

# Architecture des circuits Architecture des ordinateurs

## Protopoly

**Florent de Dinechin**

*avec des figures de*

R. Bergasse, N. Bonifas, N. Brunie, P. Boutilier,  
A. Derouet-Jourdan, J. Detrey, A. Friggeri, M. Feurgard, B. Grenet, F. Givors,  
T. Jolivet, C. Keller, E. Lassalle, P.E. Meunier,  
P. Robert, G. Salagnac, O. Schwander, T. Risset, P. Vannier

version du 28 novembre 2025



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	Des livres sérieux, parce que ce poly, bon . . . . .	9
1.2	Historique du calcul mécanique . . . . .	10
1.3	Objectifs du cours . . . . .	12
1.4	La conception d'ASR est hiérarchique . . . . .	14
1.5	Quelques ordres de grandeur . . . . .	14
1.5.1	L'univers est notre terrain de jeu . . . . .	14
1.5.2	La technologie en 2018 . . . . .	15
1.5.3	Calcul contre stockage et déplacement de données . . . . .	16
1.5.4	Ya pas que les PC dans la vie . . . . .	16
<b>I</b>	<b>L'information et comment on la traite</b>	<b>19</b>
<b>2</b>	<b>Coder l'information</b>	<b>21</b>
2.1	Généralités . . . . .	21
2.1.1	Information et medium . . . . .	21
2.1.2	Information analogique . . . . .	21
2.1.3	Information numérique . . . . .	21
2.1.4	Coder le temps . . . . .	22
2.2	Coder des nombres entiers . . . . .	22
2.2.1	Numération de position pour les entiers naturels . . . . .	22
2.2.2	Ordres de grandeurs en puissances de 2 . . . . .	23
2.2.3	Parenthèse pratique : superposition des codages . . . . .	23
2.2.4	Codage des entiers relatifs . . . . .	23
2.2.5	Et pour aller plus loin . . . . .	25
2.3	Coder du texte . . . . .	25
2.4	Coder les images et les sons . . . . .	26
2.5	Codes détecteurs d'erreur . . . . .	26
2.6	Codes correcteurs d'erreur . . . . .	26
2.7	Compression d'information . . . . .	27
2.8	Les QR-codes . . . . .	27
<b>3</b>	<b>Transformer l'information : circuits combinatoires</b>	<b>29</b>
3.1	Algèbre booléenne . . . . .	29
3.1.1	Définitions et notations . . . . .	29
3.1.2	Expression booléenne . . . . .	30
3.1.3	Dualité . . . . .	30
3.1.4	Quelques propriétés . . . . .	30
3.1.5	Universalité . . . . .	30
3.1.6	Fonctions booléennes . . . . .	30
3.2	Circuits logiques et circuits combinatoires . . . . .	31
3.2.1	Signaux logique . . . . .	31
3.2.2	Circuits logiques . . . . .	31
3.2.3	Portes de base . . . . .	31
3.2.4	Circuits combinatoires . . . . .	32
3.2.5	Construction bottom-up de circuits combinatoires . . . . .	32
3.2.6	Construction top-down de circuits combinatoires . . . . .	33
3.2.7	Métriques d'un circuit combinatoire . . . . .	33

3.3	Quelques constructions algorithmiques de circuits combinatoires . . . . .	34
3.3.1	Les portes de base à deux entrées . . . . .	34
3.3.2	Le multiplexeur, abstraction du si-alors-sinon . . . . .	34
3.3.3	Le multiplexeur d'ordre $k$ . . . . .	35
3.3.4	Pour le calcul sur les entiers . . . . .	36
3.3.5	Conclusion . . . . .	38
3.4	D'une fonction booléenne à un circuit combinatoire . . . . .	38
3.4.1	Par les formes canoniques . . . . .	38
3.4.2	Par les arbres de décision binaire . . . . .	39
3.5	Application : construction des circuits arithmétiques de base . . . . .	40
3.5.1	Addition/soustraction binaire . . . . .	40
3.5.2	Multiplication binaire . . . . .	40
3.5.3	Division binaire . . . . .	40
3.6	Conclusion . . . . .	40
3.7	Annexe technologique contingente : les circuits CMOS . . . . .	40
3.7.1	Transistors et processus de fabrication . . . . .	40
3.7.2	Portes de base . . . . .	41
3.7.3	Vitesse, surface et consommation . . . . .	42
<b>4</b>	<b>Memoriser l'information</b>	<b>43</b>
4.1	Mémoriser un bit . . . . .	43
4.1.1	Le registre 1-bit, ou flip-flop . . . . .	44
4.1.2	Construction : du verrou (latch) au registre (Flip-Flop) . . . . .	44
4.1.3	Variantes : registre avec reset et / ou write enable . . . . .	45
4.1.4	Le registre n-bit . . . . .	45
4.2	Mémoire adressable . . . . .	46
4.2.1	Construction d'une mémoire adressable . . . . .	46
4.2.2	Construction des mémoires utilisant des porte de transmission . . . . .	47
4.2.3	Et pour des puces plus carrées . . . . .	49
4.2.4	Mémoire dynamique . . . . .	49
4.2.5	Mémoire volatile ou non volatile . . . . .	49
4.3	Mémoires à accès séquentiel : piles et files . . . . .	50
4.3.1	File ou FIFO ( <i>first in, first out</i> ) . . . . .	50
4.3.2	Pile ou LIFO ( <i>last in, first out</i> ) . . . . .	50
4.3.3	Construction de piles et files . . . . .	50
4.4	Mémoires adressables par le contenu . . . . .	51
4.5	Disques (*) . . . . .	51
4.6	Une loi fondamentale de conservation des emmerdements . . . . .	51
<b>5</b>	<b>Circuits séquentiels synchrones</b>	<b>53</b>
5.1	Quelques exemples de circuits séquentiels . . . . .	53
5.2	Restriction aux circuits séquentiels synchrones . . . . .	53
5.3	Correction et performance . . . . .	54
<b>6</b>	<b>Automates</b>	<b>57</b>
6.1	Un exemple . . . . .	57
6.2	Définition formelle d'un système de transitions . . . . .	58
6.2.1	États, transitions . . . . .	58
6.2.2	Définition en extension des fonctions de transition et de sortie . . . . .	59
6.2.3	Correction et complétude d'un système de transitions . . . . .	59
6.3	Synthèse d'un automate synchrone . . . . .	60
6.3.1	L'approximation temporelle réalisée par l'automate synchrone . . . . .	61
6.3.2	Optimisation d'un automate synchrone . . . . .	61
6.4	Comprendre les circuits séquentiels comme des automates . . . . .	61
6.4.1	La norme JTAG . . . . .	62
6.4.2	Équivalence de circuits . . . . .	63
6.5	Conclusion : l'ingénierie des automates . . . . .	63

<b>7 Transmettre</b>	<b>65</b>
7.1 Medium . . . . .	65
7.2 Liaison point a point . . . . .	65
7.2.1 Série ou parallèle . . . . .	65
7.2.2 Protocoles . . . . .	65
7.3 Bus trois états . . . . .	67
7.4 Réseaux en graphes (*) . . . . .	68
7.4.1 Les topologies et leurs métriques . . . . .	68
7.4.2 Routage . . . . .	69
7.4.3 Types de communication : point à point, diffusion, multicast . . . . .	71
7.5 Exemples de topologies de réseau (*) . . . . .	71
7.5.1 Le téléphone à Papa . . . . .	71
7.5.2 L'internet . . . . .	71
7.5.3 FPGAs . . . . .	71
7.5.4 Le bus hypertransport . . . . .	71
7.5.5 Machines parallèles . . . . .	71
<b>II Machines universelles</b>	<b>73</b>
<b>8 Jeux d'instruction</b>	<b>75</b>
8.1 Rappels . . . . .	75
8.2 Vocabulaire . . . . .	75
8.3 Travaux pratiques . . . . .	76
8.3.1 Le jeu d'instruction de votre PC . . . . .	76
8.3.2 Le jeu d'instruction de votre téléphone portable . . . . .	77
8.4 Instruction set architecture . . . . .	77
8.5 Que définit l'ISA . . . . .	78
8.5.1 Types de données natifs . . . . .	78
8.5.2 Instructions de calcul . . . . .	78
8.5.3 Instructions d'accès mémoire . . . . .	79
8.5.4 Instructions de contrôle de flot . . . . .	81
8.5.5 Les appels système, les interruptions, et l'atomicité . . . . .	82
8.5.6 Quelques ISA . . . . .	82
8.6 Codage des instructions . . . . .	83
8.7 Adéquation ISA-architecture physique . . . . .	84
8.8 Un peu de poésie pour finir . . . . .	85
<b>9 Construction d'un processeur RISC</b>	<b>87</b>
9.1 Un jeu d'instruction RISC pas terrible mais à la mode . . . . .	87
9.2 Plan de masse . . . . .	88
9.3 Construction de l'automate de commande . . . . .	89
<b>10 Entrées/sorties, interruptions, exceptions, et appels systèmes</b>	<b>91</b>
10.1 Interfaces d'entrée/sorties . . . . .	91
10.2 Interruptions . . . . .	92
10.3 Instructions atomiques . . . . .	92
10.4 Modes d'exécution . . . . .	93
10.5 Exceptions . . . . .	93
10.6 Appel système . . . . .	94
10.7 Instructions de synchronisation . . . . .	94
<b>11 Exécution parallèle</b>	<b>95</b>
11.1 Pipeline d'exécution . . . . .	95
11.2 Exploitation du parallélisme d'instruction : exécution superscalaire . . . . .	99
11.2.1 Architecture superscalaire (*) . . . . .	99
11.2.2 VLIW ou superscalaire (*) . . . . .	101
11.3 Deux jeux d'instructions VLIW récents . . . . .	101
11.3.1 IA64 . . . . .	101
11.3.2 Kalray K1 . . . . .	102
11.4 Exploitation du parallélisme de processus . . . . .	102

