaître plusieurs. Les systèmes de fichiers les plus courants sont les suivants :

File Allocation Table (FAT)

File Allocation Table 32 (FAT32)

New Technology File System (NTFS)

High Performance File System (HPFS)

#### □ **FAT**

Le système de fichiers FAT est celui qui est utilisé par DOS, Windows 3.x et normalement par Windows 95. Il peut également être utilisé par Windows NT et OS/2.

Le système de fichiers FAT se caractérise par l'utilisation d'une table d'allocation de fichiers (File Allocation Table, FAT) et de clusters qui représentent la plus petite unité de stockage de données. Ils consistent en plusieurs secteurs de disque. Le système FAT est utilisé pour enregistrer les clusters qui sont utilisés, ceux qui ne le sont pas et à quel emplacement se trouvent les fichiers. La table d'allocation de fichiers est au cœur de ce système de fichiers et elle est dupliquée pour protéger ses données.

Le système de fichiers FAT utilise aussi un répertoire racine qui peut contenir un nombre maximal d'entrées de répertoire autorisées et qui doit se trouver en un endroit spécifique sur le volume. Avec les systèmes d'exploitation qui utilisent le système de fichiers FAT, le répertoire racine est représenté par une barre oblique inverse (\) et c'est le répertoire qui est affiché en premier lieu lors de l'amorçage du système d'exploitation.

Lorsque vous créez un fichier ou un sous-répertoire, les informations relatives à ce fichier ou sous-répertoire sont stockées dans le répertoire racine sous forme d'entrée de répertoire.

Ainsi, une entrée de répertoire FAT contient des informations telles que le nom et la taille du fichier, la date et l'heure à laquelle les dernières modifications ont été apportées au fichier, le numéro du cluster de départ (c'est-à-dire le cluster qui contient la première portion du fichier) et les attributs du fichier (caché, système, etc.).

Le système de fichiers FAT peut gérer un maximum de 65 525 clusters. Par conséquent, la taille du cluster utilisé dépend de la quantité de l'espace volume disponible, la taille maximale d'un volume FAT étant de 2 Gigaoctets. Quelle que soit la taille du volume, la taille du cluster doit être suffisamment grande pour que tout l'espace disponible puisse être inclus dans les 65 525 clusters. Plus l'espace disponible est grand, plus la taille des clusters doit être grande.

#### □ **FAT32**

FAT32 est le système de fichiers utilisé par les versions mises à jour de Windows 95 (version 4.00.950B ou plus récente). Au moment de la parution de ce manuel, FAT 32 n'était disponible que pré-installé sur les ordinateurs de certains fabricants. (Microsoft a l'intention de lancer une mise à jour de Windows 95 avec laquelle FAT32 sera disponible sur une base générale). DOS, Windows 3.1, Windows NT et la version d'origine de Windows 95 ne reconnaissent pas les volumes FAT32 et ne peuvent donc pas être amorcés à partir des fichiers d'un volume FAT32. Il se peut que la situation évolue avec Windows NT.

FAT32 est une amélioration du système de fichiers FAT et il est basé sur des entrées de tables d'allocation de fichiers de 32 bits plutôt que des 16 bits actuellement utilisés par le système de fichiers FAT. Le système FAT32 est de ce fait capable de gérer des volumes de beaucoup plus grande taille (jusqu'à 2 téraoctets).

Le système de fichiers FAT32 utilise des clusters plus petits que le système FAT (par exemple des clusters de 4Ko pour des volumes ayant jusqu'à 8Go), contient des doubles des enregistrements d'amorçage et se caractérise par un répertoire racine qui peut être de n'importe quelle taille et situé n'importe où dans le volume.

#### ❑ **NTFS**

Le système de fichiers nouvelle technologie NTFS, (New Technology File System) n'est accessible qu'à partir du système d'exploitation Windows NT. NTFS n'est pas recommandé avec des disques d'une capacité inférieure à 400 Mo car il utilise une grande quantité d'espace pour les structures du système.

La table de fichiers permanente, MFT (master file table) constitue la structure centrale du système de fichiers NTFS, qui conserve des copies multiples de la partie critique de la table de fichiers de manière à vous protéger d'une perte éventuelle des données..Informations de base

Le système NTFS utilise des clusters pour stocker des fichiers de données, mais il n'y a pas de dépendance entre la taille de ces clusters et celle du volume. Il est possible de spécifier un cluster d'une taille inférieure à 512 octets, quelle que soit la taille du volume. L'utilisation de clusters de plus petite taille réduit la quantité d'espace disque perdue ainsi que la quantité de fragmentation de fichiers, c'est-à-dire lorsque les fichiers sont répartis sur un grand nombre de clusters non contigus, ralentissant ainsi l'accès aux fichiers. Le système de fichiers NTFS offre donc une bonne performance pour les grandes unités. Le système de fichiers NTFS gère la réparation automatique de secteurs défectueux qui sont détectés et signalés de manière à ce qu'ils ne soient pas utilisés.

#### ❑ **HPFS**

Le système de fichiers haute performance HPFS (High Performance File System) est le système de choix pour OS/2 et il est également géré par des versions plus anciennes de Windows NT. Contrairement au système de fichiers FAT, HPFS trie les répertoires sur la base des noms de fichiers et utilise une structure plus efficace pour organiser le répertoire. Ce qui se traduit le plus souvent par un accès aux fichiers plus rapide qu'avec les volumes FAT. Le système de fichiers HPFS rend, par ailleurs, l'utilisation de l'espace disque plus efficace que les systèmes de fichiers FAT.

HPFS attribue les données à des secteurs plutôt qu'à des clusters. Pour conserver la trace des secteurs utilisés et de ceux qui ne le sont pas, HPFS organise un volume en bande de 8 Mo, avec une allocation de bitmaps de 2 Ko entre les bandes. Les bandes améliorent la performance car les têtes de lecture/écriture n'ont pas à revenir à la piste zéro chaque fois que le système d'exploitation a besoin d'accéder à des informations relatives à l'espace volume.

### **Compréhension des partitions**

Une partition est une division physique de votre disque dur. Une fois qu'un disque a été physiquement formaté, et lorsque le formatage logique a été effectué, il peut être divisé en partitions distinctes.

#### ❑ **Pourquoi utiliser des partitions multiples ?**

De nombreux disques sont formatés de manière à ne contenir qu'une seule grande partition, ce qui ne fournit pas la sécurité optimale des données, et ne vous permet pas d'organiser vos fichiers de manière à pouvoir les trouver facilement, ni d'utiliser votre espace disque le plus efficacement possible.

Si vous souhaitez installer plusieurs systèmes d'exploitation sur un disque ou utiliser l'espace disque le plus efficacement possible, ou encore sécuriser au maximum vos fichiers ou séparer physiquement des données afin de faciliter la recherche de fichiers et la sauvegarde de données, il faut que vous compreniez comment utiliser des partitions multiples de différents types.

#### ❑ **Types de partition**

Il existe deux types de partitions principaux : les partitions principales et étendues. Les partitions étendues peuvent être elles-mêmes divisées en partitions logiques.

Il peut y avoir jusqu'à quatre partitions principales sur un disque dur, l'une de celles-ci pouvant être une partition étendue. Vous pouvez donc ainsi avoir quatre partitions principales ou trois partitions principales et une partition étendue.

#### ❑ **Partitions principales**

Une partition principale peut contenir n'importe quel système d'exploitation ainsi que des fichiers de données, tels que des applications et des fichiers utilisateur. Une partition principale est formatée logiquement pour utiliser un système de fichiers compatible avec le système d'exploitation sur lequel il est installé.

Si vous créez des partitions principales multiples, seule une partition peut être active à la fois. Lorsqu'une partition principale est active, les données des autres partitions principales ne sont pas accessibles, c'est-à-dire que l'on ne peut accéder aux données d'une partition principale (à toutes fins pratiques) que par le biais du système d'exploitation installé dans cette partition..Informations de base

S'il vous faut installer plus d'un seul système d'exploitation sur votre disque dur, vous devrez probablement créer des partitions principales multiples car la plupart des systèmes d'exploitation ne peuvent être amorcés qu'à partir d'une partition principale.

#### ❑ **Partitions étendues**

Les partitions étendues ont été inventées pour fournir un moyen de contourner la limite arbitraire de quatre partitions. Une partition étendue est essentiellement un " conteneur " dans lequel vous pouvez continuer à diviser physiquement votre espace disque en créant un nombre illimité de partitions logiques.

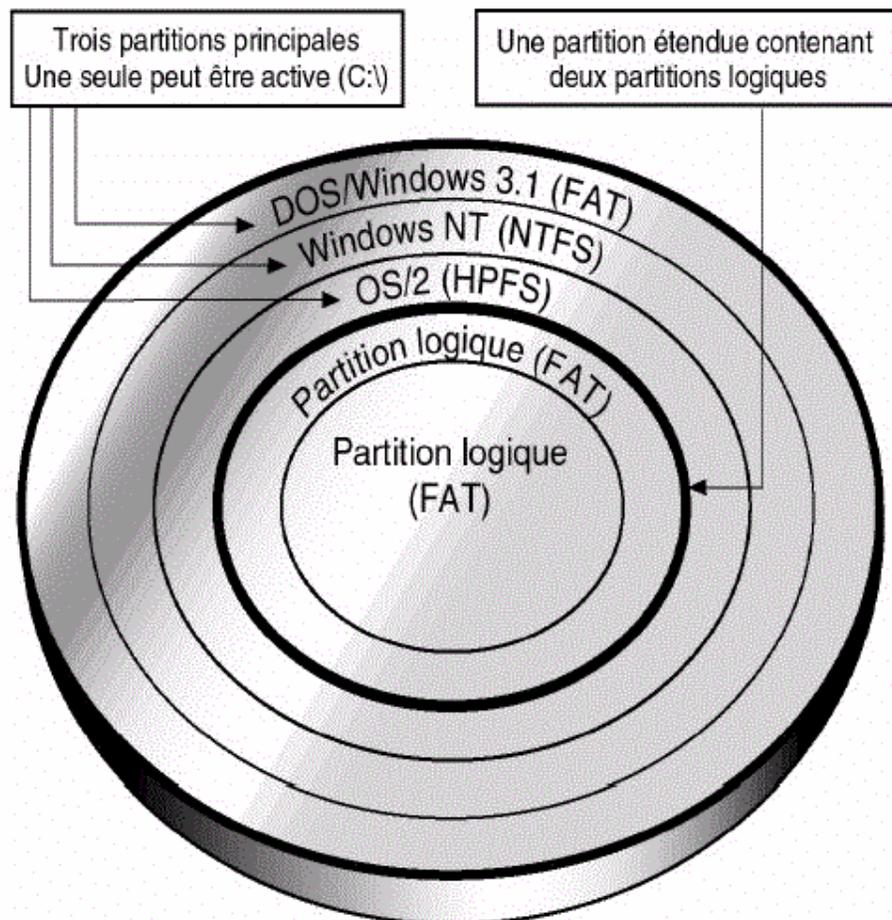
Une partition étendue ne contient pas directement de données. Vous devez créer des partitions logiques à l'intérieur d'une partition étendue. Ce sont les partitions logiques qui contiendront les données. Les partitions logiques doivent être logiquement formatées et chacune d'elle peut posséder un système de fichiers différent. Lorsque le formatage logique a été effectué, chaque partition logique constitue un volume disque distinct.

#### ❑ **Partitions logiques**

Les partitions logiques ne peuvent exister qu'à l'intérieur d'une partition étendue et sont destinées à ne contenir que des fichiers de données et des systèmes d'exploitation qui peuvent être amorcés à partir d'une partition logique (comme OS/2, par exemple). Les systèmes d'exploitation, tels qu'OS/2, qui peuvent être amorcés à partir d'une partition logique, sont normalement installés dans une partition logique, ce qui permet de conserver les partitions principales pour d'autres utilisations.

La figure ci-dessous illustre un disque dur qui contient quatre partitions, trois partitions principales et une partition étendue. La partition étendue a été divisée en deux partitions logiques. Chaque partition principale a été formatée pour utiliser différents systèmes de fichiers (FAT, NTFS et HPFS). Les partitions logiques ont toutes deux été formatées pour utiliser le système de fichiers FAT.

Bien que la figure ci-dessous indique toutes les partitions sur un seul côté du plateau, en utilisation réelle les partitions seraient probablement réparties des deux côtés de plusieurs



plateaux.

### Compréhension des lettres d'unité

Lorsque vous amorcez votre ordinateur, le système d'exploitation amorcé affecte des lettres d'unité (C:, D:, etc.) aux partitions principales et logiques de chaque disque dur. Pour faire référence aux fichiers contenus dans la partition, vous, ainsi que le système et les applications, utilisez les lettres d'unité affectées par le système d'exploitation.