11.4.1	Échange de messages versus partage de mémoire . . . . .	102
11.4.2	Modèle de programmation : thread versus process . . . . .	102
11.4.3	Grandeur et misère de la mémoire partagée . . . . .	103
11.4.4	Architecture : Multifilature ( <i>multithreading</i> ) . . . . .	103
11.4.5	Architecture : processeurs multicoeurs . . . . .	104
11.5	Mécanismes architecturaux pour l'exclusion mutuelle . . . . .	104
11.6	Conclusion : parallélisme partout . . . . .	104
11.6.1	Taxonomie 1 : la classification de Flynn . . . . .	105
11.6.2	Taxonomie 2 : la classification de Raina . . . . .	105
<b>12</b>	<b>Du langage au processeur (*)</b>	<b>107</b>
12.1	Introduction : langages interprétés, langages compilés, processeurs virtuels . . . . .	107
12.2	L'arrière-cuisine du compilateur . . . . .	107
12.2.1	Variables et expressions . . . . .	107
12.2.2	Désucrage des opérations de contrôle de flot . . . . .	108
12.2.3	Tableaux . . . . .	108
12.2.4	Chaînes de caractères . . . . .	110
12.2.5	Structures . . . . .	111
12.3	Application binary interface . . . . .	111
12.3.1	Procédures et compilation séparée . . . . .	112
12.3.2	Récursivité et pile . . . . .	112
12.3.3	Variables locales et passage de valeurs sur la pile . . . . .	113
12.3.4	Vision globale de la mémoire . . . . .	113
12.3.5	Sauvegarde des registres . . . . .	113
12.3.6	En résumé : l'enregistrement d'activation . . . . .	114
12.3.7	En résumé : code à générer par le compilateur . . . . .	115
12.4	Compilation des langages objets . . . . .	115
<b>13</b>	<b>Hiérarchie mémoire</b>	<b>117</b>
13.1	Mémoire cache . . . . .	117
13.1.1	Principes de localité . . . . .	117
13.1.2	Scratchpad versus cache . . . . .	117
13.1.3	Cache hit et cache miss . . . . .	118
13.1.4	Hiérarchie mémoire . . . . .	118
13.1.5	Construction d'un cache . . . . .	118
13.1.6	Statistiques et optimisation des caches (*) . . . . .	121
13.1.7	Entre le cache et la mémoire physique (*) . . . . .	121
13.1.8	Les caches dans un multicœur à mémoire partagée (*) . . . . .	122
13.2	Mémoire virtuelle . . . . .	122
13.2.1	Vue générale . . . . .	122
13.2.2	Avantages de la mémoire virtuelle . . . . .	123
13.2.3	Aspects architecturaux . . . . .	124
13.2.4	Dans le détail : table des pages (*) . . . . .	124
13.2.5	Cache d'adresses virtuelles ou cache d'adresses physiques ? (*) . . . . .	124
13.3	Une mémoire virtuelle + cache minimale . . . . .	125
13.3.1	Instructions spécifiques à la gestion mémoire . . . . .	125
<b>14</b>	<b>Conclusion : vers le système d'exploitation (*)</b>	<b>127</b>
14.1	Le rôle de l'OS (du point de vue d'un prof d'archi) . . . . .	127
14.2	Multiutilisateur ou multitâche, en tout cas multiprocessus . . . . .	128
14.2.1	Partage du temps . . . . .	128
14.2.2	Partage des entrées/sorties . . . . .	129
14.2.3	Partage des ressources d'exécution . . . . .	129
<b>III</b>	<b>Annexes</b>	<b>131</b>
<b>A</b>	<b>Rudiments de complexité</b>	<b>133</b>
A.1	Les fonctions $2^n$ et $\log_2 n$ . . . . .	133
A.2	Raisonner à la louche . . . . .	134

<b>B VHDL, un langage pour la synthèse de circuits</b>	<b>135</b>
B.1 Flot de synthèse . . . . .	135
B.2 Décrire un circuit par un langage . . . . .	136
B.2.1 Description de circuit $\neq$ programmation . . . . .	136
B.2.2 Mais au fait, quelle est la sémantique d'un circuit numérique ? . . . . .	136
B.2.3 Description comportementale ou description structurelle . . . . .	136
B.2.4 VHDL synthétisable . . . . .	137
B.3 Le langage VHDL par l'exemple . . . . .	137
B.3.1 Entités, architectures et composants . . . . .	137
B.3.2 Variables, paramètres, signaux et ports . . . . .	137
B.3.3 Instances . . . . .	139
B.3.4 Types de base . . . . .	139
B.3.5 Concurrent statement . . . . .	140
B.3.6 Processes . . . . .	140
B.3.7 Génération de circuit . . . . .	143
<b>C Anatomie d'un Unix</b>	<b>145</b>
C.1 Appels systèmes . . . . .	145
C.2 Processus . . . . .	145
C.3 Signaux . . . . .	146
C.4 Fichiers . . . . .	146
C.5 Protection . . . . .	146
<b>D Systèmes de fichiers</b>	<b>149</b>
D.1 L'abstraction du fichier . . . . .	149
D.2 Implémentation d'un système de fichiers . . . . .	150
D.2.1 Allocation contigüe . . . . .	150
D.2.2 Organisation en FAT . . . . .	150
D.2.3 I-nodes . . . . .	151
D.2.4 Taille des blocs . . . . .	151
D.2.5 Gestion des blocs libres . . . . .	152
D.3 Cohérence et fiabilité . . . . .	152
D.4 Performance . . . . .	152
D.5 Ce que je zappe . . . . .	152
<b>E Le réseau</b>	<b>153</b>
E.1 Introduction . . . . .	153
E.1.1 Besoins . . . . .	153
E.1.2 Architecture globale . . . . .	153
E.1.3 En résumé : les défuntes couches OSI . . . . .	154
E.2 Les couches de l'internet . . . . .	154
E.2.1 IP . . . . .	155
E.2.2 Routage (commutation de paquets) . . . . .	155
E.2.3 UDP . . . . .	156
E.2.4 TCP . . . . .	156



# Chapitre 1

## Introduction

L'informatique c'est la science du traitement de l'information.

Un ordinateur est une *machine universelle de traitement de l'information*. Universelle veut dire : qui peut réaliser toute transformation d'information *réalisable*<sup>1</sup>

Ce cours va vous apprendre ce qu'est un ordinateur et comment on le construit.

### 1.1 Des livres sérieux, parce que ce poly, bon.

Le bouquin de référence de support de ce cours est *Computer Organization & Design, the hardware-software interface*, de Patterson et Hennessy (qui ont eu le prix Turing en 2017). Aux dernières nouvelles ils en sont à la 5ème édition, et derrière il y a des éditions pour différents processeurs. Celui à la mode c'est la version RISC-V (sorti en 2018).

Des mêmes auteurs mais dans le désordre, *Computer Architecture : A Quantitative Approach*, 6ème édition (pour une certaine valeur de 6) est plus avancé : il parle de l'optimisation des paramètres d'un processeur. Il y a eu des traductions en français de vieilles éditions. Attention, ce livre était bien dans ses premières éditions, puis il a grossi, est devenu illisible, enfin il s'est allégé du précédent et est redevenu lisible – mais pointu.

Plus vieux, mais très bien, il y a aussi *Structured Computer Organization*, par Andrew Tanenbaum, qui je crois s'est arrêté en 2005 à la 5ème édition. Il existe en français aussi sous le titre *Architecture de l'Ordinateur*.

Il y a aussi *Computer Architecture and Organisation* de Murdocca et Heuring, ainsi que *Computer Organization and Architecture* de William Stallings (11ème édition en 2019, notre record), tous deux très bien, d'ailleurs ils ont le même titre.

**Parenthèse culturelle** Le mot "Organisation" dans les titres de bouquins ci-dessus décrit le *comment de la construction d'un ordi*. Le mot "Architecture" décrit le *quoi*, sauf chez Hennessy et Patterson qui utilisent le mot "Design". Page 77 on retrouvera le mot *Architecture* avec ce sens précis dans le terme **Instruction Set Architecture**.

Cette nuance est lourdement expliquée dans les premières pages des 3 bouquins ci-dessus qui ont comme titre "Computer X and Y". Malheureusement, votre prof a tendance à utiliser le terme "Architecture" à la fois pour Computer Architecture et Computer Organization. Il n'est pas le seul.

---

1. Il existe des calculs non réalisables, et des calculs non réalisables en temps raisonnable. C'est pas que ce n'est pas important, mais on ne s'y intéresse pas dans ce cours.

## 1.2 Historique du calcul mécanique

**néolithique** Invention du système de numération unaire, de l'addition et de la soustraction (des moutons). Car en latin, *calculus* c'est un petit caillou.

**antiquité** Systèmes de numération plus évolués :

**systèmes alphabétiques** (Égypte, Grèce, Chine) :

chaque symbole a une valeur numérique fixe et indépendante de sa position. Des pièces de monnaie, quoi. Les chiffres romains sont un mélange de unaire et d'alphabétique.

**systèmes à position** (Babylone, Inde, Mayas) :

c'est la position du chiffre dans le nombre qui donne sa puissance de la base.

Les bases utilisées sont la base 10 (Égypte, Inde, Chine), la base 20 (Mayas), la base 60 (Sumer, Babylone), ...

*Ces inventions sont guidées par la nécessité de faire des calculs*

Essayez donc de décrire l'algo de multiplication en chiffres romains... Par contre la base 60 c'est bien pratique, pourquoi ?

**≈-1000** Invention du calculateur de poche (l'abaque ou boulier) en Chine.

**+1202** Le génial système de numération indo-arabe arrive en Europe par l'Espagne.

**1623** Sir Francis Bacon (Angleterre) décrit le codage des nombres dans le système binaire qu'il a inventé dans sa jeunesse.

**1623** Wilhelm Schickard (Tübingen) invente la roue à chiffres qui permet de propager les retenues.

**1624** Il construit la première calculette "occidentale"

**1645** Blaise Pascal en fabrique une mieux qui peut propager les retenues sur les grands nombres.

**1672** Gottfried Wilhelm Leibniz construit une machine à calculer 4 opérations

**1679** Leibniz encore développe l'arithmétique binaire (*De Progressione Dyadica*) mais sans l'idée que cela pourrait servir à quelque chose.

**1728** Première utilisation connue des cartes perforées (Falcon?)

**1741** Jacques de Vaucanson invente la mémoire de programme dans le contexte des métiers à tisser. Il utilise des rouleaux de fer-blanc perforés.

**1792** Claude Chappe est le précurseur de la couverture 5G en tapissant la France d'antennes-relais utilisées pour les télécommunications sans fil.

**1808** Joseph Marie Jacquard utilise du carton perforé, c'est moins cher que le fer-blanc.  
Surtout, il est Lyonnais.

**1822** Charles Babbage se lance dans sa *difference engine*, qui sert à calculer des polynômes.

**1833** Karl Friedrich Gauß et Wilhelm Weber, à Göttingen, sont les précurseurs de l'internet en s'envoyant à distance sur un fil électrique du texte codé par un genre de code Morse.

**1833** Babbage laisse tomber car il a une meilleure idée, l'*analytical engine*, programmable par cartes perforées, inconstruisible avec les moyens de l'époque.

**1843** Ada Lovelace invente (pour la machine de Babbage) le langage de programmation. Par ailleurs elle suggère que de telles machines peuvent être utilisées aussi pour manipuler du texte ou composer de la musique. On peut donc affirmer qu'elle est la précurseuse à la fois de Word et de la musique techno.

**1854** Georges Boole (Angleterre) met en place la logique mathématique désormais dite booléenne.

**1854** Christopher L. Sholes (USA) : première machine à écrire utilisable.

**1914** Leonardo Torres y Quevedo construit le premier automate qui joue aux échecs et, bien avant Deep Blue, gagne à tous les coups (dans un cas légèrement simplifié : la finale roi et tour contre roi seul).

**1915** Le même Leonardo décrit la construction d'automates (électromécaniques) à nombre d'états arbitraires dans *Ensayos sobre Automática* (Essais sur l'automatique. Sa définition. Etendue théorique de ses applications.)

**1920** Il construit les premiers calculateurs dont l'entrée est un clavier et la sortie une imprimante.

**1928** Ackermann montre qu'il y a des fonctions entières qui ne peuvent être calculées par un nombre fini de boucles (alors qu'on n'a pas encore inventé la boucle).

- 1900-1935** Développements en électronique : tube cathodique, tube à vide 3 électrodes, enregistrement sur support magnétique.
- 1936** Thèse d'Alan M. Turing sur une machine abstraite universelle. Début de la théorie de la calculabilité.
- 1936** Konrad Zuse (Allemagne) construit le premier calculateur binaire (Z1 : mécanique, Z2 : à relais).
- 1937** John V. Atanasoff (USA) a l'idée d'utiliser le binaire pour des calculateurs.
- 1941** Le Z3 de Zuse est le premier calculateur (presque) universel programmable
  - 1400 relais pour la mémoire, 600 pour le calcul
  - 64 mots de mémoire
  - arithmétique binaire en virgule flottante sur 22 bits
  - programmation par ruban perforé de 8 bits
  - une seule boucle (d'où le "presque" universel)
- 1943** Tommy Flowers, à Bletchley Park (là où bosse aussi Turing) construit le premier calculateur à tubes électroniques. Il n'est pas universel, il ne sait que déchiffrer des messages secrets. On n'apprendra son existence que dans les années 1970.
- 1946** L'ENIAC est le premier calculateur électronique plus ou moins programmable. Comparé au Z3 c'est un boulier (en moins pratique), mais il va bien plus vite.
- 1949** Turing et von Neumann, ensemble, construisent le premier ordinateur universel électronique, le Manchester Mark I. L'innovation c'est la mémoire partagée programme et donnée.
- 1969** Grace Hopper conçoit le premier compilateur. Jusque là on programmait uniquement dans le langage de la machine.
- 1969** l'*Apollo Guidance Computer* pose deux hommes sur la Lune. C'est un ordinateur embarqué à base de circuits intégrés, muni d'un clavier, d'un écran et même d'un joystick, et d'une interface réseau sans fil. Il possède aussi un système d'exploitation multi-tâches temps réel (conçu sous la direction de Margaret Hamilton). Comparé à votre smartphone, il ne manque donc que le haut-parleur, avantageusement remplacé par une vingtaine de moteurs fusées.
- 1982** Bien sûr, ce ne sont ni Ada ni Margaret ni Grace, mais deux barbus, Ken et Dennis, qui auront le prix Turing pour cette sale blague que sont le langage C et UNIX.
- Depuis** on n'a pas fait tellement mieux. Enfin si, mais les innovations sont difficiles à expliquer *avant* le cours d'ASR...

### 1.3 Objectifs du cours

Notre but est de construire un ordinateur moderne, superbement représenté sur la figure 1.1.

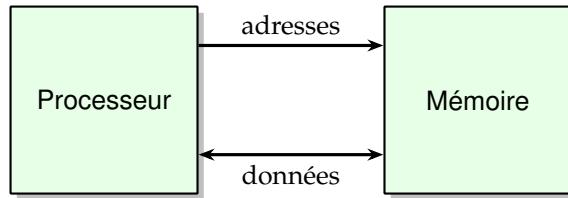


FIGURE 1.1 – *S'il te plaît, dessine-moi un ordinateur*

Et il y a une idée capitale cachée dans ce dessin, et que personne n'avait eu avant la bande à von Neumann : *C'est la même mémoire qui contient le programme et les données*.

C'est génial car cela permet par exemple

- le *système d'exploitation* : un programme qui prend un paquet de données sur un disque dur, les met en mémoire, et ensuite décide que ces données forment un programme, et exécute ce programme.
- le *compilateur* qui prend un texte (des données) et le transforme en un programme...

La mémoire est un ensemble de cases mémoires numérotées par leur *adresse*. Dans une case on peut avoir une quantité fixe, finie d'information – par exemple un nombre fixe de bits. A part cela, le contenu d'une case mémoire n'est pas typé a priori : l'information contenue dans une case peut être interprétée n'importe comment<sup>2</sup>.

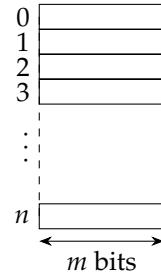


FIGURE 1.2 – *Vue logique de la mémoire*

Le processeur réalise le *cycle de von Neumann* :

1. Lire une case mémoire d'adresse PC (envoyer l'adresse à la mémoire, et recevoir en retour la donnée à cette adresse)
2. Interpréter cette donnée comme une instruction, et l'exécuter
3. Augmenter PC pour passer à la case suivante
4. Recommencer

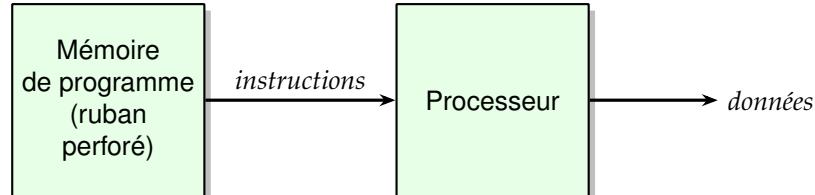
PC c'est pour *program counter*, bande de gauchistes. La fréquence de votre ordinateur favori, c'est la fréquence à laquelle il exécute ce cycle.

On détaillera plus tard la notion d'instruction. Certaines des instructions que peut interpréter le processeur font des calculs, d'autres vont interagir avec la mémoire (lecture ou écriture), d'autre enfin vont modifier le PC lui-même : ainsi le programme peut sauter des paquets d'instructions, ce qui permet l'exécution conditionnelle if-then-else, ou bien revenir en arrière, ce qui permet les boucles.

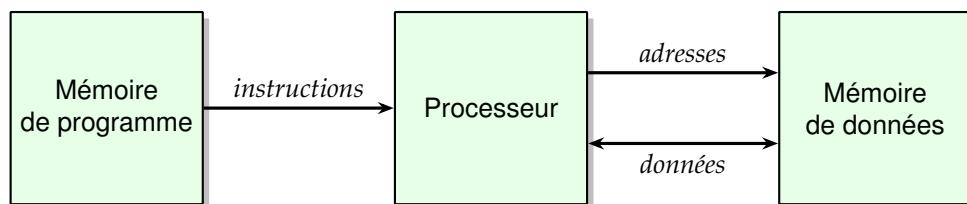
Dans ce cours, on va essayer de se concentrer sur les techniques qui sont indépendantes de la technologie. En principe on pourra construire notre ordinateur en utilisant l'électronique actuelle, mais aussi en Lego, ou bien avec des composants moléculaires opto-quantiques à nanotubes de carbone.

2. La citation d'Ada Lovelace, qui montre que c'est elle qui a inventé l'iPod : *[The Analytical Engine] might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine... Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.*

**Parenthèse : une perspective historique sur les machines programmables** Dans un orgue de barbarie, il y a un programme qui produit des données (de la musique) mais il est linéaire : il n'y a pas besoin d'adresse du côté du programme, on se contente d'avancer le ruban perforé. A Lyon on construisait des métiers à tisser qui assemblaient des centaines de fils en un tissu avec un dessin, selon un programme également stocké sur ruban perforé. Conceptuellement c'est le même dessin :

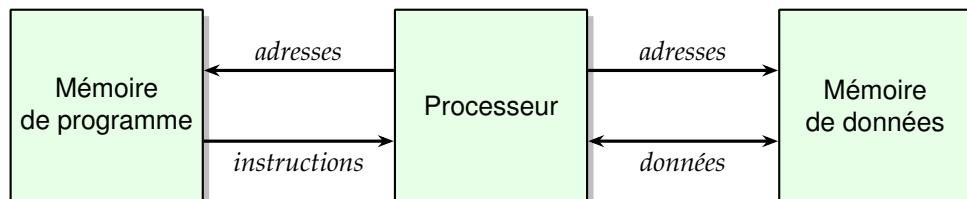


Il était donc naturel de séparer la mémoire de programme et la mémoire de donnée dans les premiers ordinateurs. On appelle cela une architecture de Harvard, à cause du Harvard Mark 1, un ordinateur électromécanique opérationnel en 1944 dont la mémoire de programme était un ruban perforé. Chez la concurrence, le Zuse Z3 (opérationnel en 1942) était également programmé par ruban perforé. En plus de cette mémoire de programme, les deux possédaient une mémoire de données contenant essentiellement des nombres. Voici donc le dessin qui décrit ces prototypes héroïques :



(il paraît que le Zuse pouvait faire des boucles mais je ne sais pas comment. Pour l'Eniac, vous remplacez la mémoire de programme par un gros tableau plein d'interrupteurs.)

Dans ce qu'on appelle architecture de Harvard, la mémoire de programme est également adressable, ce qui permet boucles et sauts. C'est le dessin ci-dessous, une évolution naturelle des calculateurs précédents.



Il faut comprendre que c'est par contraste avec l'architecture de Harvard que l'architecture de von Neumann marque une rupture.

(La suite de cet encadré n'est pas supposée compréhensible à première lecture : essayez de la comprendre en phase de révision.) On continue de fabriquer des architectures de Harvard pour des raisons pratiques : bien que le dessin ci-dessus semble plus compliqué que la figure 1.1, il est plutôt plus simple à construire. Surtout, il permet de dimensionner séparément la mémoire de programme et la mémoire de données, c'est à dire le nombre de bits que porte chacun des 4 fils du dessin ci-dessus. Cette architecture est donc utilisée dans des processeurs limités/embarqués/specialisés :

- processeurs de traitement de signal (petite mémoire de programme et instructions sur peu de bits, grosse mémoire de donnée avec des nombres de 24 ou 48 bits)
- microcontrôleurs conçus pour être programmés “de l'extérieur”, donc on s'en fiche des avantages du modèle de von Neumann. Exemples bien connus des bricoleurs/hackers : les microcontrôleurs PIC, les AVR utilisés par les Arduino.

Enfin, bien que les processeurs modernes soient des machines de von Neumann, leur micro-architecture (leur organisation) a tendance à séparer le cache de programme du cache de données, essentiellement pour des raisons de performance (ainsi on peut accéder aux deux en même temps). Il y a toutefois des mécanismes qui maintiennent l'illusion d'une machine de von Neumann.

## 1.4 La conception d'ASR est hiérarchique

Selon le bon vieux paradigme *diviser pour régner*, on est capable de construire l'objet très complexe qu'est votre PC par assemblage d'objets plus simples. Au passage, on utilise différents formalismes (ou différentes abstractions) pour le calcul (arithmétique binaire, fonctions booléennes, etc), pour la maîtrise des aspects temporels (mémoires, automates, etc), pour les communications (protocoles)... La figure 1.3 décrit partiellement cet empilement.