Le système d'exploitation change l'affectation des lettres d'unité lorsque vous ajoutez ou supprimez un second disque dur, ou lorsque vous ajoutez, supprimez ou copiez une partition sur n'importe quel disque. L'affectation des lettres d'unité peut également changer en fonction du système d'exploitation que vous amorcez, ou encore si vous reformatez une partition au moyen d'un autre système de fichiers. Si les lettres d'unité changent, la configuration de votre système peut, en

certaines parties, ne plus être valide, comme par exemple, les commandes de démarrage d'une application qui sont basées sur une lettre d'unité.

Pour éviter des modifications de configuration et pour résoudre les problèmes liés à la configuration, il faut comprendre plusieurs éléments, à savoir, comment le système d'exploitation affecte des lettres d'unité, les divers types de problèmes qu'entraînent un changement de lettre d'unité, ce que vous pouvez faire lors du partitionnement pour éviter des changements de lettre d'unité, et en dernier lieu, comment résoudre les problèmes de configuration découlant de changements inévitables.

#### □ Comment le système d'exploitation affecte des lettres d'unités

Il est important de comprendre l'ordre dans lequel un système d'exploitation affecte des lettres d'unité

Des lettres d'unité sont d'abord affectées aux partitions principales dans l'ordre dans lequel elles apparaissent sur les disques durs. La lettre d'unité C: est affectée à la partition principale active sur le premier disque dur, la lettre D: est ensuite affectée à la première partition principale reconnue sur le disque dur suivant et ainsi de suite, jusqu'à ce qu'une lettre ait été affectée à la première partition principale reconnue sur tous les disques durs.

Des lettres d'unité sont ensuite affectées à toutes les partitions logiques dont le système d'exploitation reconnaît le système de fichiers, en commençant par celles qui se trouvent sur le premier disque dur et en procédant dans l'ordre.

Une lettre d'unité est enfin affectée aux lecteurs de CD-ROM et à tout autre type de

Examinons quelques exemples qui illustrent comment les lettres d'unité sont affectées.

Considérons tout d'abord un ordinateur avec un disque dur sur lequel DOS est installé. La figure ci-dessous illustre le partitionnement du disque.

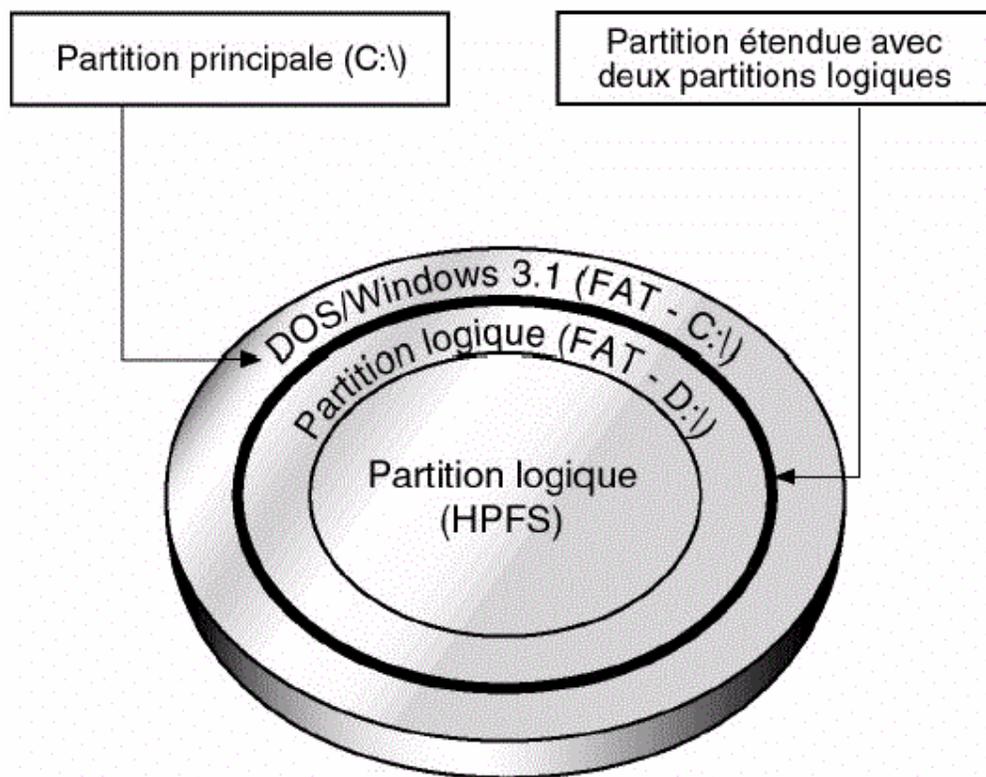


Figure 4

Le disque est partitionné en une partition principale et une partition étendue qui elle-même contient deux partitions logiques. La partition principale est formatée avec le système de fichiers FAT dans lequel DOS a été installé. La première partition logique est formatée avec le système de fichiers FAT que DOS reconnaît, mais la deuxième partition logique est formatée avec le système de fichier HPFS que DOS ne reconnaît pas.

Sur ce disque, DOS affecte la lettre d'unité C: à la partition principale et la lettre d'unité D: à la première partition logique, mais n'affecte pas de lettre d'unité à la deuxième partition logique car il ne reconnaît pas le système de fichiers de cette unité (HPFS).

Considérons maintenant un ordinateur identique à celui du premier exemple, à la différence près qu'un second disque dur a été installé. Cette configuration est illustrée dans la figure ci-dessus.

## Disque dur SCSI, technologie RAID

Les données (et son support principal, le disque dur) sont primordiales dans toutes applications informatiques. Dans le cas de serveurs réseaux, deux directions sont utilisées pour augmenter la vitesse et garantir la fiabilité en cas de panne d'un disque dur: l'utilisation de solutions SCSI pour le stockage et l'implantation du RAID. Les deux sont généralement implantés en même temps.

### Technologie SCSI.

Les connexions SCSI (Small Computer System Interface) ont par rapport aux périphériques IDE plusieurs avantages.

1. Le **nombre de périphériques** connectés sur la même nappe est supérieur: jusque 15.  
L'adresse se fait par pontage ou sélecteur rotatif
2. Une connexion SCSI est **multi-session**. En clair, le disque peut exécuter plusieurs applications en même temps (ou presque) et l'écriture (ou la lecture) d'un fichier ne doit pas attendre la fin de l'opération précédente. Ceci est le cas par exemple lorsque l'on écrit un gros fichier sur le disque dur, la lecture suivante ne doit pas attendre pour démarrer. Le SCSI est donc capable de gérer plusieurs requêtes de lecture /écriture en parallèle (jusque 255), au contraire de l'IDE qui ne peut traiter qu'une opération (complètement) à la fois. Ceci ne fonctionne qu'avec un système d'exploitation compatible (Win NT, Win 2000 ou XP Pro, Novell), à l'exclusion de Win 95 et consort.
3. Les périphériques SCSI peuvent être **internes ou externes**. Le transfert de données entre deux périphériques se fait directement par DMA entre les deux périphériques.
4. Un disque dur SCSI inclut automatiquement un **contrôle d'erreur** des données.

Tout ceci explique que les vitesses soient plus élevées qu'en IDE (même à taux de transfert équivalent), avec des prix qui varient en même temps.

La norme SCSI a évolué, mais on rencontre encore quasiment toutes les normes selon les appareils à connecter. **NARROW** se réfère à un bus 8 bits, **WIDE** se réfère à un bus 16 bits.

Normes SCSI	Taux transfert MB/s maximum	Largeur bus (en bits)	Longueur maximum câble SCSI			Nombre conducteurs	Nb. max. connexion (non compris la carte contrôleur)
			SE	LVD	HVD		
SCSI (NARROW) <b>SCSI 1</b>	5	8	6 m	-	-	25	7
FAST NARROW SCSI	10	8	3 m	-	-	50	7
Fast Wide SCSI <b>SCSI 2</b>	20	16	3 m	12 m	25 m	68 ou 80	15
Ultra SCSI Narrow	20	8	3 m	-	-	50	3

Ultra SCSI Narrow	20	8	1,5 m	-	-	50	7
Wide Ultra SCSI <b>SCSI 3</b>	40	16 bits	3 m	-	-	68 ou 80	3
Wide Ultra SCSI	40	16 bits	1,50	-	-	68 ou 80	7
Wide Ultra SCSI <b>SCSI 3 DIFFERENTIEL</b>	40	16 bits	-	12 m	25 m	68 ou 80	15
Ultra 2 SCSI (Narrow)	40	8 bits	6 m	-	-	50	7
Wide <b>Ultra 2 SCSI</b>	80	16 bits	-	12 m	25 m	68 ou 80	15
Wide Ultra 160 <b>Ultra 3 SCSI ou SCSI 5</b>	160	16 bits	-	12 m	-	68 ou 80	15
Wide Ultra 320 SCSI	320	16 bits	-	12 m	-	68 ou 80	15
<b>Par comparaison</b>							
E-IDE-ATA133	133	16 bits	-	80 fils (connecteur 40)	2		

Le taux de transfert, la longueur des cordons, le nombre de fils et de périphériques diffèrent selon la norme.

Ce qu'il est important de connaître pour commander un cordon ou un terminateur :

1. Le nombre de fils et le modèle des connecteurs.
2. La norme SE, LVD ou HVD.

**Tous les autres renseignements sont superflus sur le plan des connecteurs.**

**SCSI 1 (NARROW SCSI):** Il est codé sur 8 bits seulement

**SCSI 2 (WIDE SCSI) :** Codé sur 16 bits il autorise un taux de transfert jusqu'à 20 MB/s



Carte SCSI 2 Adaptec AH 3940. Cette carte est identique à une 2940 mais possède 2 canaux internes distincts (RAID)



Connecteur SCSI2 interne

**SCSI 3 (ULTRAWIDE SCSI) :** il permet d'atteindre des taux de transfert de 40 MB/s

- SCSI 3 DIFFERENTIEL : il utilise du câble de très bonne qualité appairé et véhicule par paires deux signaux, le signal utile étant la différence entre les deux. Il existe 2 types de chaînes différentielles :
- **le HVD (High Voltage Differential)** qui travaille sous 5 volts et permet d'atteindre des longueurs de chaîne de 25 m alors que l'UltraWide ne permet pas de dépasser 6 m. Utilisé surtout dans des configurations professionnelles.
- **le LVD (Low Voltage Differential)** qui travaille sous 3,3 volts et permet d'atteindre 12 m de longueur de chaîne. Il utilise les nappes internes LVD, cordons LVD, terminateurs externes LVD et les terminateurs internes LVD.

Le connecteur est spécifique à chacune des 3 normes : les cordons et les nappes internes n'ont pas la même impédance en Ultra Wide, en HVD ou en LVD. Les terminateurs aussi sont différents. Il ne faut donc pas mélanger les différents types de cordons et terminateurs entre eux, surtout le HVD incompatible avec le reste. Toutefois la norme LVD permet de connecter des périphériques non LVD sur la carte hôte et à l'inverse de connecter des composants LVD sur une carte hôte non LVD. Bien entendu les débits ne seront pas dans ce cas ceux du LVD.

**La norme Ultra 2 SCSI LVD** est une extension du SCSI 3. Sur les connecteurs, les signaux sont véhiculés dans deux fils à la fois, le signal utile étant la différence entre les deux. Cette méthode est également utilisée en liaison Ethernet (entre autre). Les cordons LVD doivent de haute qualité et à la bonne impédance. La vitesse maximum théorique est de 80 MB/s.