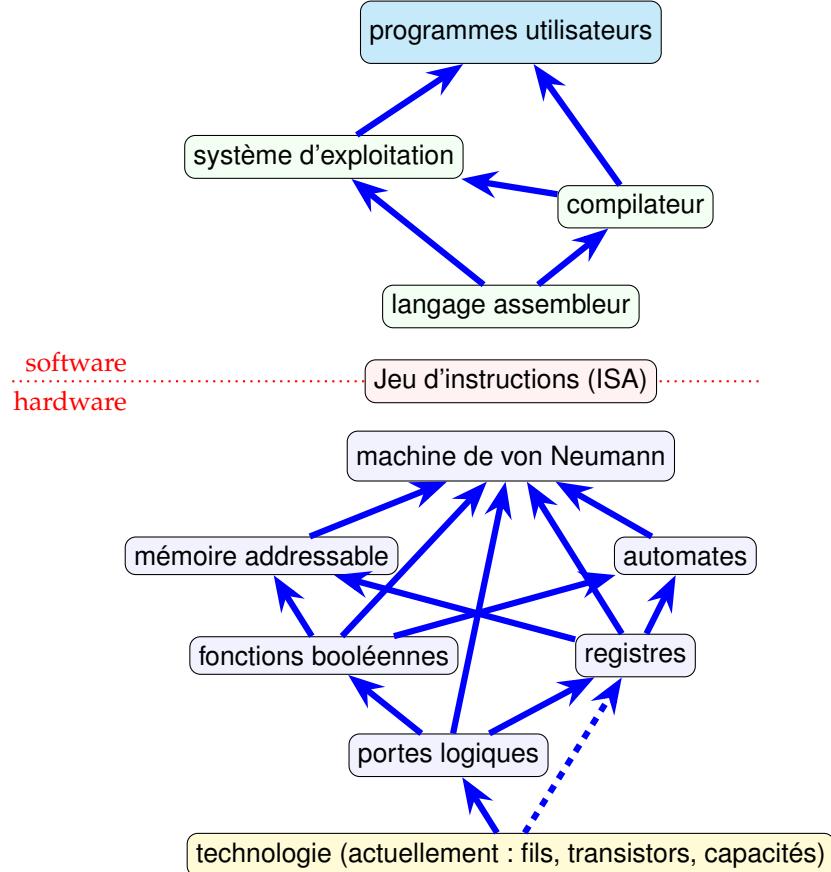


FIGURE 1.3 – En gros, le plan du cours (il faut partir du bas)

Dans ce cours on va remonter cette figure de base en haut (construction *bottom-up*), mais avec des détours et des zig-zags.

Maintenant qu'on a une recette pour faire des systèmes complexes, voyons les limites pratiques à cette complexité.

## 1.5 Quelques ordres de grandeur

### 1.5.1 L'univers est notre terrain de jeu

Il y a quelques limites aux ordinateurs qu'on saura construire, par exemple :

$$-\frac{\text{Masse de l'univers}}{\text{Masse du proton}} \approx 10^{78}$$

(ce qui limite le nombre de transistors sur une puce. Actuellement on est à  $10^{10}$ , il y a de la marge).

— Vitesse de la lumière :  $3 \cdot 10^8 \text{ m/s}$

— Distance entre deux atomes :  $\approx 10^{-10} \text{ m}$ .

Taille de la maille du cristal de silicium :  $\approx 5 \cdot 10^{-10} \text{ m}$

(Actuellement, un fil ou un transistor font quelques dizaines d'atomes de large : en 2018, on vend de processeurs "en technologie 7 nm", donc une quinzaine de mailles de cristal... Le problème c'est surtout les couches d'isolant utilisées, actuellement entre 2 et 3 atomes d'épaisseur, elles isolent de moins en moins)

— Par conséquent, le temps minimum physiquement envisageable pour communiquer une information est de l'ordre de  $\approx 10^{-18} \text{ s}$ .

- On a fait plus de la moitié du chemin, puisque nos transistors commutent à des fréquences de l'ordre de  $10^{12}$  Hz.
- Actuellement, un signal n'a pas le temps de traverser toute la puce en un cycle d'horloge.
- Actuellement, on manipule des charges avec une résolution correspondant à une centaine d'électrons. Cela permet à votre téléphone portable de gérer des fréquences au Hz près dans les 600MHz

### 1.5.2 La technologie en 2018

Les ordinateurs actuels sont construits en assemblant des transistors (plein<sup>3</sup>). La figure 1.4 montre les progrès de l'intégration des circuits électroniques. Il y a une loi empirique, formulée par un dénommé Moore chez Fairchild dans les années 60 (après il a rejoint Intel) et jamais démentie depuis, qui décrit l'augmentation exponentielle de la quantité de transistors par puce (la taille de la puce restant à peu près constante, de l'ordre de 1cm<sup>2</sup>).

La formulation correcte de la loi de Moore est : le nombre de transistors qu'on sait intégrer sur une puce économiquement viable double tous les deux ans.

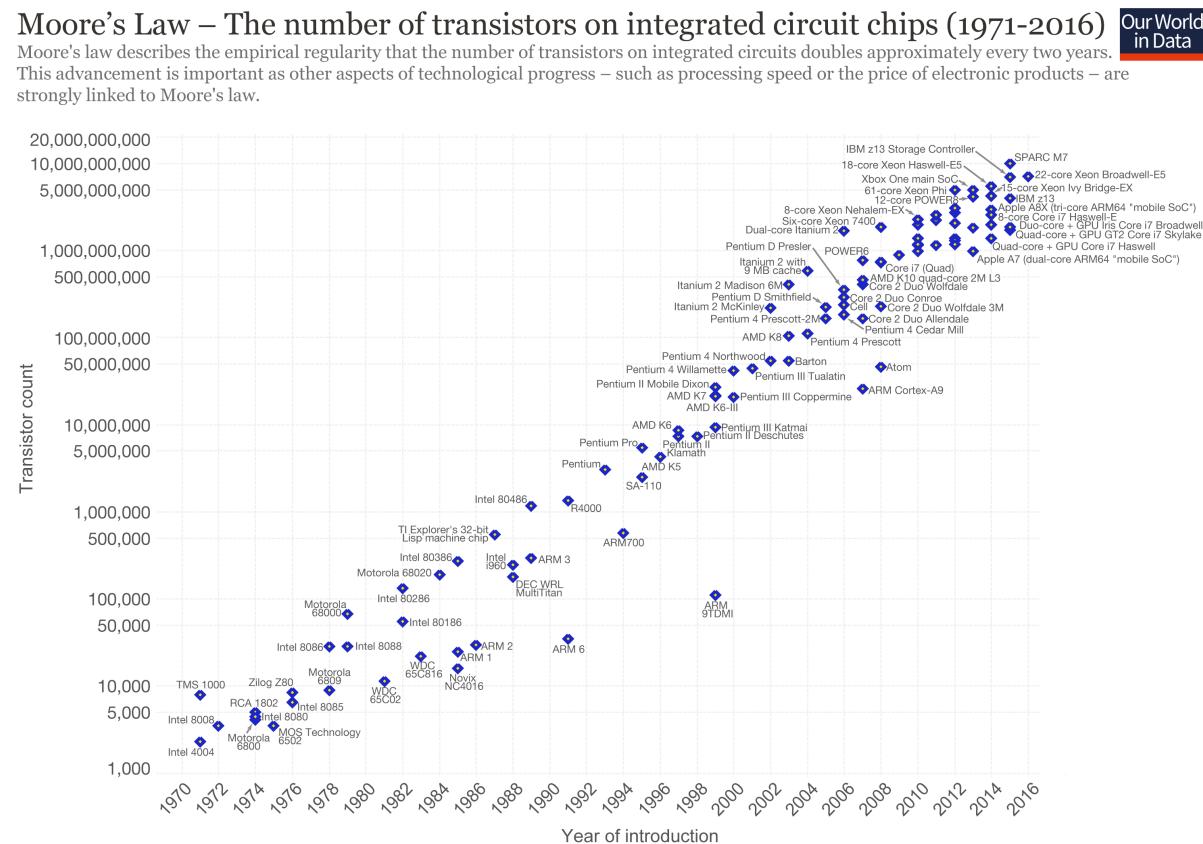


FIGURE 1.4 – La “loi” de Moore (©Wikipedia/creative commons), d'ailleurs vous trouvez une version plus récente sur Wikipedia

Pour illustrer le chemin accompli : le premier processeur (l'Intel 4004, en bas à gauche de la figure 1.4) comptait 2300 transistors. De nos jours on pourrait mettre 2300 processeurs 32-bit complets dans le budget de transistors d'un processeur de téléphone portable.

Ce doublement tous les deux ans se traduit par exemple directement par un doublement de la mémoire qu'on peut mettre dans une puce. Mais ce n'est pas tout : les transistors étant plus petits, ils sont aussi plus rapides (j'expliquerai peut-être avec les mains pourquoi, admettez en attendant). La puissance de calcul des processeurs peut donc augmenter donc plus vite que ce facteur deux tous les deux ans (les processeurs sont aussi de plus en plus complexes, ce qui tire dans l'autre sens).

Enfin, c'était vrai jusqu'aux années 2000 : en réduisant la taille du transistor, on en mettait plus sur la puce, ils étaient plus rapides, et ils consommaient moins. Avec les dernières générations (dites sub-microniques, c'est-à-dire en gros que le transistor fait moins d'un micron), cela se passe moins bien :

3. A partir de 2004 on a produit plus de transistors par an dans le monde que de grains de riz. Je ne sais pas si c'est une bonne nouvelle.

on réduit toujours le transistor, cela permet toujours d'en mettre plus par puce, mais ils sont de moins en moins plus rapides<sup>4</sup> et consomment de plus en plus (en 2007, sur les 150W que consommait le processeur d'un PC de gamerz, il y en avait déjà 50 qui partaient en courants de fuite). Une raison facile à comprendre est qu'on arrive à des couches d'isolant tellement fines et à des fils tellement proches que les électrons passent par effet tunnel d'un fil à l'autre. Il y a d'autres raisons.

En pratique, les experts ne donnent pas 10 ans de plus à la loi de Moore pour le silicium<sup>5</sup>. Heureusement il y a de la marge de progression après la fin de la loi de Moore, et c'est par de meilleures architectures<sup>6</sup> (le sujet de ce cours).

Autre évolution de la technologie : l'investissement (en terme de megadollars) pour passer d'une génération technologique à la suivante double aussi avec chaque génération... Cela se traduit par des regroupement de firmes pour partager ces coûts, et à ce train, il n'y aura plus qu'un seul fabricant de circuits intégrés dans dix ans.

Tout ceci pour dire que d'ici à ce que les plus brillants d'entre vous fassent à leur tour un cours d'architecture, la technologie sera sans doute en train de changer profondément. Vers quoi, personne ne sait vraiment<sup>7</sup> En tout cas je vais essayer de ne pas trop perdre du temps sur la technologie actuelle. Par contre je la connais assez pour répondre à vos questions.

Un dernier aspect technologique : les transistors de plus en plus petits sont de moins en moins fiables, et il faut le prévoir quand on fait des circuits. Par exemple, on évaluait en 2007 que les rayons cosmiques font statistiquement changer un bit par puce et par mois... Et cela ne s'arrange pas : plus on intègre, moins il faut d'énergie pour faire changer un bit. Ce sont des "soft errors", la puce n'est pas endommagée. Les mémoires actuelles sont munies de dispositifs d'autocorrections.

### 1.5.3 Calcul contre stockage et déplacement de données

D'abord, la consommation électrique est un problème, et pas que pour les téléphones portables. Le *supercomputing center* de l'UCSD consomme 1/4 de l'électricité du campus. Allez chercher sur internet combien il faut d'énergie pour faire une addition avec une calculette, et avec la barre Google, et jurez qu'on ne vous y reprendra plus.

En vieille techno 45 nm, il fallait

- 0.05 pJ pour faire une addition 32 bits
- 8 pJ pour bouger une donnée de 32 bits d'1mm sur la puce
- 320 pJ pour sortir le même mot de la puce.

Pour 50W, on peut

- faire  $10^{15}$  additions 32 bits par seconde
- lire de la DRAM à 160Go/s, soit "seulement"  $4.10^{10}$  mots 32bits par seconde...

Tous ces chiffres ne seront pas demandés à l'examen, par contre retenez que le calcul n'est pas forcément ce qui est coûteux, surtout s'il reste localisé sur un coin microscopique de puce. Balader les données de droite à gauche, sur des distances macroscopiques<sup>8</sup>, demande une énergie macroscopique. Bref, évaluer un ordinateur en ne comptant que les opérations qu'il fait, c'est un peu à côté de la plaque.

### 1.5.4 Ya pas que les PC dans la vie

Juste trois exemples :

Bien qu'on vous fasse surtout programmer sur des PC, les processeurs les plus vendu au monde ne sont pas de la famille Intel, ce sont les membres de la famille ARM, conçue pour coûter pas cher et consommer peu d'énergie, et qui équipe la plupart des téléphones portables (sans compter les GameBoy, les oreillettes bluetooth, les machines à laver, etc). Il se vend en 2021 plus d'un cœur ARM par humain et par mois. Il y a des processeurs ARM peu puissants qui consomment quelques miliwatts, et des plus gros dont la performance est comparable à celle du x86 quelques années auparavant<sup>9</sup> (voir la figure 1.4, mais avec une consommation électrique bien plus faible).

4. En fait on saurait les faire fonctionner plus vite, mais on ne sait pas comment évacuer la chaleur produite dans ce cas...

5. Cela fait presque 15 ans que cette phrase est dans le poly et qu'elle reste vraie. Cela dit, le consortium chargé de l'autoréalisation de cette prophétie autoréalisatrice, l'association ITRS pour *International Technology Roadmap for Semiconductors*, a baissé les bras en 2015...

6. [http://www.theregister.co.uk/2016/09/22/the\\_evolution\\_of\\_moore\\_s\\_law\\_suggests\\_hardware\\_is\\_eating\\_software/](http://www.theregister.co.uk/2016/09/22/the_evolution_of_moore_s_law_suggests_hardware_is_eating_software/)

7. [https://www.theregister.com/2021/08/05/moores\\_law\\_what\\_next/](https://www.theregister.com/2021/08/05/moores_law_what_next/)

8. Notez au passage que si on mettait bout à bout tous les fils qui relient les différents transistors de votre téléphone portable, on arriverait à quelques centaines de km.

9. Les Macophiles vous diront même que le processus Apple M1 sorti en 2020 surpassé les processeurs intel, mais méfiez-vous de ces gens-là.

Une vieille télé Philips de 2002 contenait une puce (Viper2) devant traiter 100 GOps ( $10^{11}$  opérations par seconde). Cette puce incluait 4 microprocesseurs programmables, 250 RAMs, 60 autres blocs de calculs divers, 100 domaines d'horloge différents, et en tout 50Mtransistors tournant à 250 MHz (source : M. Duranton, Philips, Euromicro DSD 2006).

Un mobile 3G doit calculer entre 35 et 40 Gops (milliards d'opérations par seconde) pour traiter un canal de 14.4 Mbps, en mangeant moins d'1W. Rajoutez au moins un zéro par génération (4G, 5G).

Avec tout cela, votre Core i7 tout neuf est ridicule, avec ses 90W il atteint à peine le TOps (téraprécisions par seconde). Bien sûr, on ne parle pas des mêmes opérations : le téléphone portable mouline de petits nombres de quelques bits, alors que le PC travaille sur de gros nombres de 32 ou 64 bits. De plus, le PC est généraliste. Cela se traduit par le fait qu'il dépense 6% de son énergie à calculer, et le reste à déplacer les données et les instructions à traiter (source : Dally et al dans *IEEE Computer* de juillet 2008, et cela n'a fait qu'empirer depuis).

Cela dit, il paraît qu'en 2025, la consommation de veille des objets internet sera la consommation du Portugal en 2012. Je ne suis pas certain que ce soit une bonne nouvelle.



## **Première partie**

### **L'information et comment on la traite**



# Chapitre 2

## Coder l'information

### 2.1 Généralités

#### 2.1.1 Information et medium

La notion d'information est une notion abstraite. On appelle medium un support physique de l'information, que ce soit pour la stocker (CD, journal) ou la transmettre (fil, onde radio, écran télé).

Le medium a un coût, l'information aussi, il faut bien distinguer les deux. Quand on achète un CD<sup>1</sup>, on paye 1 euro le medium, et 10 euros l'information. Quand on le pirate, on économise le coût de l'information, mais si on le grave on n'économise pas le coût du medium. Quand on convertit un CD légalement acheté en mp3 sur son disque dur, on a dupliqué l'information, et on n'est pas deux fois plus riche pour autant. C'est une notion que l'industrie du disque a du mal à intégrer.

Formellement, l'information est un bien non rival. Les économistes définissent un bien *rival* comme un bien dont la jouissance par l'un exclut la jouissance par un autre (une baguette de pain, une fois qu'elle est mangée par l'un, ben y en a plus pour l'autre). Toute l'économie s'étant construite sur des biens rivaux, y intégrer une économie de l'information n'est pas évident, et à ce jour cette question n'est pas réglée.

#### 2.1.2 Information analogique

L'information peut être discrète ou continue. La nature est pleine d'informations continues, que l'on sait enregistrer sur des supports analogiques (disque vinyle, photo argentique, etc), et transmettre par fil, par radio, etc. Les machines qui traitent ce genre d'information sont dites analogiques. Les traitements les plus courants sont *amplification* et *filtrage*.

Quelques exemples de *calculateurs analogiques* plus généraux :

- les horloges astronomiques ;
- Mr Thompson/Lord Kelvin a construit à base de roues dentées une machine à prédire les marées dans chais plus quel port ;
- dans les années 50, il y avait à Supelec une “salle de calcul analogique” où l'on pouvait brancher des amplis ops entre eux pour simuler, plus vite, des phénomènes physiques. Par exemple, on a pu ainsi simuler tout le système de suspension de la Citroen DS avant de fabriquer les prototypes.

Ces derniers temps<sup>2</sup>, on s'est rendu compte que c'est plus propre et plus facile de *discretiser* l'information d'abord et de la traiter ensuite avec un calculateur numérique.

Jusqu'aux années 90, il ne restait d'analogique dans les circuits intégrés que les parties qui devaient produire des ondes à hautes fréquence : les téléphones mobiles de la première génération se composaient d'une puce numérique, et d'une puce analogique branchée sur l'antenne. De nos jours on peut faire un téléphone en une seule puce : les transistors actuels commutent tellement vite qu'on peut tout faire en numérique. Un avantage supplémentaire est que les composants analogiques ne peuvent pas être miniaturisés autant qu'on veut : dans la surface typique d'une inductance, on peut mettre 4 processeurs ARM complets en technologie de 2013.

#### 2.1.3 Information numérique

Si elle est discrète, l'unité minimale d'information est le *bit* (contraction de *binary digit*), qui peut prendre juste deux valeurs, notées 0 ou 1 (mais vous pouvez les appeler vrai et faux, ou  $\emptyset$  et  $\{\emptyset\}$ , ou yin

1. Oui, cette phrase a été écrite à une époque où on achetait des CD pour écouter de la musique.  
2. Lire : un peu avant votre naissance

et yang si cela vous chante)<sup>3</sup>.

Pour coder de l'information utile on assemble les bits en vecteurs de bits. Sur  $n$  bits on peut coder  $2^n$  informations différentes.

Une information complexe sera toujours un vecteur de bits. Le code est une relation entre un tel vecteur et une grandeur dans un autre domaine.

Le précepte central du codage est : **un bon code est un code qui permet de faire facilement les traitements utiles.**

## 2.1.4 Coder le temps

Le temps physique est continu, on peut le discréteriser. Parmi les instruments de mesure du temps, les plus anciens utilisent une approche analogique du temps continu (clepsydre, sablier). Les plus récents utilisent une approche numérique/discrète : horloge à balancier, à quartz... Ici aussi, le passage au monde discret (par le balancier) permet ensuite de transformer cette information sans perte : les roues dentées de l'horloge ont un rapport qui est une pure constante mathématique, ce qui permet de construire un dispositif dans lequel une heure fait toujours exactement 3600 secondes. Il ne reste dans une telle horloge qu'un seul point de contact avec le monde analogique : le balancier, qui donne la seconde.

Et tout vaché de chouette idée qu'elle est, une montre à quartz numérique c'est juste pareil. Il y a entre le sablier et l'horloge normande la même différence révolutionnaire qu'entre le vinyle et le CD (passage du traitement analogique au traitement numérique de l'information). Par contre, la montre à quartz n'est pas fondamentalement différente de l'horloge normande.

Plus pratiquement, on codera toujours un *instant* par un *changement* d'information. C'est vite dit mais il faudra s'y arrêter longtemps... en temps utile.

## 2.2 Coder des nombres entiers

Il y a plein de manières de coder les entiers en binaire. Toute bijection de  $\{0, 1\}^n$  dans un ensemble d'entiers fait l'affaire.

Suivons le précepte ci-dessus. Quels sont les traitements utiles qu'on peut vouloir faire sur les entiers ? On aime par exemple avoir la relation d'ordre, et faire des opérations comme addition, multiplication...

Le codage en unaire fut sans doute le premier. On y représente un nombre  $n$  par un tas de  $n$  cailloux (ou batons, ou "1"). Le gros avantage de ce codage est que l'addition de deux nombres est réalisable par une machinerie simple qui consiste à mélanger les deux tas de cailloux. La comparaison peut se faire par une balance si les cailloux sont suffisamment identiques, sinon vous trouverez bien quelque chose. Plus près de nous, j'ai lu récemment un papier sérieux proposant des circuits *single electron counting* : il s'agit de coder  $n$  par  $n$  électrons dans un condensateur. La motivation est la même : l'addition de deux nombres consiste à vider un condensateur dans l'autre. La comparaison se ramène à une différence de potentiel.

Le gros problème du codage unaire est la représentation des grands nombres. Dans l'exemple du *single electron counting*, compte tenu des imperfections des composants utilisés, les nombres ne doivent pas dépasser la douzaine si on veut être capable de les discriminer. Notre précepte nous dit que ce codage n'est pas bon pour coder les grands nombres.

On a inventé les codages de position comme par exemple celui que vous avez appris en maternelle. La version la plus simple est le codage binaire : un tuple  $(x_0, x_1, \dots, x_n)$  représenté par exemple un entier  $X = \sum_{i=0}^{n-1} 2^i x_i$ . Commençons par ceux-ci.