#### **La norme Ultra 3 SCSI ou Ultra 160/m ou SCSI 5 (SCSI PARALLEL INTERFACE SPI-3):**

L'Ultra 160/m est une implantation spécifique de la norme Ultra 3 SCSI et ne retient que 3 éléments de cette norme :

1. Taux de transfert doublé par rapport à l'Ultra 2 SCSI : 160 Mo/s au lieu de 80 Mo/s.
2. Test physique du bus SCSI par le contrôleur, au démarrage, permettant de déterminer la vitesse de travail en fonction des différents éléments de la chaîne SCSI. Il va sans dire que la qualité des cordons et terminateurs jouera un rôle décisif dans la vitesse globale de la chaîne.
3. Contrôle de redondance cyclique (CRC) qui permet le contrôle d'erreurs dans la transmission des données. Ce sont les caractéristiques essentielles qui caractérisent l'Ultra 160/m, le "/m" signifie que cette norme est manageable (test physique et CRC). La désignation utilisée pour l'Ultra 2 SCSI, LVD, est impropre car l'Ultra 3 SCSI est lui aussi LVD. Il utilise les mêmes connecteurs, donc cordons et terminateurs sont identiques, mais doivent être plus encore ici d'excellente qualité, compte tenu des débits très élevés. On peut aussi utiliser 15 périphériques sur une longueur totale de 12 m. Il est manageable car une carte Ultra 160/m saura gérer des périphériques connectés dans la même chaîne avec des vitesses différentes, en respectant leurs vitesses respectives : 80 Mo/s pour l'Ultra 2 SCSI ET 160 Mo/s pour le 160/m. Ainsi c'est la règle du maillon le plus faible qui détermine la vitesse globale de la chaîne disparaît. A l'inverse un contrôleur Ultra 2 SCSI pourra aussi gérer les 2 normes en même temps, mais à une vitesse unique de 80 Mo/s. Conclusion : totale compatibilité entre les 2 normes.

#### **La norme Ultra 320 SCSI (SCSI PARALLEL INTERFACE SPI-4) :**

Cette nouvelle norme enrichit tout en gardant les spécifications de la précédente norme Ultra 160/m :

- Transfert d'unités d'information (IU transfert ou packetization). Les informations indépendantes du flux de données, par exemple les commandes échangées entre la carte hôte et le disque dur, sont transférées à la vitesse nominale soit 320 Mo/s.
- Multiplexage des tâches d'entrées/sorties sans attendre la phase du BUS FREE. Continuité du flux de données sans phases d'inertie et exploitation optimisée des canaux disponibles. - La chaîne de commandes pour l'envoi d'un paquet de données se simplifie et les différentes étapes sont moins nombreuses, d'où une amélioration des taux de transferts.
- Rectification du signal de données par rapport au signal d'horloge (skew compensation). Skew est la différence de temps dans l'acheminement de deux signaux différents en provenance d'un même émetteur vers deux cibles différentes situées dans le même bus de traitement. Il peut s'agir d'une carte hôte émettant des signaux différents vers deux disques durs situés sur la même chaîne SCSI. Pour maintenir la logique du système, le délai entre les deux signaux est arbitré par un signal d'horloge. La vitesse de traitement fait que tous les périphériques Ultra 320 accomplissent une compensation en mode réception car un écart d'une nanoseconde peut faire la différence entre un signal valide et une transmission incorrecte.



DB 25

Centronix 50 mâle externe (plus de pin que le connecteur centronix imprimante standard)



SUB 50 mâle



Connecteur DB68HD

### Connexion de périphériques SCSI

Les types de périphériques qui sont connectés en SCSI sont

- Disque dur (interne ou externe)
- Sauvegarde sur bandes (DAT, DLT, ...) en interne ou externe
- Scanner (externe)

Le choix du numéro de périphérique se fait soit par pontage, soit par un sélecteur rotatif. L'adresse doit être unique sur un même câble ou plutôt sur un même contrôleur. En effet, les périphériques internes et externes partageant généralement le même contrôleur.

La fin de la liaison des périphériques externes doit se terminer par une résistance de terminaison spécifique au type de connexion SCSI. En effet, pour les périphériques externes, la connexion est chaînée. On débute du contrôleur vers le périphérique. Le câble suivant passe du périphérique au périphérique suivant. Dans certains périphériques, la terminaison est incluse dans le périphérique (pontage à insérer pour l'activer).

Les câbles standards internes reprennent généralement seulement 3 connecteurs mais des modèles reprenant plus de périphériques se trouvent dans le commerce.

Généralement, on ne mélange pas les périphériques SCSI de différents types, néanmoins des terminaisons spéciales permettent de chaîner des périphériques NARROW (50 pin) avec des ULTRA WIDE (68 pin), tant en interne qu'en externe.

### **RAID (Redundant Array of Independent Disk).**

Cette possibilité nécessite un contrôleur disque dur particulier et ... toujours un système d'exploitation spécifique réseau lourd (une fois de plus, évitez Win 95 / 98). Adaptec fournit des contrôleurs RAID IDE, mais ce n'est pas la majorité des cas pour des questions de performances. Certains systèmes d'exploitation réseau permettent également, dans une moindre mesure, d'effectuer du **RAID logiciel**. Ceci n'est pas conseillé puisque c'est le processeur qui remplace le travail des contrôleurs, d'où un ralentissement.

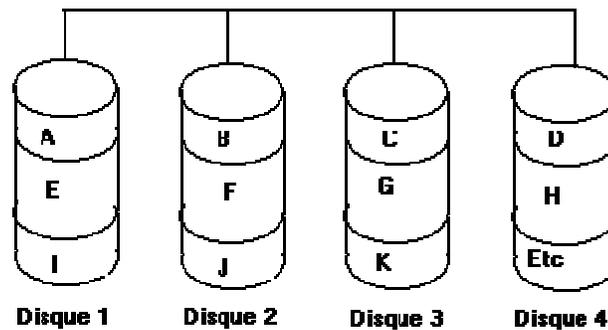
Le RAID permet d'associer plusieurs unités de disques durs de même type en une seule grappe. En cas de défaillance d'un disque, les versions de RAID autres que 0 permettent de reconstruire les données perdues à partir des autres unités de la grappe. L'implantation de cette technologie peut être de nature matérielle, logicielle ou passer par une combinaison des deux. Les solutions proposées peuvent être regroupées en six grandes familles, du Raid 0 au RAID 5.

#### **RAID 0 (striping)**

Le **RAID 0** n'apporte aucune sécurité des données, il augmente seulement le taux de transfert des informations. L'information est découpée à la suite de l'autre sur l'ensemble des disques durs (4 disques dans le cas ci-dessous mais plus généralement 2) L'unité qui regroupe les disques bénéficie de la somme des débits de chaque disque. Un accès (opération de lecture) ou un transfert (opération d'écriture) s'effectue en simultané sur l'ensemble de la grappe sur un mode parallèle. La panne d'un seul disque dur provoque la perte de toutes les données.

La capacité globale est équivalente à la somme des capacités de chaque disques durs (de capacités identiques).

Signal de départ A B C D E F G H I J ...

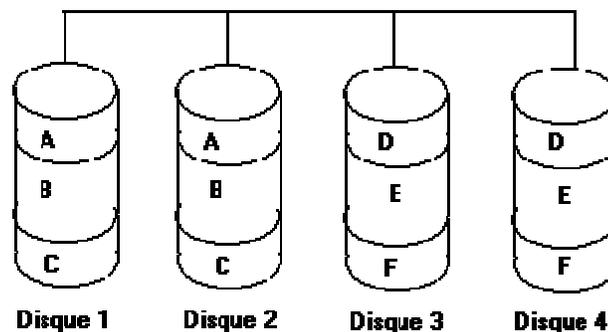


## RAID 0

### RAID 1 (Mirroring)

Dans les liaisons RAID 1, les données sont dupliquées intégralement sur un second disque ou sur un second groupe de disques durs. Les performances en lecture sont doublées grâce à l'accès simultané aux deux disques (à condition d'avoir deux contrôleurs séparés). Cette méthode réduit néanmoins la capacité de stockage. Il correspond au Mode logiciel "**Miroir**" de Windows NT4 ou Windows 2000 Server.

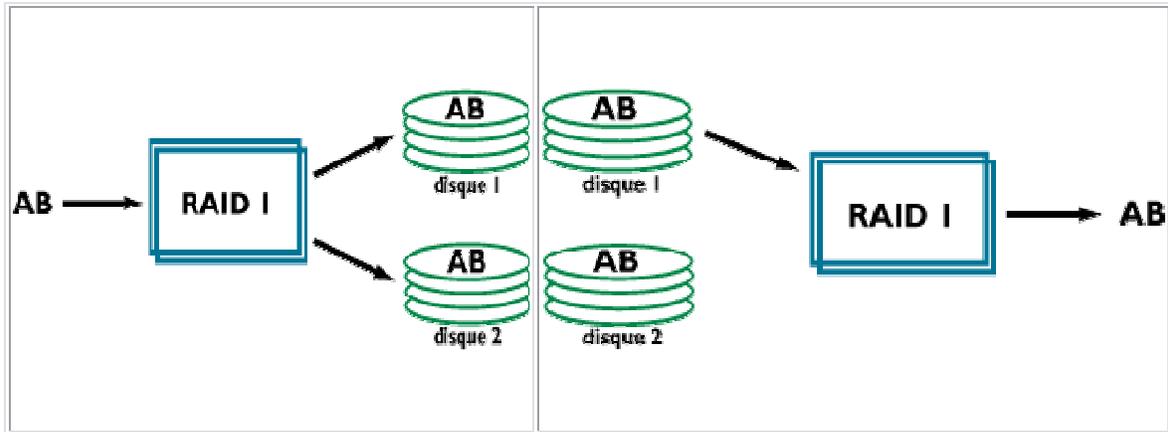
Signal de départ A B C D E F G H I J ...



## RAID 1

Ecriture

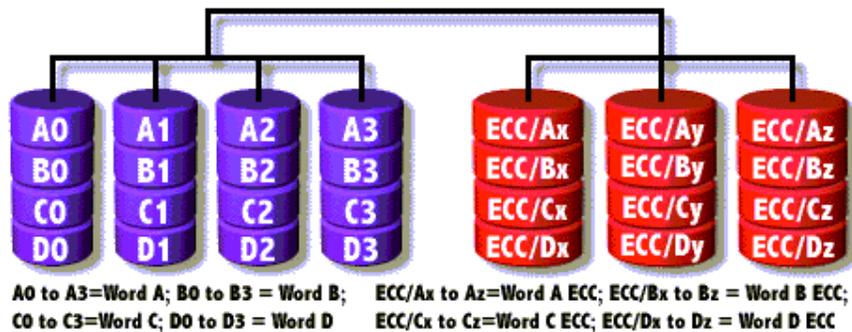
Lecture



### RAID 2

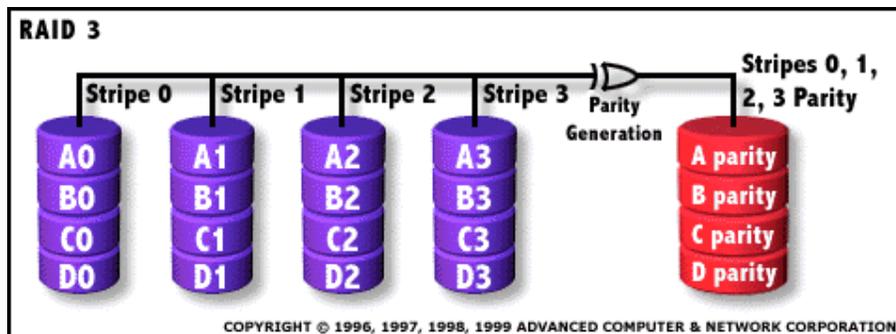
**RAID 2** repose sur une grappe avec plusieurs disques de parité et une synchronisation des accès. Cette technologie est peu utilisée de par sa complexité et le surcoût qu'elle entraîne. Il est identique au RAID 0 avec un contrôle des données intégré. On utilise généralement 3 disques de contrôle pour 4 disques de données. La méthode de correction est l'ECC. Néanmoins, tous les disques durs SCSI incluent ce contrôle d'erreur, d'où le caractère anecdotique de ce mode.

### RAID 2



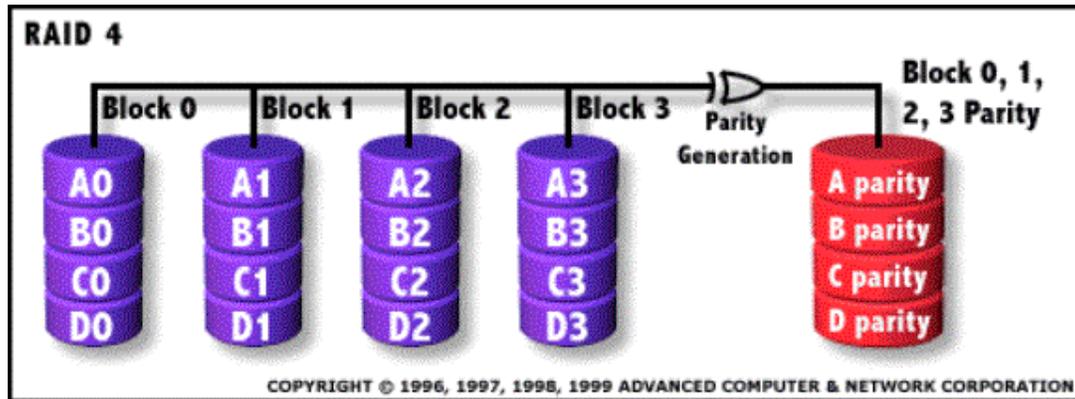
### RAID 3

Le RAID 3 est basé sur des grappes de disques identiques avec une unité de stockage réservée au stockage du bit de parité. Si le disque de parité tombe en panne, on se retrouve en RAID 0. La sécurité des données est peu implantée. Attention, le disque de parité est 2 fois plus utilisé que les autres disques.