### 2.2.1 Numération de position pour les entiers naturels

Soit  $\beta \in \mathbb{N}$ ,  $\beta > 1$ , une base. Tout  $n \in \mathbb{N}$  peut être représenté de manière unique par sa *représentation positionnelle en base  $\beta$*  :

$$(x_{p-1}x_{p-2} \cdots x_1x_0)_{\beta} := \sum_{i=0}^{p-1} x_i \beta^i.$$

Les  $x_i \in \{0, 1, \dots, \beta - 1\}$  sont les *chiffres* de l'écriture de  $n$  en base  $\beta$ ,  $p$  est le nombre de chiffres nécessaires pour écrire de l'entier naturel  $n$ . On attribue un symbole à chaque chiffre : chiffres 0 et 1 en binaire, chiffres de 0 à 9 en décimal, chiffres de 0 à F en hexadécimal.

3. En français, "bit" est un peu un gros mot, c'est fâcheux. J'ai donc proposé à l'Académie Française de le remplacer par la contraction équivalente de *chiffre binaire*, ce qui donne *chibre*. Bizarrement, l'adoption officielle de ce terme traîne un peu.

### Changement de base par sommation des puissances

La définition permet directement d'effectuer des changements de base : si on connaît l'écriture  $(x_{p-1} \dots x_0)_\beta$  de  $n$ , et qu'on veut connaître cette écriture en base  $\gamma$ , il suffit de convertir les chiffres  $x_i$  en base  $\gamma$ , puis de calculer  $\sum_{i=0}^{p-1} x_i \beta^i$  en base  $\gamma$ .

### Changement de base par divisions euclidiennes

Une autre technique de changement de base est celle des divisions euclidiennes successives. Rappelez et justifiez cette méthode

Le reste dans la division euclidienne d'un entier naturel  $n$  par  $\beta$  donne le chiffre de poids faible dans la représentation de  $n$  en base  $\beta$ . En effet,

$$\begin{aligned} n &= x_{p-1} \cdot \beta^{p-1} + x_{p-2} \cdot \beta^{p-2} + \dots + x_2 \cdot \beta^2 + x_1 \cdot \beta^1 + x_0, \\ &= \underbrace{(x_{p-1} \cdot \beta^{p-2} + x_{p-2} \cdot \beta^{p-3} + \dots + x_2 \cdot \beta^1 + x_1)}_{\text{quotient}} \cdot \beta + \underbrace{x_0}_{\text{reste}}, \end{aligned}$$

avec  $0 \leq x_0 < \beta$ .

En d'autres termes,  $n \bmod \beta = x_0$ . On peut donc obtenir tous les chiffres de l'écriture d'un entier naturel  $n$  par des divisions euclidiennes successives, en s'arrêtant au premier quotient nul : on obtient les chiffres de poids faibles d'abord.

## 2.2.2 Ordres de grandeurs en puissances de 2

$$2^{10} = 1024 \approx 10^3$$

d'où les Ki=2<sup>10</sup> (se lit kibi mais l'abus de langage kilo est courant et accepté), Mi=2<sup>20</sup>, Gi=2<sup>30</sup>, etc.

Il y a même depuis 1999 une norme internationale qui définit tout cela : [https://en.wikipedia.org/wiki/Binary\\_prefix](https://en.wikipedia.org/wiki/Binary_prefix)

Il faut savoir faire de tête les conversions à la louche (c'est-à-dire aux 2% de différence entre 1000 et 1024 près). Exemple d'exercice d'une grande utilité pratique : une mole d'atomes de carbone c'est  $6.02 \times 10^{23}$  atomes et cela pèse 12 gramme. En supposant qu'on trouve une technologie qui stocke un bit par atome de carbone, est-ce que 2<sup>64</sup> bits de mémoire pèsent plus ou moins d'un gramme ?

## 2.2.3 Parenthèse pratique : superposition des codages

Lorsqu'on construit un calculateur, on est souvent appelé à superposer des codages. Pas de grande théorie ici, mais une application du précepte (le bon codage c'est celui qui permet le bon traitement) à plusieurs niveaux de granularité. Quelques exemples :

- le boulier code des grands nombres en base 10, et chaque chiffre est représenté en unaire. C'est un codage très bien adapté à la technologie d'il y a 2000 ans : on peut reconnaître les chiffres d'un coup d'œil; on peut les additionner d'un mouvement du doigt; etc.
- Les calculettes bon marché fonctionnent également en base 10, avec chaque chiffre codé en binaire.
- Mon papier sur *single electron counting* reconnaît que cela ne fonctionne que pour un nombre limité d'électrons dans un condensateur, et utilise ces condensateurs comme chiffres dans une représentation de position classique.

## 2.2.4 Codage des entiers relatifs

Si on veut gérer des entiers relatifs, on peut ajouter un bit qui contient l'information de signe. C'est ce que vous avez appris en 6ème. Un défaut de cette technique est que vous avez du apprendre deux algorithmes complètement différents pour l'addition et la soustraction.

Nous introduisons ici un codage un peu plus zarbi, mais qui permet une implémentation très simple de la soustraction dès lors qu'on a l'addition. Je commence par le présenter en décimal pour vous en convaincre.

### En décimal : complément à 10 sur trois chiffres

On va travailler en base 10 avec des nombres à 3 chiffres. N'avoir que trois chiffres signifie qu'on doit ignorer tous les chiffres qui sortent de ce cadre. Mathématiquement, cela s'appelle travailler modulo 1000 (où  $1000 = 10^3$ ). Par exemple  $999+3 = 1002$  mais on n'a pas la place pour le 1 de gauche donc sur trois chiffres  $999+3=002$ .

Pour coder -1, demandons nous ce que vaut 000-001 modulo 1000. Eh bien c'est aussi  $1000-1=999$ . Le code naturel de -1 sur trois chiffres est donc 999.

Plus généralement, on va dire que

- les codes 000 à 499 représentent les entiers positifs habituels
- 999 représente -1. Remarque : c'est 1000-1.
- 998 représente -2. Remarque : c'est 1000-2.
- ...
- 501 représente -499. Remarque : c'est 1000-499.
- 500 représente -500 (vous avez compris)

Et maintenant, pourquoi c'est un bon codage (on répète en choeur le précepte) :

- Le signe n'est plus explicite, mais il suffit de regarder le chiffre de gauche pour le connaître : de 0 à 4, on a affaire à un positif, de 5 à 9 c'est un négatif. Quand on transposera tout cela en binaire, le bit de gauche sera directement le bit de signe.
- Notre algo d'addition du CE1 marche pour les négatifs comme pour les positifs, à condition d'ignorer la retenue qui sort. Illustration : calculons  $17 + (-2)$ . On envoie un étudiant au tableau pour qu'il pose  $017 + 998$ . Après un laps de temps plus ou moins long, il trouve 1015. On ignore la retenue sortante, c'est-à-dire qu'on reste sur 3 chiffres, et on a bien calculé 15.
- Pour calculer  $abc-xyz$  on calcule  $abc+(1000-xyz)$ . Comme  $1000-xyz$  c'est encore compliqué, on calcule en fait  $abc+(999-xyz)+1$ . Et là tout devient simple :
  - La soustraction  $999-xyz$  se fait chiffre à chiffre (pas de propagation de retenue). On obtient le code  $pqr$ .
  - Le calcul  $abc+pqr+1$ , c'est l'algo d'addition du CE1, le +1 consistant juste à ajouter systématiquement une retenue à la colonne de droite, qui n'en avait pas.

Et donc on a bien ramené la soustraction à l'addition à peine modifiée. Et elle marche pour les positifs comme pour les négatifs (illustration en envoyant un autre étudiant d'excellence calculer  $17-42$  puis  $-17 - 42$ ).

### En binaire : complément à 2 sur $p$ chiffres

Transposons tout ceci en binaire sur 4 bits (Figure 2.1) :

- Les codes 0000 à 0111 représentent les entiers 0 à 7 comme si de rien n'était
- Le code 1111 représente -1 (remarque : c'est  $1000_2 - 1$ )
- ...

Remarque : ignorer la retenue sortante c'est travailler modulo  $1000_2$ , c'est pourquoi la figure est un disque.

Et en binaire tout est plus simple :

- Le signe est directement donné par le bit de gauche.
- pour calculer  $abcd-xyzt$  on calcule  $abcd+(1111-xyzt)+1$ .
- la soustraction chiffre à chiffre est un *non* logique.
- et l'algo d'addition du CE1 marche très bien.

**Remarque** Un autre point de vue est de dire que le bit le plus à gauche a le poids qu'il aurait en binaire classique, mais *avec un signe négatif*. Par exemple 1111 représente  $(-8)+4+2+1=-1$ . Ceci s'explique, encore une fois, par le modulo.

Généralisation : en complément à 2 sur  $p$  bits, le vecteur de bits  $(c_{p-1}c_{p-2} \dots c_1c_0)$  représente la valeur  $-c_{p-1}2^{p-1} + \sum_{i=0}^{p-2} c_i 2^i$ .

Autrement dit, le bit le plus à gauche a un poids négatif  $(-2^{p-1})$ . Vérifiez que  $n \geq 0$ ssi  $c_{p-1} = 0$ , et  $n < 0$ ssi  $c_{p-1} = 1$ .

**Important** Comme le bit de gauche est différent des autres, il ne faut jamais parler juste de complément à 2, mais toujours préciser lourdement "sur  $p$  bits". C'est différent du binaire non signé, où la taille importe peu.

**Conséquences pratiques** Avec un type entier signé dans votre langage favori, si ce type est implémenté par du complément à 2 (typiquement sur 16, 32 ou 64 bits) l'addition de deux positifs peut vous donner un résultat négatif... Si vous n'y prenez pas garde, votre boucle `for` risque de devenir une boucle infinie.

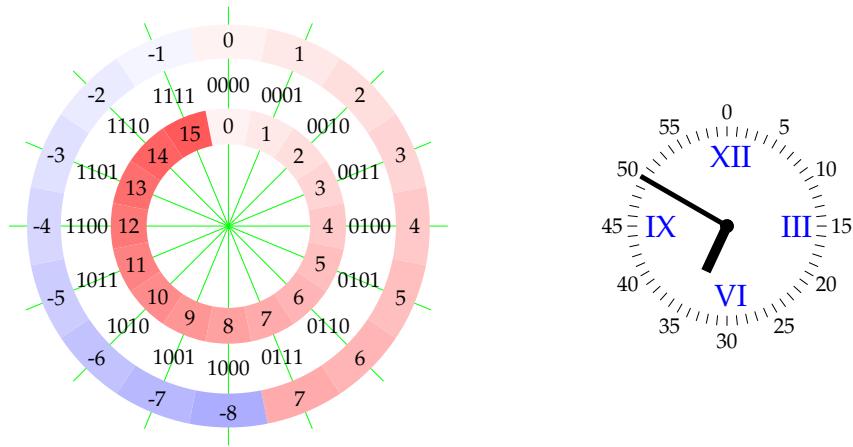


FIGURE 2.1 – Illustration du complément à 2 sur 4 bits (complément à  $2^4$ ). Pour chaque code, on donne son interprétation non signée ( cercle intérieur) et son interprétation en complément à 2 ( cercle extérieur). Remarquez l'analogie avec l'horloge : on peut dire “sept heures moins dix” ou “six heures cinquante” parce que  $50 = -10$  modulo 60.

**Détection des dépassement de capacité dans l'addition en complément à 2** L'idée est en gros la suivante :

- Si les deux entrées ont des signes différents, il ne peut y avoir de dépassement de capacité.
- Si elles ont le même signe, il y aura un dépassement si et seulement si le signe du résultat n'est pas le signe des entrées.

**Et pourquoi cela s'appelle complément à 2 ?** Une hypothèse à la fois plausible et morale est que cela aurait du s'appeler *complément à  $2^p$* , et que quelqu'un a été négligent. Ce serait moral parce que, encore une fois, la mention de  $p$  est indispensable quand on parle de complément à 2.

## 2.2.5 Et pour aller plus loin

... On peut utiliser une représentation de position avec des *chiffres négatifs*. Par exemple, du décimal avec les chiffres  $\{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$ . En notant plutôt le chiffre  $-5$  par  $\bar{5}$ , on compte comme ceci :  $1, 2, 3, 4, 5, \bar{4}, \bar{3}, \bar{2}, \bar{1}, 10, 11, 12\dots$

Vos algorithmes d'addition et de multiplication marchent toujours, ils sont basés sur la formule  $\sum_{i=0}^p x_i 10^i$ .

Gros avantage : fini les tables de 6, 7, 8 et 9 ! C'est une idée de Cauchy, dans une note publiée dans les Comptes-Rendus de l'Académie des Sciences intitulée *Sur les moyens d'éviter les erreurs de calculs numériques*. On remarque que, pour la beauté de la symétrie, on a 11 chiffres, plus 10. Du coup, le code est devenu *redondant* : 5 peut aussi s'écrire  $\bar{5}$ . Comme chaque fois qu'on a plus de liberté, cela peut être un avantage ou un inconvénient.

Plus près de nous, le diviseur des processeurs intel modernes utilisent en interne une représentation de position en base 16 avec bien plus que 16 chiffres, signés. La motivation est que cela permet de calculer la division plus vite (toujours le précepte). Mais c'est un peu trop pointu pour être détaillé ici.

## 2.3 Coder du texte

Second exemple : le texte. On a un alphabet de 26 lettres, plus les chiffres, plus les majuscules et la ponctuation. Tout cela tient en moins de  $128 = 2^7$  caractères, donc on peut coder chaque caractère sur 7 bits. C'est ce que fait le code ASCII.

On l'a déjà dit : un bon code est un code qui permet de faire facilement les traitements utiles. Par exemple, pour le texte,

- il faut que l'ordre alphabétique se retrouve dans les codes, considérés comme des entiers ;
- idem pour les chiffres ;
- il faut qu'on puisse passer de la majuscule à la minuscule en ne changeant qu'un bit ;

ASCII a été inventé par des Américains. Il a été étendu de diverses manières pour d'autres langues, en passant à 8 bits, ce qui permet 128 caractères supplémentaires. Par exemple, il existe deux codes 8 bits populaires pour le Russe, chacun cohabitant avec les caractères de l'anglais :

- un code qui, lorsqu'on annule le bit  $8^4$ , projette chaque lettre russe sur la lettre anglolatine qui lui correspond le mieux phonétiquement;
- un code qui est dans l'ordre alphabétique de l'alphabet cyrillique.

La norme Unicode a tenté d'unifier dans un seul code, sur 16 bits, l'ensemble des écritures de la planète. En pratique, certains caractères nécessitent deux codes de 16 bits, et on peut l'utiliser à travers un encodage sur 8 bits (appelé UTF8) compatible avec le bon vieil ASCII. Bref, c'est une usine à gaz, à l'image de la complexité des écritures des langages de l'humanité. Vous en verrez peut-être des morceaux en TD.

## 2.4 Coder les images et les sons

On décompose une image en pixels, et on code pour chaque pixel sa couleur. Une couleur est la somme de trois composantes, par exemple vert, bleu, rouge. En codant chaque composante sur 8 bits, on peut déjà coder plus de couleurs que ce que l'œil humain peut distinguer.

Cela donne des quantités d'information astronomiques pour chaque image. On verra comment on peut *compresser* cette information.

Pour le son, on peut le regarder à l'oscilloscope et discréteriser la courbe obtenue. La nouveauté est qu'il faut discréteriser dans le temps (on dit *échantillonner*, en yaourt *sample*), et dans l'amplitude.

Avec tout cela, on sait même coder des vidéos.

Remarque : dans un fichier de dessin vectoriel (CorelDraw, SVG (*scalable vector graphics*) – un standard du web), on ne code pas des images, mais des figures construites à partir de primitives telles que point, droite, cercle... C'est en général plus compact, et on peut zoomer à volonté sans voir apparaître de vilains gros pixels.

Exercice : sur votre linux, ouvrez `/usr/share/icons` (cela dépend des distributions) et promenez vous dedans. Vous trouverez des icônes de différentes tailles, et des icônes en SVG (dans des répertoires appelés `scalable` ou `symbolic`). Ouvrez les (par exemple dans un navigateur web) et zoomez dessus. Comparez les tailles des fichiers avec les icônes définies par des pixels.

A ma connaissance il n'y a pas de format de fichier vectoriel pour les dessins animés. Il y a des startups qui se perdent.

## 2.5 Codes détecteurs d'erreur

On peut ajouter à chaque octet un bit de parité qui vaut 1 s'il y a un nombre impair de 1 dans l'octet, et 0 sinon. À la réception d'un tel paquet de 9 bits, on peut vérifier cette parité. Cela permet de détecter une erreur de transmission d'un bit<sup>4</sup>, mais cela ne permet pas de la corriger.

## 2.6 Codes correcteurs d'erreur

Le *code auto-correcteur de Hamming* permet de corriger un bit erroné dans une information de 4 bits  $(i_3, i_2, i_1, i_0)$ . On code cette information au moyen de 3 bits supplémentaires  $p_2, p_1$  et  $p_0$  : le codage est  $(i_3, i_2, i_1, p_2, i_0, p_1, p_0)$ , où :

- $p_0$  est le bit de parité paire du sous-mot  $(i_3, i_1, i_0, p_0)$ ,
- $p_1$  est celui du sous-mot  $(i_3, i_2, i_0, p_1)$ , et
- $p_2$  est celui de  $(i_3, i_2, i_1, p_2)$ .

À la réception de  $(i_3, i_2, i_1, p_2, i_0, p_1, p_0)$ , on vérifie ces trois parités, et on définit  $t_k$  par :  $t_k = 0$  si  $p_k$  est correct et  $t_k = 1$  sinon.

Par magie,  $(t_2, t_1, t_0)$  est alors le rang dans  $(i_3, i_2, i_1, p_2, i_0, p_1, p_0)$  du bit erroné, écrit en binaire : 000 si l'information est correcte, 001 si  $p_0$  est erroné, 010 si  $p_1$  est erroné, 011 si  $i_0$  est erroné, 100 si  $p_2$  est erroné, 101 si  $i_1$  est erroné, 110 si  $i_2$  est erroné ou 111 si  $i_3$  est erroné.

On parle ici du code  $(7, 4)$  : 7 bits transmis pour 4 bits d'information effective. On peut généraliser ces idées.

Il y a bien sûr des limitations : le code de Hamming ne permet de corriger qu'un bit d'erreur, pas deux bits simultanés. Et cela coûte presque autant que ce qu'on veut corriger.

Les codes correcteurs d'erreur sont utilisés en particulier dans

4. Cela arrivait aux temps héroïques de l'internet : votre message pouvait passer par un ordinateur anglo-saxon ou configuré comme tel, qui ne ressentait pas le besoin de transmettre plus de 7 bits par octet. On en trouve par exemple des traces dans le protocole *ftp file transfer protocol* qui offre toujours deux modes, "ASCII" et "binary".

5. ... ou d'ajouter une erreur : jadis les anglo-saxons utilisaient volontiers le bit 7 des octets, qui ne leur sert à rien, pour y mettre un bit de parité. Forcément, nos accents se perdaient.

- Les barrettes de mémoire avec ECC (*error correcting code*) dans le nom, ce qui explique qu'elles coûtent plus cher.
- La transmission de données, autrefois sur fil, actuellement par radio.

## 2.7 Compression d'information

Notions : compression avec perte ou sans perte.

Algos : on espère que vous en verrez quelques uns en cours d'algo.

## 2.8 Les QR-codes

Ah mais c'est que c'est très compliqué.



# Chapitre 3

## Transformer l'information : circuits combinatoires

On a déjà mentionné la préhistoire analogique. Remarquez que le calcul analogique à Supelec utilisait déjà l'approche hiérarchique : il fallait se ramener à des amplis ops.

Mais désormais on manipule de l'information sous sa plus simple expression : codée en binaire.

Et pour transformer de l'information binaire, on va utiliser les outils offerts par l'algèbre booléenne.

Vérifiez que vous comprenez la différence entre *algèbre*, *calcul*, *expression* et *circuit* booléens.

- L'*algèbre* booléenne va définir un certain nombre d'*opérations* et de *propriétés*.
- On utilisera les *opérations* de l'*algèbre* pour construire des *expressions* booléennes, et on utilisera les propriétés de l'*algèbre* pour manipuler ces expressions.
- La définition la plus simple d'une *fonction* booléenne sa table de vérité. Elle donne, pour chaque valeur possible des entrées, les valeurs de sortie correspondantes.
- Mais on peut également définir une *fonction* booléenne par une *expression* booléenne. Pour une fonction donnée, il existe une infinité d'*expressions* booléennes : on passe de l'une à l'autre par des identités de l'*algèbre*, comme  $\overline{a \vee b} = \overline{a} \wedge \overline{b}$ .
- Dans une technologie donnée, on sait implémenter certaines opérations de l'*algèbre* par des *opérateurs*, qu'on appelle alors volontiers des *portes logiques*.
- En assemblant ces portes, on peut construire des *circuits logiques* qui implémentent une expression, donc une fonction booléenne donnée. C'est ce qu'on fera en 3.2, en se posant notamment des questions de coût et de performance.

### 3.1 Algèbre booléenne

#### 3.1.1 Définitions et notations

L'*algèbre* booléenne définit

- un ensemble  $\mathbb{B}$  à deux éléments, muni de
- une opération unaire involutive (la négation ou complément)
- et deux<sup>1</sup> opérations binaires de base (ET et OU, aka conjonction et disjonction, aka inf et sup) commutatives, associatives, ayant chacun un élément neutre et l'autre élément absorbant, et distributives l'une par rapport à l'autre .