## Le RAID 4

**RAID 4** se différencie du RAID 3 par une gestion asynchrone des unités. Même si les accès fonctionnent en parallèle sur les différentes unités, le disque de parité est plus souvent sollicité. Ceci implique en RAID 3 un goulot d'étranglement. La seule différence avec le RAID 3 est la structure des données implantée sur les disques.



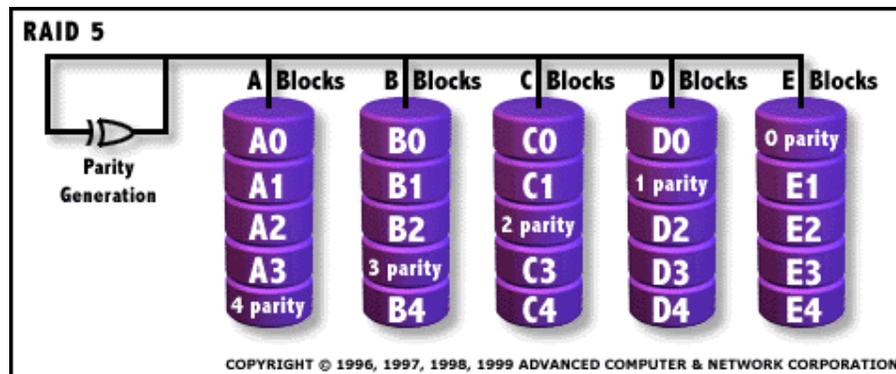
En **écriture**, les données sont décomposées en blocs de petite taille et répartis sur les différents disques composant le RAID 4. Simultanément, le contrôle de parité s'inscrit sur le disque dédié à cet effet.

Ecriture	Lecture :
<p>The diagram shows data 'AB' entering a RAID 4 controller. The controller distributes the data to three data disks labeled 'disque A', 'disque B', and 'disque C'. Simultaneously, the parity is calculated and written to the 'disque de parité' (parity disk), which contains 'A+B'.</p>	<p>The diagram shows data being read from a RAID 4 controller. The controller reads data from three data disks labeled 'disque 1', 'disque 2', and 'disque de parité' (containing 'A+B').</p>
<p><b>Avantages :</b></p>	<p><b>Inconvénients :</b></p>
<ul style="list-style-type: none"> <li>• Tolérance de panne et parité centralisée sur un disque dédié.</li> <li>• Parité : processus algorithmique permettant au système de reconstituer une donnée défectueuse ou manquante à partir de l'information de parité mémorisée au cours de l'écriture.</li> <li>• Un sous-système en RAID 4 présente un rapport capacité/ performance/ investissement intéressant.</li> <li>• Les performances en lecture des piles de disques de niveau RAID 4 sont excellentes (équivalentes au RAID 0).</li> </ul>	<ul style="list-style-type: none"> <li>• Le défaut majeure provient de la <b>mise à jour</b> des données de parité qui <b>dégrade les performances</b> de Raid 4. Pour cette raison, le RAID 5 est toujours préférés aux systèmes RAID de niveau 4.</li> </ul>

- Puisqu'il n'y a pas duplication des données, mais uniquement **enregistrement des données de parité correspondantes**, le coût par méga-octet d'une solution de niveau RAID mode 4 demeure raisonnable.

## Le RAID 5

**RAID 5** s'apparente au RAID 4 avec un bit de parité distribué sur l'ensemble des unités de la grappe, supprimant ainsi le fameux goulot d'étranglement tout en bénéficiant des performances de la gestion asynchrone. Ce mode correspond à l'agrégat par bandes avec parité sous NT4 ou 2000 Server. Les disques travaillent tous autant. Lors d'un échange à chaud (hot plug) d'un disque dur, les données sont recrées à partir des autres disques durs.



Avantages:	Inconvénients:
<ul style="list-style-type: none"> <li>○ Bonne tolérance aux erreurs</li> <li>○ Enormément d'implantations commerciales</li> <li>○ Hot-spare</li> <li>○ Hot-plug</li> </ul>	<ul style="list-style-type: none"> <li>○ 3 disques au minimum</li> <li>○ En cas de problème, remise en ordre assez lente</li> <li>○ En cas de panne d'un disque dur, on revient en mode 0.</li> </ul>

### Applications:

- Les applications qui utilisent les E/S de manière aléatoire sur de petits volumes
- Typiquement les serveurs de bases de données

### Orthogonal RAID 5

**L'Orthogonal RAID 5**, technique logicielle développée par IBM, ce mode utilise un contrôleur par disque. Dans le langage courant, il est assimilé à un RAID 5.

### Les autres RAID

D'autres système RAID sont proposés. Ce ne sont que des RAID 5 évolués. Ces modes restent rares du fait d'une architecture complexe et un coût élevé.

- **RAID 6** utilise une double parité. Ce système permet de travailler avec 2 disques durs défectueux, avec forcément une perte de performances.
- **RAID 7** met en jeux plusieurs disques de données couplé avec 1 ou plusieurs disques de parité. le contrôle des données et de la mémoire cache et le calcul de la parité se fait par un microcontrôleur. Ceci donne des performances jusque 50 % plus rapides que dans les autres modes RAID. Cette solution est une marque déposée de Storage Computer Corporation
- **RAID 10** implante le striping (RAID 0) couplé au Mirroring (RAID 1). Une excellente mais coûteuse solution de tolérance de panne. Ce système est le plus sûr et le plus rapide. Néanmoins, sa difficulté de mise en oeuvre et le prix le rendent peu utilisé.

### RAID hardware et software.

Le RAID peut être géré de manière hardware ou logiciel. La gamme des systèmes d'exploitation "professionnels" de Microsoft peuvent gérer le RAID de manière logicielle: Windows NT, 2000, 2003 et 2008 en version serveur gèrent le RAID 0, 1, 5

- Netware gère en mode natif (logiciel) le RAID 1
- Linux gère les RAID 0, 1, 4 et 5
- Les séries "amateurs" (DOS, WIN95/98/Me et XP Home) ne gèrent pas le RAID en mode natif.

Cette manière de procéder permet au système d'exploitation d'utiliser plusieurs disques en mode RAID sans hardware dédié. Par contre, en utilisant un hardware (avec le pilote associé) adéquat, **tous les systèmes d'exploitation peuvent travailler en RAID en théorie**. Pour rappel, les systèmes d'exploitation "amateurs" de Microsoft ne tirent pas non plus partie des possibilités de l'SCSI, notamment pour le transfert par DMA simultané de données entres disques durs.

### Solutions RAID hardware.

Diverses cartes sont proposées sur le marché de solutions RAID pour disques durs SCSI (notamment ADAPTEC). Actuellement les cartes mères standards incluent directement des solutions RAID en IDE ou S-ATA. Cette dernière possibilité permet d'installer les RAID 0 et RAID 1.

## Les IRQ et les DMA

Au fur et à mesure que l'on ajoute un périphérique, on doit s'assurer qu'il n'empiète pas sur d'autres déjà en place. Chaque périphérique doit obtenir l'accès à certaines ressources système pour en avoir le contrôle ou pour échanger des données. Cependant aujourd'hui avec le bus PCI et le Plug and Play, la configuration des cartes est bien plus facile.

### Les interruptions

Les ordinateurs sont des machines multitâches, quand ils sont occupés à sauvegarder ou imprimer, ils peuvent faire autre chose à condition d'être prêt à recevoir ces nouvelles sollicitations. Pour cela l'unité centrale utilise les **interruptions, IRQ (Interrupts ReQuests)**. Ce mécanisme permet aux périphériques de signaler eux-mêmes leur besoin d'accès au système. Quand un périphérique active son interruption, il demande au système d'arrêter ce qu'il fait pour s'occuper de sa demande. Si deux appareils sont sur la même interruption, il y a conflit.

### Les différentes interruptions.

0	L'horloge système	8	Horloge en temps réel
1	Le clavier	9	Cascade à IRQ 2
2	Cascade à IRQ 9	10	Libre
3	Com 2 et Com 4	11	Libre
4	Com 1 et Com 3	12	Souris PS/2 /Import
5	LPT2	13	Coprocasseur mathématique
6	Contrôleur de disquettes	14	Contrôleur de disque dur
7	LPT1	15	Second contrôleur de disque dur.

**Les numéros les plus faibles ont la priorité la plus importante**, c'est pourquoi 0 et 1 sont réservés à l'horloge système et au clavier. Par conséquent, la mise en cascade de IRQ 2 pour aller vers IRQ 9 donne une priorité plus grande à IRQ 10 qu'à IRQ 5, c'est pourquoi une carte réseau sera installé de préférence sur IRQ 10 ou 11.

### Les adresses E/S (Entrée/Sortie).

Si le périphérique après avoir demandé une interruption obtient l'accès au système, il lui faut un lieu pour communiquer ses données. Ceci se fait en lui attribuant des blocs de mémoires. Deux types d'attribution de mémoire peuvent être faits.

- **Les adresses E/S** (aussi appelés ports), sont les petites adresses situés dans la mémoire base du système.
- **Les adresses de la mémoire partagée**, qui permettent aux périphériques d'utiliser plus de mémoire vive du système.

Les adresse E/S se trouvent dans la mémoire de rang 100H jusqu'à 3FFH. Un port sera de 8 à 32 octets. Malheureusement la documentation des périphériques (cartes additionnelles) ne précise pas toujours la taille du port que le périphérique requiert.

### Les attributions habituelles des ports E/S.

Adresse de base	Dispositif	Rang d'adresse typique.
200	Port de jeu	200-20F
260	LPT2	260-27F

2E8	Com 4	2E8-2EF
2F8	Com 2	2F8-2FF
300	Réglage habituel pour de nombreuses cartes réseaux.	300-31F
330	Adaptec et adaptateurs SCSI	330-33F
360	LPT1	360-37F
3C0	Ecran EGA VGA	3C0-3DF
3E8	Com 3	3E8-3EF
3F8	Com 1	3F8-3FF
370	LPT2	370-38F

Les adresses de mémoire partagée se trouvent au-delà des 640 Ko du Dos. On les sélectionne dans les rangs C0000h à EFFFFh.

Les plages A0000h-BFFFFh sont utilisées par l'affichage vidéo, et F0000-FFFFh par la Rom Bios.

### Canaux DMA. (Direct Memory Access)

Ils permettent aux périphériques de communiquer directement avec la mémoire sans faire appel à l'unité centrale. 8 canaux DMA sont disponibles. L'adresse DMA 0 est réservé au rafraîchissement de la mémoire DRAM.

### Le PLUG And PLAY

Au fur et à mesure que l'on ajoute un périphérique, sur un ordinateur, on doit s'assurer qu'il n'empiète pas sur d'autres déjà en place, Dans ce but, faut attribuer aux différents composants de l'ordinateur des autorisations d'accès aux ressources du système, par l'intermédiaire des IRQ, des E/S, des DMA..

Ainsi quand on voulait installer une nouvelle carte, (son, modem, réseau), il fallait, par un réglage de cavalier, déclarer l'interruption, utilisée. Mais pour éviter tout conflit on devait connaître les adresses d'entrée/sortie, l'IRQ et le DMA encore disponibles

### 1993 Naissance du Plug and Play.

Tout repose sur le Bios qui au démarrage interroge chaque carte sur ses ressources qui lui sont nécessaires en vue de répartir celles dont il dispose. Evidemment cela ne fonctionne qu'avec

- **Un Bios PnP**
- **Des cartes PnP capables de communiquer avec le Bios.**

Au démarrage du PC, le Bios examine les cartes non PnP pour lesquelles les interruptions sont attribuées de manière fixe, puis il se renseigne auprès des composants des cartes PnP afin de leur affecter automatiquement les IRQ et les canaux DMA encore libres.

Chaque carte PnP possède dans une mémoire de 256 octets les informations dont le Bios a besoin. Une fois la répartition effectuée, ces paramètres sont inscrits dans la **NVRAM**, une mémoire non volatile, en tant qu'informations supplémentaires à prendre en compte à chaque démarrage du P.C. C'est l'**ESCD** (Extended System Configuration Data). Puis à chaque démarrage le Bios effectue une comparaison avec cet enregistrement pour voir s'il n'y a pas eu de modification.

Par la suite les systèmes d'exploitation PnP comme Windows 95/98 lisent l'ESCD et configurent leur pilote en conséquence.

### ***Mais en pratique.***

Windows 95 étant un système d'exploitation PnP, en liaison avec le Bios PnP, si dans le Setup (section PNP/PCI) on a déclaré l'existence de ce système, le Bios fait un strict minimum de réservation de ressources et laisse à Windows le gros du travail.

De plus Windows peut mettre à jour l'**ESCD**, (qui normalement ne dépend que du Bios) Windows permet la réservation de ressources (IRQ, DMA) qui dans ce cas ne seront plus affectées aux périphériques PnP.

Tout ce travail réalisé par Windows peut éventuellement être désactivé dans le gestionnaire de périphérique, sous l'icône Périphérique système/Bios Plug and Play, [Désactiver les mises à Jour NVRAM/ESCD], mais ceci est déconseillé car Windows fonctionne mieux que le Bios. Windows 98 reprend les mêmes fonctionnalités mais avec des routines d'installation qui reconnaissent encore mieux le matériel.

### **Faire cohabiter une carte non PnP avec un système PnP.**

Dans la plus part des cas, Windows 95 et encore plus Windows 98 sont capable de détecter la carte. En cas d'échec, la carte ISA non PnP sera configurée à partir de ses cavaliers, puis dans le Bios PnP, ces ressources devront être réservées dans la section PNP/PCI par l'option Legacy ISA ou Used by ISA.

### **Faire cohabiter du matériel PnP avec un Bios PnP et un système non PnP.**

Dos, Windows 3.11, ne sont pas des systèmes PnP, ils ignorent tout de la détection automatique des périphériques, par contre Windows NT sans être un système PnP peut grâce à ses fichiers d'information INF identifier les périphériques déclarés compatible NT (Microsoft indique ces périphériques dans la **HCL (Hardware Compatibility List)**)

Dans ce cas l'absence d'un système non PnP doit être signalé au Bios et c'est le Bios qui se charge d'affecter les ressources et d'enregistrer l'ESCD.

### **Faire cohabiter du matériel PnP avec un Bios non PnP et un système non PnP.**

On est de moins en moins confronté à ce problème, mais on le rencontre parfois sur d'ancien PC sur lesquels ont été installé une carte son PnP.

Comme les périphériques PnP ont une mémoire qui contient les informations nécessaires à leur configuration, il faut un utilitaire pour les configurer et émuler le PnP. L'utilitaire le plus répandu est ICU (Isa Configuration Utility), il analyse l'équipement matériel du PC, puis simule un Bios PnP afin d'affecter à la carte les bonnes ressources. Cette configuration est ensuite enregistrée

dans un fichier sur le disque dur. Un pilote installé dans le fichier Config.sys est chargé de contrôler cette émulation PnP.

Souvent les constructeurs de cartes proposaient aussi leurs propres utilitaires, ainsi Creative Labs proposait CTCM, un utilitaire pour affecter les ressources de la carte Sound Blaster quand elle devait fonctionner sous Dos sur une machine dont le Bios n'était pas PnP.

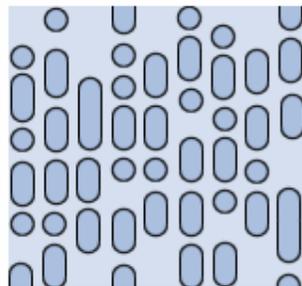
### Présentation du DVD

Le **DVD** (*Digital Versatile Disc*, plus rarement *Digital Video Disc*) est une «alternative» au disque compact (CD) dont la capacité est six fois plus importante (pour le support DVD de moindre capacité, simple face, simple couche). Le format DVD a été prévu afin de fournir un support de stockage universel alors que le CD était originalement prévu en tant que support audio uniquement.

Le format DVD est prévu afin de rendre les données adressables et accessibles aléatoirement (de manière non séquentielle). Il possède une structure complexe, permettant une plus grande interactivité, mais nécessitant l'utilisation de microprocesseurs évolués.

Le format DVD était porté originalement (dès le 15 septembre 1995) par un consortium de dix sociétés du monde multimédia (Hitachi, JVC, Matsushita, Mitsubishi, Philips, Pioneer, Sony, Thomson, Time Warner et Toshiba). A partir de 1997, un nouveau consortium, baptisé «DVD Forum» a succédé au consortium initial.

Un DVD peut facilement être confondu avec un CD dans la mesure où les deux supports sont des disques en plastique de 12 cm de diamètre et de 1.2 mm d'épaisseur et que leur lecture repose sur l'utilisation d'un rayon laser. Toutefois, les CD utilisent un laser infrarouge possédant une longueur d'onde de 780 nanomètres (nm) tandis que les graveurs de DVD utilisent un laser rouge avec une longueur d'onde de 635 nm ou 650 nm. De plus, les lecteurs de CD utilisent généralement une lentille dont la focale vaut 0,5, alors que les lecteurs de DVD sont basés sur une lentille ayant une focale de 0,6. Ainsi, les DVD possèdent des alvéoles dont la taille minimum est de  $0,40\mu$  avec un espacement de  $0,74\mu$ , contre  $0,834\mu$  et  $1,6\mu$  pour le CD.



L'intérêt du DVD est essentiellement sa capacité de stockage, ce qui en fait le support par excellence pour la vidéo. Un DVD de 4,7 Go permet ainsi de stocker plus de deux heures de

vidéo compressée en MPEG-2 (*Motion Picture Experts Group*), un format qui permet de compresser les images tout en gardant une grande qualité d'image.

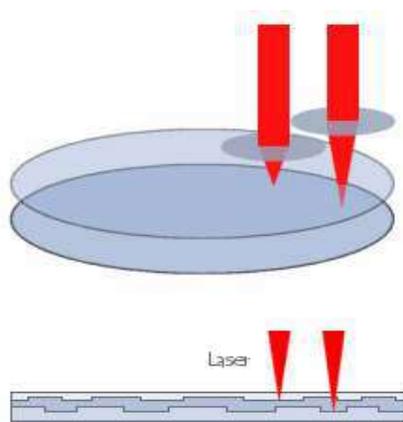
### Structure physique

Les DVD existent en version «simple couche» et «double couche» (en anglais «Dual Layer», noté *DL*). Ces derniers sont constitués d'une couche translucide semi réfléchissante à base d'or et d'une couche réflexive opaque à base d'argent, séparées par une couche de liaison (*bonding layer*). Pour lire ces deux couches le lecteur dispose d'un laser pouvant changer d'intensité en modifiant sa fréquence et sa focale :

- avec une intensité faible le rayon est réfléchi sur la surface dorée supérieure,
- avec une intensité plus élevée le rayon traverse la première couche et est réfléchi sur la surface argentée inférieure.

La couche inférieure possède toutefois une densité moindre. De plus l'information y est stockée «à l'envers» sur une spirale inversée, afin de limiter le temps de latence

lors du passage d'une couche à une autre



Par ailleurs, les DVD existent en version simple face ou double face, à la manière des disques vinyles. Dans le second cas, l'information est stockée de part et d'autre du support.

On distingue généralement 4 grandes familles de supports DVD, possédant différentes capacités selon leurs caractéristiques physiques :

Type de support	Caractéristiques	Capacité	Temps musical équivalent	Nombre de CD équivalent
CD		650Mo	1h14 min	1
DVD-5	simple face, simple couche	4.7 Go	9h30	7
DVD-9	simple face double couche	8.5 Go	17h30	13
DVD-10	double face, simple couche	9.4 Go	19h	14

DVD-17	double face, double couche	18 Go	35h	26
--------	-------------------------------	-------	-----	----

### Formats standard de DVD

Les spécifications officielles des DVD se déclinent en cinq livres :

- le livre A (*Book A*) pour le DVD-ROM ;
- le livre B (*Book B*) pour le DVD Vidéo ;
- le livre C (*Book C*) pour le DVD Audio ;
- le livre D (*Book D*) pour le DVD inscriptible (DVD-R) et le DVD réinscriptible (DVD-RW). Le format DVD-R est un format inscriptible une seule fois (*Write-Once*), tandis que le format DVD-RW est un format réinscriptible, permettant l'effacement et la modification de données grâce à une phase capable de changer d'état ;
- le livre E (*Book E*) pour le DVD réinscriptible (également DVD-RAM, pour *DVD Random Access Memory*). Le DVD-RAM est un support réinscriptible utilisant une technologie du changement de phase pour l'enregistrement. Les DVD-RAM sont en réalité des cartouches composées d'un boîtier et d'un DVD. Certaines cartouches sont amovibles, afin de permettre la lecture d'un DVD-RAM dans un lecteur DVD de salon.

### Formats standard de DVD enregistrables

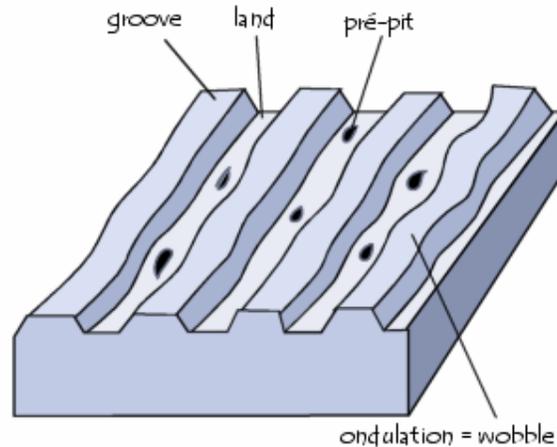
Il existe actuellement trois formats de DVD enregistrables :

- **DVD-RAM** de Toshiba © et Matsushita ©. Il s'agit d'un format essentiellement utilisé au Japon.
- **DVD-R / DVD-RW**, porté par le DVD Forum. Les DVD au format DVD-R sont enregistrables une seule fois tandis que les DVD au format DVD-RW sont réinscriptibles à raison d'environ 1 000 enregistrements. Le format DVD-R, ainsi que le format DVD-RW, permet d'obtenir une capacité totale de 4.7 Go.
- **DVD+R / DVD+RW**, porté par Sony et Philips au sein de la DVD+RW Alliance, regroupant, en plus des deux précédentes, les sociétés Dell, Hewlett-Packard, Mitsubishi/Verbatim, Ricoh, Thomson et Yamaha.

Ces trois formats sont incompatibles entre eux, malgré des performances équivalentes. Le format DVD-RAM ne fera pas l'objet de détails, dans la mesure où il est principalement utilisé au Japon. Les formats DVD-R(W) et DVD+R(W) sont par contre largement utilisés en Europe.

### DVD-R/RW

Le format DVD-R/DVD-RW est basé sur une technique dite du «**pré-pits**». À la manière des CD inscriptibles (CD-R), les DVD inscriptibles et réinscriptibles utilisent une «pre-groove» (spirale préalablement gravée sur le support), ondulant selon une sinusoïdale appelée *wobble*. La pre-groove permet de définir le positionnement de la tête d'enregistrement sur le support (appelé *tracking*) tandis que la fréquence d'oscillation permet au graveur d'ajuster sa vitesse. Les informations d'adressage (position des données) sont par contre définies grâce à des cuvettes pré-gravées sur le support, dans les creux (appelés *land*) entre les sillons du disque (appelés *groove*), baptisées «land pré pits» (abrégé en LPP).

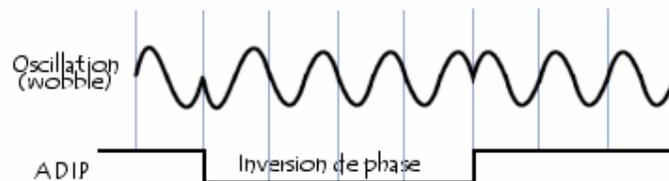


Les pré-pits constituent ainsi un second signal servant au positionnement des données. Lorsque le laser rencontre un pré-pit, un pic d'amplitude apparaît dans l'oscillation, indiquant au graveur où la donnée doit être gravée. Les spécifications du DVD-R précise qu'un pré-pit doit posséder une longueur d'au moins une période (1T).

Le format DVD-R/DVD-RW propose des fonctionnalités de gestion des erreurs, essentiellement logicielles (appelées *Persistent-DM* et *DRT-DM*).

### DVD+R/RW

Le format DVD+R/DVD+RW utilise une spirale dont l'oscillation (wobble) possède une fréquence beaucoup plus élevée que les DVD-R (817,4 kHz pour les DVD+R contre 140,6 pour les DVD-R) et gère l'adressage grâce à une modulation de la phase de l'oscillation, c'est-à-dire un codage par inversion de phase appelé **ADIP** (*ADdress In Pre-groove*). L'inversion de phase a lieu toutes les 32 périodes (32T).



Le format DVD+RW offre une fonctionnalité de correction d'erreurs appelée DVD+MRW (*Mount Rainier for DVD+RW* abrégé en *Mt Rainier for DVD+RW*) permettant de marquer les blocs défectueux. De plus, si des données lisibles existent sur ce bloc, un mécanisme permet de les déplacer sur un bloc sain et met à jour la table d'allocation des fichiers (on parle alors de *Logical to Physical Address Translation*).

En outre, une vérification en arrière-plan est prévue dans les spécifications, permettant de vérifier les erreurs présentes sur le disque, lorsque le lecteur est inactif. L'utilisateur peut néanmoins lire le support ou l'éjecter à tout moment, auquel cas les vérifications continueront où elles s'étaient arrêtées dès que le lecteur sera à nouveau en veille.

### Différences entre DVD+ et DVD-

D'une manière générale la méthode d'adressage utilisée par les DVD+R (modulation de phase) possède une meilleure résistance aux perturbations électromagnétiques que la méthode des pré-pits. En effet, lors de la gravure, le graveur doit également lire les pré-pits afin de positionner correctement les données sur le support. Or, la lumière émise par le laser peut provoquer des perturbations.

D'autre part, étant donné la période correspondant à la longueur d'un pré-pit (1T), les pré-pits sont d'autant plus difficiles à détecter que la vitesse de lecture est élevée. Il n'est donc pas étonnant que le premier graveur 16x commercialisé ait été au format DVD+RW.

Ainsi, le format DVD+R(W), bénéficiant de spécifications plus récentes, propose de meilleures performances ainsi que des fonctionnalités supplémentaires. En contrepartie, le format DVD-R(W) est ratifié par le DVD Forum et correspond au format initial, si bien qu'une majorité de lecteurs (notamment de salon) sont compatibles.

Les graveurs de DVD supportent généralement les deux formats de disques. En conclusion, compte tenu de sa meilleure compatibilité avec les platines de salon, le format DVD-R(W) est à privilégier pour la création de DVD Vidéo, tandis que le format DVD+R(W) comporte des avantages pour la création de DVD de données.

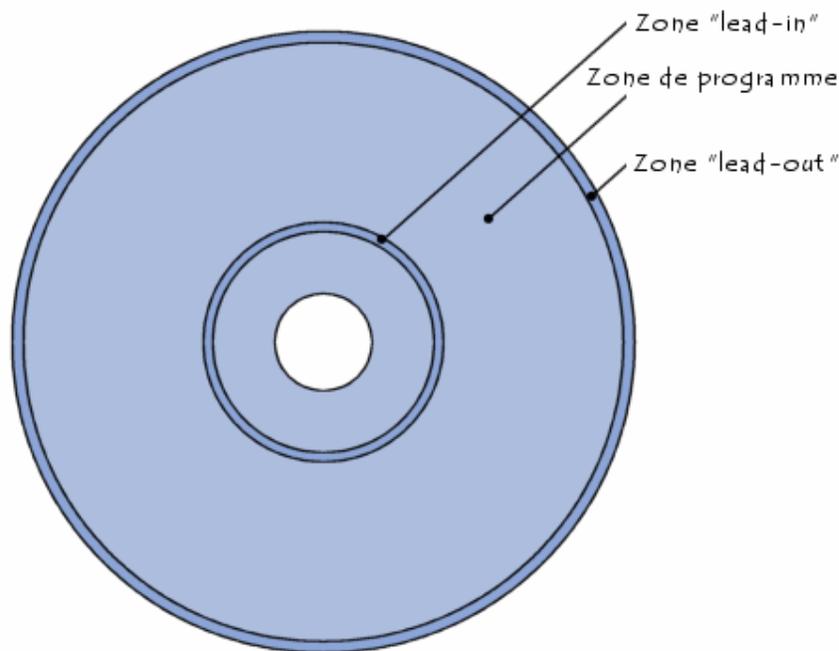
## DVD DL

Le terme «**DVD DL**» (*DVD Dual Layer*) désigne les DVD enregistrables double couche. Ces supports, proposant une plus grande capacité de stockage que les DVD simple couche, utilisent une technologie proche des DVD-9 (DVD pressés double couche).

## Structure logique

Un DVD est essentiellement constitué de trois zones représentant la «*zone d'information*» (*information area*) :

- La zone **Lead-in Area** (parfois notée *LIA*) contenant uniquement des informations décrivant le contenu du support (ces informations sont stockées dans la **TOC**, *Table of Contents*, traduisez *table des contenus*). La zone *Lead-in* sert au lecteur à suivre les creux en spirale afin de se synchroniser avec les données présentes dans la *zone programme*.
- La zone **Programme** (*Program Area*) est la zone contenant les données.
- La zone **Lead-Out** (parfois notée *LOA*), contenant des données nulles (du silence pour un DVD audio), marque la fin du CD.



Un DVD enregistrable contient, en plus des trois zones décrites ci-dessus, une zone appelée *PCA* (*Power Calibration Area*) et une zone *RMA* (*Recording Management Area*) situées avant la zone *Lead-In*.

La *PCA* peut être vue comme une zone de test pour le laser afin de lui permettre d'adapter sa puissance au type de support. C'est grâce à cette zone qu'est possible la commercialisation de supports vierges utilisant des couches réfléchissantes et des colorants organiques différents. A chaque calibration, le graveur note qu'il a effectué un essai. Un maximum de 99 essais par media est autorisé.

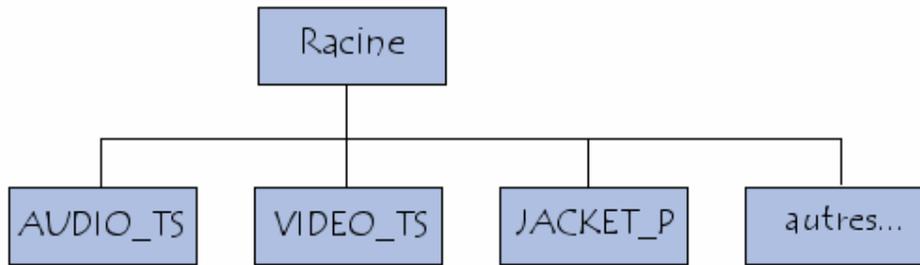
### Système de fichiers et répertoires

Les DVD utilisent le système de fichiers **UDF** (*Universal Disk Format*). Afin de maintenir une certaine compatibilité avec d'anciens systèmes d'exploitation, un système de fichiers hybride, appelé «UDF Bridge», supportant l'UDF et le système de fichiers ISO 9660 utilisé par les CD-ROM, a été mis au point. Il est toutefois important de noter que les lecteurs de DVD Vidéo et de DVD Audio ne supportent que le système UDF.

### Structure d'un DVD Vidéo

Un DVD vidéo peut contenir des données destinées à des platines de salon ainsi que des données additionnelles pouvant être lues sur un ordinateur.

Un DVD-Video possède une organisation hiérarchique de ses répertoires permettant de contenir les données vidéo et audio. Il repose habituellement sur la structure suivante :



Le répertoire principal, nommé *VIDEO\_TS* (pour *Video Title Sets*), a pour vocation de contenir les fichiers du DVD Vidéo. Le répertoire *AUDIO\_TS* concerne les DVD-Audio mais sa présence est parfois demandée par certains lecteurs DVD de salon. *JACKET\_P* contient les images des jaquettes du DVD. Il est enfin possible d'y adjoindre d'autres répertoires, pouvant être lus sur un ordinateur.

Un DVD vidéo est composé d'un certain nombre d'éléments, présents dans le répertoire *VIDEO\_TS* :

- un **gestionnaire vidéo (VMG, pour *Video Manager*)**. Le VMG contient généralement la ou les vidéos d'introduction, ainsi que le menu donnant accès aux autres titres vidéo (y compris les sous-menus).
- un ou plusieurs **ensembles de titres vidéo (VTS, pour *video titles sets*)**, contenant les titres vidéo.

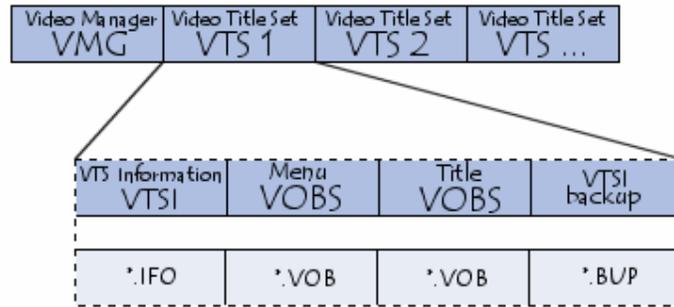
Les «titres vidéo» correspondent à des films, des vidéos ou des albums. Un titre est composé d'un «ensemble d'objets vidéo» (**VOBS, *Video Object Block Sets***), chacun composé :

- d'un «fichier de contrôle» (appelé **VTSI, pour *Video Title Set Information***), et contenant les données de navigation.
- d'un ou plusieurs objets vidéo (**VOB, *Video Object Block***). L'objet vidéo (VOB) est l'élément de base du DVD. Il contient des données vidéo, audio et des images multiplexées, au format MPEG2. Ainsi, un fichier *.VOB* peut être lu par un lecteur vidéo logiciel en changeant son extension en «*MPG*». Les spécifications du DVD imposent que chaque fichier VOB ne dépasse pas un giga-octet. Chaque VOB est lui-même composé de «cellules» (*Cells*), représentant les différents clips vidéo ou audio composant le VOB : par exemple des chapitres vidéo ou les chansons d'un album.
- d'une copie du *VTSI* (**VTSI Backup**).

Un DVD peut contenir jusqu'à 99 titres (*VTS*), chacun subdivisé jusqu'à 10 chapitres.

Ainsi, le répertoire *VIDEO\_TS* contient habituellement trois types de fichiers possédant les extensions suivantes :

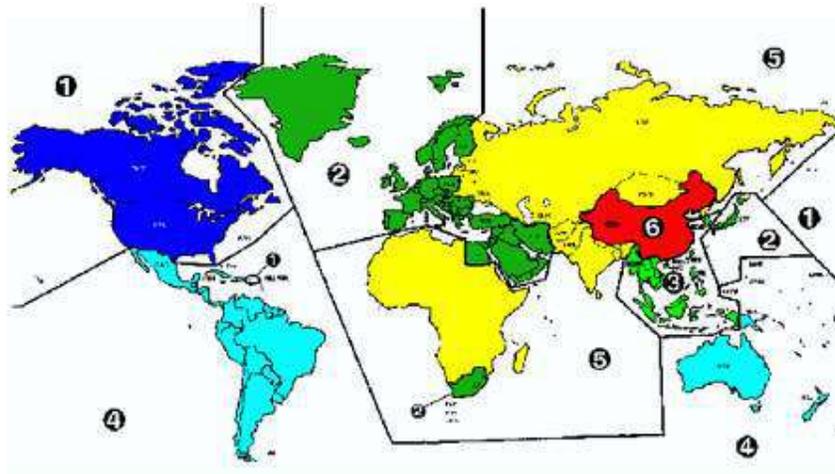
- **IFO** contenant les informations de navigation (il correspond au Video Manager).
- **VOB** (*Video Object Block*) contenant les flux vidéo, les différents canaux audio ainsi que les sous-titres d'un titre vidéo.
- **BUP** (*BUP* signifiant *Backup*), contenant une sauvegarde des fichiers IFO, au cas où ils seraient illisibles.



Le fichier particulier nommé *VIDEO\_TS.IFO* (*IFO* signifiant *information*) contient les informations nécessaires au lecteur pour l'affichage du menu principal. Il est accompagné du fichier *VIDEO\_TS.VOB*, contenant le clip d'animation d'introduction, ainsi que d'un fichier de sauvegarde (nommé *VIDEO\_TS.BUP*).

### Zones

Les DVD Vidéo sont conçus pour ne pouvoir être consultés que dans certaines régions du monde : il s'agit du découpage en zone (prévu initialement pour limiter la diffusion des copies illicites). Il est ainsi théoriquement impossible de lire un DVD d'une zone en étant situé dans une autre. Néanmoins, la quasi-totalité des lecteurs de DVD pour ordinateurs et une grande partie des lecteurs de salon peuvent être «dézonnés» grâce à des utilitaires.



Lecteur de bande

### Introduction.

La sauvegarde sur bande utilise une bande magnétique similaire à une cassette audio. Le prix de ces cassettes magnétiques est généralement faible. Les programmes associés permettent de choisir les dossiers / fichiers à sauvegarder (restaurer) et de les sauvegarder automatiquement à certaines dates / heures. Si ces systèmes ont eut leurs heures de gloire, les capacités des disques durs actuels couplés avec la vitesse de transfert de ces sauvegardes (sauf systèmes DAT professionnels que nous verrons en deuxième année) limitent leur utilisation.

Ces systèmes de bande permettent les sauvegardes complètes, y compris les bases de registres des différentes versions de Windows 95-NT. On distingue 3 types de sauvegarde. Une sauvegarde intelligente est un mélange de ces 3 types

**Sauvegarde complète:** sauve l'ensemble de disque dur. Cette sauvegarde est très sûre, mais est longue.

**Sauvegarde incrémentale:** ne sauvegarde que les fichiers qui ont été modifiés depuis la dernière sauvegarde. Une restauration nécessite donc de récupérer d'abord une sauvegarde complète et ensuite de reprendre les restaurations incrémentales.

**Sauvegarde différentielle:** copie tous les fichiers depuis le dernier backup complet ou incrémental. En effet, ce mode de sauvegarde ne modifie pas le bit d'archive des fichiers.

### Les types.

On trouve sur le marché trois gros types différents de sauvegardes sur bande: QIC/Travan, DAT (4 et 8 mm) et DLT.

Le système QIC se caractérise par les plus faibles capacités (moins de 4 GB) et des vitesses de sauvegarde relativement lentes. Le QIC souffre de plus de la multitude de format incompatibles entre-eux. QIC est réservé aux applications personnelles.

La technologie DAT est la plus répandue. Elle se scinde en 2 catégories: 4 et 8 mm, en référence à la largeur de la bande. Les 4 mm ont une capacité de 2 voire 4 GB. Les 8 mm ont une capacité supérieure à 5 GB. Les performances des DAT sont correctes, mais les entretiens sont coûteux: les sauvegardes DAT utilisent une tête mobile qui usent les bandes.

La technologie DLT se distingue des modes QIC et DAT par des capacités élevées (plus de 10 GB en mode natif) et des vitesses d'exécution supérieures. Plus chères, elles sont également plus fiables. Leur tête de lecture est fixe et les cartouches plus résistantes. Les têtes peuvent lire et écrire des données simultanément sur plusieurs canaux (bits en parallèle). En utilisant une compression de 2:1, certains DLT assurent un débit théorique de 3 MB / s. Par comparaison, les DAT et QIC approchent à peine les 1,5 MB / secondes. L'utilisation de têtes fixes permet également une plus longue longévité de l'appareil (10.000 heures en moyennes pour 2.000 en technologie DAT). Revers de la médaille, ces appareils sont chers à l'achat et sont donc réservés à des applications lourdes de type serveurs. Les technologies DAT, DLT et autres seront vues en détail dans le cours de deuxième: sauvegarde et backup réseau.

La compatibilité des média ne peut se faire que dans une même famille. Méfiez-vous particulièrement des QIC, où les différents constructeurs proposaient des bandes de capacité similaires, sinon égales, mais incompatibles entre-elles (ex: Hewlett Packard et IOMEGA). De plus, le logiciel entre en compte pour l'écriture des données.

### Les disquettes de grande capacité.

#### Introduction

Avec sa faible capacité de stockage, la disquette est dépassée. Quelques firmes ont tenté de développer un standard. Parmi eux: IOMEGA et son ZIP accompagné de la firme SYQUEST

sont les plus connus, bien que fin 1999, SYQUEST soit pratiquement en faillite. En 2003, Syquest ne vend plus que les bandes et supports et ne développe plus de nouveaux produits.

Tous ces lecteurs utilisent un procédé magnéto-optique. Ceci permet une fiabilité élevée du média.

### **IOMEGA Zip 100 MB, 250 MB et 750 MB.**

D'abord sorti en version SCSI (interne et externe), le zip 100 MB s'est ensuite connecté sur le contrôleur lecteur de disquette dans ses versions internes. Actuellement, les zip 100 MB interne sont connectés sur le port IDE.

IOMEGA a sorti une version Zip 100 + externe, connectable sur un port SCSI ou parallèle. Ce modèle a posé plusieurs problèmes d'incompatibilité avec une imprimante connectée à la suite.

Le lecteur IOMEGA était pratiquement un standard en Mac, il est souvent utilisé par les imprimeurs. De ce fait, l'utilisation d'un PC pour créer des fichiers d'impression vers les imprimeurs est pratiquement l'idéal.

En 1998, IOMEGA a sorti une version 250 MB de son lecteur ZIP. Si la version parallèle a existée, elle est remplacée par une version USB 1.0. En interne, la connexion se fait en IDE ou en SCSI.



Dernière nouveauté en 2002, la version 750 MB. Cette version peut être interne (IDE) ou externe (USB2.0, compatible 1.1 ou firewall sortie en janvier 2003). Les lecteurs Zip 750 MB sont totalement compatibles avec les versions 250 MB (lecture et écriture). Par contre, s'ils peuvent lire les disquettes Zip 100 MB, ils ne peuvent les utiliser en écriture.

### **LS-120 MB de Syquest.**

Le LS-120 de Syquest utilisait des média de 120 MB. Ses mauvaises performances n'ont pas décidé les constructeurs de PC de l'installer directement, malgré la possibilité de lire des disquettes 1,44 MB standard. Il est néanmoins souvent repris comme possibilité de démarrage dans les BIOS

### **Mémoires sur ports USB (stick USB)**

Une autre méthode d'utiliser les ports USB pour la sauvegarde consiste à utiliser des mémoires flash appelées clés USB directement sur le port USB 1.1 ou 2.0. Les capacités vont jusqu'à 8 GB. La vitesse est assez lente en USB 1.1. Ces périphériques sont directement reconnus par Windows 2000, XP et suivants (pas Win98 et forcément pas Win95 et NT4, incompatibles USB).

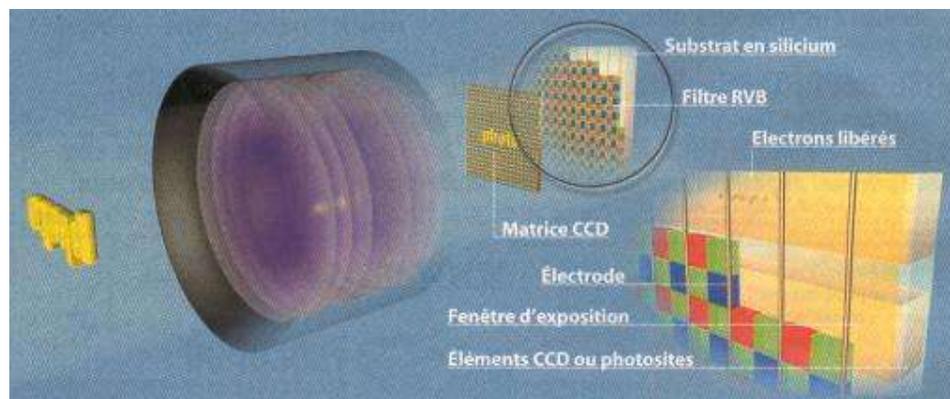
### **Disque dur externe.**

Avec l'arrivée des liaisons firewire et plus encore avec l'USB 2.0 fin 2002, une majorité des constructeurs sortent des disques externes de grande capacité (jusqu'à 2 TB). Le prix est néanmoins quasi le double par rapport à un disque dur interne équivalent plus rapide. Vous pouvez également utiliser des adaptateurs, sortes de boîtiers connectés en USB ou firewire où l'on insère un disque dur standard. Des modèles existent pour disques durs IDE 3"5 et 2"5 (spécifiques notebook). Ces appareils ont remplacé les Jazz d'Omega et les Sparq de Syquets, obsolètes.

### Acquisition image: appareil photo numérique et scanner.

#### Appareil photo-numérique

Le signal lumineux est transmis à une cellule photo-sensible appelée CCD (Charged Coupled Device) qui transforme l'énergie (lumière) en une série d'impulsions électriques. La charge électrique est directement proportionnelle à la lumière captée. Une charge nulle produit du noir, une charge maximum produisant du blanc. A ce stade, le CCD produit toutes les nuances de gris (le nombre de niveaux de gris est néanmoins réduit par les fabricants). En décomposant le contenu en rouge, jaune et bleu par des filtres de couleurs, on obtient par exemple: 256 nuances de bleus \* 256 nuances de vert \* 256 nuances de rouge = 16.777.216 combinaisons de couleurs possibles.



La technique la plus connue détermine la couleur par interpolation des pixels avoisinants: si les pixels rouges, bleus et jaunes donnent la même valeur, ils deviennent gris. Si le pixel rouge est maximum et les 2 autres nulles, la couleur est rouge, et ainsi de suite. Malheureusement, cette solution a quelques limites pour des couleurs spéciales.

Les appareils photo-numériques sont appelés MEGAPIXEL avec plus d'un million (Mega) de pixels dans le CCD, par opposition avec les appareils dits soft-display. En quoi consiste le MEGA-PIXEL? Pour obtenir la résolution réelle de l'image, il faut prendre la résolution maximum donnée par le fabricant (par exemple 1600 \* 1200), la diviser par 4 (un filtre rouge, 2 filtres verts et un filtre bleu), ce qui donne le megapixel réel. Dans notre exemple, nous obtenons 0,48 megapixels au lieu de 1,9 megapixels. La taille de chaque pixel est de 0,15 microns, en multipliant par 4, on obtient la taille réelle soit 60 Microns par bit (point). En général, un tel pixel est codé sur 12 bits, ce qui pour une image donne rapidement des fichiers de 12 MB. Une dia de 35 mm contient 100 fois plus d'informations. Ceci donne donc une image de 400 MB. Au point de vue physique, l'appareil est donc largement au-dessus des appareils photo-numériques.

Les appareils actuels implantent différentes techniques supplémentaires pour la prise de photo comme par exemple:

- **Zoom optique, logiciel:** Le zoom optique est le zoom réel (les plus performants acceptent jusque 18 X). Le zoom logiciel utilise une extrapolation des points et perd de la qualité.
- **Afficheur LCD** pour voir directement les photos ou même pour configurer l'appareil.
- **Mode Macro:** cette fonction permet de prendre des photos d'objets petits et proches.
- **Mode panoramique:** permet de prendre plusieurs photos d'un panorama et de les coller entre-elles.
- **Mode caméra:** permet de filmer de petits films avec votre appareil photo.
- **Anti-mouvement:** correction automatique lors du déplacement de l'appareil en mode prise de vue.
- **Zoom supplémentaires (téléobjectifs).** permet d'augmenter le niveau du zoom interne en ajoutant un zoom supplémentaire comme dans les appareils photos standards.
- mode macrophotographie, pour photographier de petits objets, insectes, ... de loin.

### Types de mémoires Flash

Si Sony a développé quelques modèles d'appareils numériques conservant les photos sur disquette directement insérées dans l'appareil, cette technique a été abandonnée, trop lente, mais surtout de capacité trop faible. La sauvegarde des photos se fait sur des mémoires flash, permettant la lecture et l'écriture comme une RAM, mais conservant les informations en l'absence d'alimentation (comme une ROM). Par contre, elles ne permettent pas un nombre d'écritures infinies. Ces mémoires sont réinscriptibles, mais conservent le contenu lorsqu'elles ne sont pas alimentées. Si on trouve des appareils contenant des mémoires de 8 MB, les nouveaux appareils montent à 8 GB (on peut en rajouter suivant les types ci-dessous). Une mémoire de 4 GB permet par exemple de sauvegarder 1240 images en 1880 X 2130 pixels au format JPG (appareil 4 Mega pixels).

### Memory Stick

Développée en 2000 par Sony et Scandisk, ce type de mémoire série n'est quasiment fabriquée que par Sony. Les dimensions standards sont de 50,0 mm (largeur) x 21,5 (Hauteur) x 2.8 mm (épaisseur).



Le boîtier inclus la ram et le contrôleur. La capacité actuelle maximum atteint les 4 GB en Pro Duo.

On distingue plusieurs types de Memory stick:

1. Standard: vitesse de transfert de 14.4 Mb/s (19.6 Mb/s maximum)
2. Pro: incompatible avec la version standard, elle offre une vitesse de transfert supérieure: (15 mb/s en standard mais jusque 160 Mbs maximum.

3. Magic stick Duo: dérivé de la version standard, elle augmente la capacité du media (en restant compatible). Insérée dans un lecteur standard, elle n'accède qu'à la moitié de la capacité. Elle est basée sur la version Pro.
4. Avec Magic Gates, Sony introduit un codage (protection des données).

### Secure Digital (SD - SDHC), la plus courante.



Développée début 2000 par Matsushita Electronic, SanDisk et Toshiba, la **SD Card** est également utilisée dans les appareils photo-numériques, Caméoscopes digitaux, PDA, ... Les dimensions sont normalisées: 24 mm de haut, 32 mm de large et 2,1 mm d'épaisseur. Le transfert des informations (mode série) utilise un connecteur de 9 broches



En version standard, la vitesse de transfert est de 2 MB/s. Une version supérieure pour les mémoires de plus de 2 GB, la Secure Digital High Capacity (**SDHC**), est produite depuis 2006. Avec ces cartes mémoires, nouvelles normes gèrent le taux de transfert avec la classe 2 (2 Mb/s), 4 (4 Mb/s) et 6 (6 Mb/s).

Ces mémoires incluent un mécanisme de protection contre le piratage (CPRM - Content Protection Media) pour la musique et les films vidéos lors de la copie sur ce média: 3 copies maximum

Un format plus petit (11 mm x 15 mm x 1 mm) est parfois utilisé sous le nom de MicroSD pour téléphonie mobile. Un adaptateur permet de les insérer dans les lecteurs SD standards.

### Compact Flash



La plus ancienne est la Compact Flash (CF en abrégé), elle existe depuis 1994. Ici aussi, le contrôleur est inclus dans le boîtier. Deux types sont développées, les types I (épaisseur de 3,3 mm) et les types 2 (épaisseurs de 5 mm). Les autres dimensions sont de 42,8 mm de largeur et

36,4 en hauteur. Les capacités standards vont jusque 8 GB, même si la norme accepte jusque 137 GB. Ce modèle disparaît.

Développées au préalable suivant la norme PCMCIA, elle utilise la même technologie de transfert (en mode parallèle) si ce n'est que le nombre de broches est de 50 au lieu de 68. Le connecteur PCMCIA accepte d'ailleurs directement les Compact Flash. La tension d'alimentation est de 3,3 Volts ou 5 volts. Le choix se fait automatiquement à la connexion.

### **XD Picture Card**

Commercialisées depuis 2002, les cartes Xd Picture Card sont principalement utilisées par les appareils numériques Olympus et Fujifilm mais dans certains dictaphones et lecteurs MP3. par rapport aux autres types de mémoires, le contrôleur n'est pas intégré, il est repris directement dans l'équipement (ce qui son prix plus chère). Prévue au départ pour remplacer les SmartMedia, les deux sont convertibles via un adaptateur spécifique.



Sorti en 2005, le **type M** utilise des mémoires MLC (Multi Layer Cell) pour augmenter la capacité jusque 8 MB, malheureusement au détriment de la vitesse de lecture / écriture par rapport à la première version.

Sorties à partir de 2006, le **type H** utilise une mémoire SLC (Single Layer Cell). Elle est la plus rapide (5 MB/s en lecture pour 4 en écriture) mais sa capacité est plus limitée que le type M.

Tous les appareils photo-numériques acceptent le type standard. Par contre, des incompatibilités existent entre les type H et M au niveaux des lecteurs.

La taille maximum actuelle est de 2 GB, la norme prévoit jusque 8 MB. Les dimensions sont de 25 mm de large, 20 mm de haut pour une épaisseur de 1,7 mm. Elles pèsent à peine 2 grammes. Le connecteur utilise 18 pins avec une tension d'alimentation de 3,3 Volts.

### **Les scanners**

Pour l'acquisition de photos, on retrouve 3 techniques distinctes: scanner à main, à défilement (identique aux fax par exemple) et à plat. Les deux premiers ont été largement supplantés par les scanners à plat. Ces scanners peuvent faire l'acquisition de différentes tailles : A4, A4+ et pour certains modèles spécifiques, jusqu'à A3 et même A2. La connexion utilisée par les scanners sont le port SCSI externe, parallèle et USB. Actuellement, ils sont quasiment tous connectés en USB 2.0 (15 MB/s), même si l'interface USB 1.1 est compatible. C'est le seul port qui accepte de se connecter à chaud (sans redémarrer l'ordinateur) si le logiciel est préalablement installé.

La majorité des fabricants les ont implantés dans les AIO qui permettent d'imprimer, scanner et copier, voire de faxer.

### **Le pilote.**

Les scanners sont pratiquement tous pilotés par un programme compatible **TWAIN**. Ceci permet à n'importe quel logiciel de traitement d'image d'utiliser le scanner via cette interface logicielle standard. Le pilote est spécifique à chaque scanner.

Actuellement, les "drivers" sont capables de bien plus que la simple acquisition d'images: détection automatique de la nature des originaux (couleur, Noir et Blanc, textes, images, dessins, ...), correction automatique des images (amélioration des images en agissant sur les tons, couleurs et netteté), gestion automatique des couleurs (restitution fidèle des couleurs sur l'imprimante et l'écran). Certains pilotes découpent le document en partie suivant le type (image, textes, ...).

### **La résolution et la palette.**

La résolution d'un scanner correspond au nombre de points que le scanner peut acquérir dans une surface donnée. Elle est exprimée en dpi (dot per inch) et peut être différente en largeur et en longueur. Cette résolution est souvent définie en résolution optique et résolution logicielle. La résolution logicielle est une simple extrapolation software des points. Ceci ne donne pas plus de détails, mais affine les différences de couleurs en faisant gonfler la taille du fichier. Pour des travaux courants, une résolution de 200 dpi (voire 150) est largement suffisante. Attention qu'un doublement de la taille de l'image par rapport à l'original réduit la résolution par 2. En cas d'agrandissement, mieux vaut doubler la résolution de départ. Pour rappel, la résolution d'un écran est limitée à 72 dpi. Les images "Internet" (publication sur site, envoi par mail pour catalogue, ...) doivent donc correspondre à cette résolution ou même moins.

Chaque scanner se distingue également par sa palette de couleurs exprimée en bit. Les scanners courants actuels ont une palette de 30 ou 36 bits. 30 bits permet de reconnaître plus d'un milliard de couleur, 36 bit permet de reconnaître un peu plus de 68 milliards de couleurs. Les scanners professionnels peuvent monter à 42 bits ou plus. La palette de couleur en niveaux de gris est généralement le 1/3 de celle couleur.

Ces 2 considérations sont trop simplistes pour différencier un modèle de scanner à l'autre. En effet, à résolution équivalente, la qualité de l'optique varie. Comme les scannages couleurs font passer l'image par 3 filtres de couleurs, la qualité de ces filtres influence fortement la qualité de l'image finale. De plus, la netteté des contours de chaque point influence également la qualité de l'image. Ceci explique qu'à caractéristiques équivalentes, les scanners se distinguent fortement des scanners de bas de gamme.

**Même si cela peut sembler bizarre**, l'impression en haute qualité avec une résolution de scannage de 200 dpi donne un meilleur résultat à l'impression qu'un scannage de 600 dpi. Pour les images Internet, la résolution habituelle de l'écran est de 72 dpi.

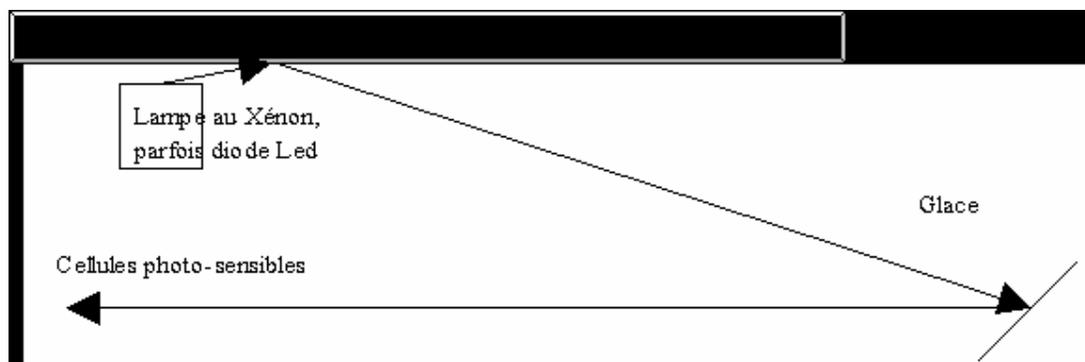
### **Détramage.**

A cause de leur mode d'impression, certains supports tels que les magazines ou certains journaux donnent des effets de quadrillage à la numérisation. Le tramage consiste à convertir les images en une série de lignes ou de points. Pour atténuer ou faire disparaître cet effet indésirable de moiré entre les points et les pixels qui rend l'image numérisée quasi inutilisable, les pilotes des

scanners disposent quasiment tous d'une option de déramage. Comme il s'agit d'un traitement logiciel interne au scanner, l'efficacité et la durée de ce traitement varient en fonction des modèles.

### Le fonctionnement d'un scanner à plat.

Tous les scanners à plat suivent le même principe de fonctionnement: une source lumineuse se déplace à la surface d'un document, la lumière produite par la source est réfléchiée et rencontre un capteur photosensible électrique. Ce capteur convertit la lumière en signal électrique, en fait une représentation analogique de l'image qui doit être convertie en signal digital à l'aide d'un convertisseur. Avant d'être transférée à l'ordinateur, l'image passe par un microcontrôleur pour mise en forme. Dans le cas de scanners couleurs, chaque partie d'acquisition est reproduite en 3 exemplaires, un par couleur.



Sur le marché, on trouve 2 types de capteurs, le plus courant est le **CCD** (Charge Coupled Device ou capteur à transfert de charge). Utilisé par les appareils photo informatiques, il utilise une technologie à semi-conducteur formée de nombreux éléments, tous capables de recueillir, stocker et transmettre les charge électriques reçues. Chaque élément CCD correspond à un élément de l'image, soit un pixel. Lorsqu'ils se retrouvent exposés à la lumière de la source (généralement une lampe Xénon), ces éléments sont chargés par des photons. Ceci crée une charge électrique mesurable qui est conservée par les CCD jusqu'à ce qu'elle soit transmise à un amplificateur de sortie où elle est convertie en signal électrique. Le capteur CCD est fixe et le cheminement de la lumière peut entraîner une déperdition du signal lumineux, une image légèrement dégradée par exemple.

Le second type de capteur, le **CIS** (Contact Image Sensor ou capteur d'images par contacts) est le plus courant actuellement, c'est le moins chère. Basé sur la même technologie que le CCD, la lumière ne chemine pas à l'intérieur du scanner car le capteur est directement en contact au

travers de la vitre avec le support à numériser. Dans ce système, une lentille cylindrique fait converger vers le capteur la lumière émise par des diodes électroluminescentes rouges, vertes et bleues. Ces diodes, la lentille et le capteur font partie du même dispositif. Ceci permet un spectre de couleurs plus propre et une plus grande fidélité chromatique, mais la netteté des images peut être moins bonne que celles des scanners à CCD. Les scanners à CCD sont généralement plus rapides que ceux à CIS. En effet, une vitesse de numérisation trop rapide en CIS rendrait l'image floue.

Comme les scanners à CIS n'incluent pas de miroirs, ni d'objectifs, ils sont facilement identifiables. De plus, ils sont nettement moins encombrants et ... nettement moins chères. Cette technologie est utilisée dans les appareils intégrés (Imprimante, scanner, photocopieuse).

Le signal capté par les capteurs (CCD ou CIS) est transformé en signal électrique digital. Comme les courbes de réponses des CCD ne sont pas parfaitement linéaire, un microcontrôleur, processeur spécialisé, assure la correction suivant les différentes couleurs. Il surveille également tout le processus et assure la conversion de l'image brute en un format compréhensible par le logiciel de retouche d'images. Outre le format de l'image, le microcontrôleur doit également s'assurer que le logiciel comprend les images. En effet, la majorité des logiciels de traitement d'images les traite en 8 bits par couleurs, alors que les scanners utilisent 10, 12 ou même 14 bits.

## **L'OCR.**

La majorité des scanners sont associés à un logiciel de reconnaissance de caractères. Ces logiciels permettent de traiter des textes numérisés comme des textes utilisables par Word, WordPro, ... Le choix de la langue de départ est primordial puisqu'en cas de difficulté, les logiciels d'OCR ont tendance à vérifier le mot le plus proche. De toute façon, une reconnaissance à 100 % n'existe pas, même avec des versions complètes du logiciel (les logiciels fournis avec le scanner sont généralement en version lite). De plus, la taille de caractère et la police influencent fortement la reconnaissance.