On utilise des mélanges variables des notations suivantes :

- Les valeurs sont notées (vrai, faux) – ou une traduction dans la langue de votre choix –, ou  $(0, 1)$ , ou  $(\emptyset, \{\emptyset\})$ , ou  $(\top, \perp)$ .
- L'opération unaire est notée NON – ou une traduction dans la langue de votre choix –, ou  $\neg$ , ou C (complément ensembliste), ou toto.
- Les opérations binaires sont notées (OU, ET) – ou une traduction dans la langue de votre choix –, ou  $(+, .)$ , ou  $(\vee, \wedge)$ , ou  $(\cup, \cap)$ .

Exercice : écrivez vous-même les trois tables de vérité avec des 0 et des 1, puis avec des yin et des yang.

En principe l'*algèbre* booléenne est complètement définie par les trois tables de vérité de ET, OU et NON. Toutes les propriétés qu'on va lister puis utiliser en découlent. Je soupçonne que de ce point de vue l'*algèbre* booléenne est plus simple que l'*algèbre* des réels...

1. Un peu plus loin, on voit qu'une seule opération suffirait, mais c'est tout de même plus confortable avec ces trois-là.

### 3.1.2 Expression booléenne

Lorsqu'on écrit des expressions booléennes, on peut utiliser des variables booléennes, les deux constantes, et des parenthèses. L'opérateur unaire est prioritaire sur les deux opérateurs binaires, qui ont une priorité équivalente :  $\neg a \vee b == (\neg a) \vee b$ .

Je déconseille la notation utilisant  $(+,. )$ . Elle présente l'intérêt d'économiser des parenthèses, puisqu'on adopte alors la priorité usuelle de  $.$  sur  $+$ . En revanche, l'arithmétique entière ou réelle a câblé dans votre cerveau des intuitions avec ces opérateurs qui seront fausses en booléen. Par exemple, les deux opérateurs booléens distribuent l'un sur l'autre. Vous serez familier de  $(a + b).c = ac + bc$ , mais vous oubliez que  $ab + c = (a + c).(b + c)$  (démonstration : comme toujours pour prouver des identités booléennes, considérer  $c = 0$  puis  $c = 1$ ). D'autres pièges vous attendent si vous utilisez cette notation. Vous voilà prévenus.

La notation à base de surlignage pour la négation est pratique, car elle permet d'économiser la plupart des parenthèses sans ambiguïté. Exemple :

$$x \wedge y = \overline{x} \vee \overline{y}$$

### 3.1.3 Dualité

L'algèbre booléenne est parfaitement symétrique : pour toute propriété, il existe une autre propriété déduite en échangeant 1 avec 0 et  $\vee$  avec  $\wedge$ . Cette idée pourra souvent vous simplifier la vie. Encore une fois, elle est particulièrement contre-intuitive avec la notation  $(+,. )$  à cause des priorités héritées de l'algèbre sur les réels.

### 3.1.4 Quelques propriétés

$0 \wedge x = 0$	$1 \vee x = 1$	(élément absorbant)
$1 \wedge x = x$	$0 \vee x = x$	(élément neutre)
$\overline{x \wedge y} = \overline{x} \vee \overline{y}$	$\overline{x \vee y} = \overline{x} \wedge \overline{y}$	(lois de De Morgan)

### 3.1.5 Universalité

On peut ramener nos trois opérations booléennes à une seule, le non-et, par application des règles suivantes :

- Négation :  $\overline{x} = \overline{x \wedge x}$
- Ou :  $x \vee y = \overline{\overline{x} \wedge \overline{y}} = \overline{\overline{x} \wedge \overline{x} \wedge \overline{y} \wedge \overline{y}}$
- Et :  $x \wedge y = \overline{\overline{x} \wedge \overline{y}} = \overline{\overline{x} \wedge \overline{y} \wedge \overline{x} \wedge \overline{y}}$

On dira que l'opération non-et est universelle. Par dualité, non-ou aussi. Cela n'est pas sans nous interpeler profondément dans notre relation à la transcendance de l'Univers, croyez-vous ? Eh bien figurez-vous que les portes de base de notre technologie CMOS seront justement NON-ET et NON-OU. Nous voilà convaincus qu'elles ne sont pas plus mauvaises que ET et OU. Plus précisément, non seulement elles forment un ensemble universel, mais en plus, pour implémenter une fonction donnée, il faudra un nombre équivalent de portes, qu'on se ramène à des NON-ET/NON-OU ou bien à des ET/OU.

### 3.1.6 Fonctions booléennes

Une fonction booléenne est une fonction d'un vecteur de bits vers un vecteur de bits :

$$f : \mathbb{B}^n \rightarrow \mathbb{B}^m$$

En général (pas toujours) on considérera une telle fonction comme un vecteur de fonctions simples  $f_i : \mathbb{B}^n \rightarrow \mathbb{B}$ , et on étudiera séparément chaque  $f_i$ .

On peut définir une fonction booléenne de diverses manières :

- par extension (en donnant sa table de vérité) – exemple une table de sinus,
- par intention (en donnant des propriétés qu'elle doit vérifier, l'unicité n'est alors pas garantie) – exemple la somme de deux chiffres BCD, voir ci-dessous,
- par une expression algébrique – exemple, la fonction NON-ET.

On peut donner une définition d'une fonction par sa table, mais avec des "don't care" pour les sorties. Exemple : la fonction d'addition de deux chiffres BCD (pour *binary-coded decimal*). Il s'agit des chiffres décimaux de 0 à 9 codés en binaire sur 4 bits. Les codes entre 10 et 15 n'étant jamais utilisés, on se fiche (don't care) de ce que fait la fonction dans ce cas. Formellement, cela revient à donner une définition intentionnelle de la fonction en question, et plusieurs fonctions booléennes différentes feront l'affaire.

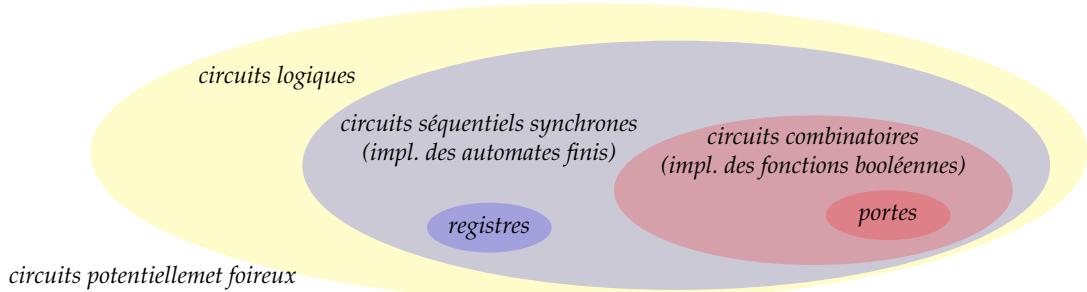


FIGURE 3.1 – Les principales classes de circuits

## 3.2 Circuits logiques et circuits combinatoires

Nous allons à présent formaliser la construction de circuits implémentant les fonctions booléennes. De tels circuits s'appellent des circuits combinatoires. C'est un cas particulier de la classe plus générale des circuits, dont une taxonomie est donnée sur la figure 3.1.

La classe des circuits combinatoires, comme la classe des circuits séquentiels synchrones que nous verrons dans les chapitres suivants, sont bien fondées sur de jolis formalismes (respectivement les fonctions booléennes, et les automates finis). En dehors de ces classes, on a essentiellement des circuits foireux, mais on verra aussi de petits objets pas très propres, mais fort utiles, comme les portes de transmission, les verrous... avec lesquels on saura construire (avec précaution) des objets tout à fait propres (en particulier des circuits combinatoires et séquentiels synchrones). Mais n'anticpons pas.

### 3.2.1 Signaux logique

Définition : un *signal logique*, c'est un dispositif physique pouvant transmettre une information binaire d'un endroit à un autre.

C'est le cas particulier le plus simple : on peut considérer des signaux qui transmettent des informations non binaires. Pour construire des circuits potentiellement foireux, on définira par exemple plus tard des signaux "trois états", les trois états étant 0, 1, et "débranché".

On dessinera un signal par un trait. Cela colle avec la technologie la plus courante, dans laquelle un signal est l'abstraction d'un fil métallique. La valeur 0 est l'abstraction d'une tension nulle, et la valeur 1 est l'abstraction d'une tension égale à la tension d'alimentation. Dans ce cas, la transmission du signal se fera par un courant électrique.

La notion de signal est toutefois plus générale que l'électronique, et fait l'objet d'un énorme domaine des sciences appelé *traitement du signal*, coincé entre informatique, physique et mathématiques, et dont je me garderai de parler.

Et pour revenir au signal logique, il est l'abstraction d'une valeur booléenne indépendamment de la technologie utilisée pour la réaliser. Vous pourrez à la fin de ce cours réaliser un ordinateur complet en lego<sup>2</sup>, dans lequel un signal binaire sera implémenté par une tige pouvant prendre deux positions.

### 3.2.2 Circuits logiques

Définition : un *circuit logique*, c'est un dispositif physique destiné à manipuler des informations binaires.

Un circuit logique se présente, vu de l'extérieur, comme une boîte noire avec des signaux d'entrée et des signaux de sortie. Rappelons que le signal sert à transmettre l'information, donc les traits sont orientés. Parfois cette orientation est implicite, comme dans le cas des portes de bases que nous voyons maintenant.

### 3.2.3 Portes de base

Définition : une *porte logique* est un circuit logique implémentant une opération booléenne élémentaire. Les figures 3.2 et 3.3 montrent les dessins standardisés pour les principales portes de base.

Exercice : construisez-les en Lego.

2. En googlant un peu vous trouverez des tentatives dans ce sens

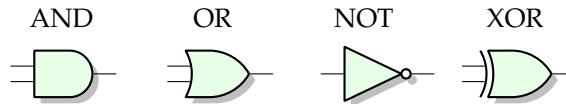


FIGURE 3.2 – Les dessins des portes ET, OU, NON et XOR (ou exclusif) en notation ANSI

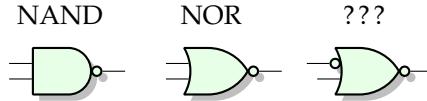


FIGURE 3.3 – En général, on peut mettre des petits ronds sur les entrées et les sorties pour dire qu'on les inverse.

Tout ensemble *universel* (au sens du 3.1.5) de portes de base fera en pratique l'affaire. Les bibliothèques de composants ne se privent pas d'offrir des portes plus complexes que les portes binaires (par exemple celle de la figure 3.4) si cela permet de gagner en performance ou en coût.



FIGURE 3.4 – Si l'opération booléenne est associative et commutative, la porte peut avoir plein d'entrée.

Tous mes dessins sont orientés de la gauche vers la droite : la porte ressemble à une flèche et induit l'orientation de ses signaux d'entrée et sortie. On peut aussi dessiner les portes à l'envers.

### 3.2.4 Circuits combinatoires

Définition : un *circuit combinatoire*, c'est un circuit logique implémentant une expression booléenne.

Le circuit combinatoire est un cas particulier de circuit logique avec une grosse restriction : *il ne possède pas de mémoire du passé*. La sortie d'un circuit combinatoire est fonction uniquement de l'entrée, pas des entrées qu'il a pu avoir dans le passé. Par exemple, une GameBoy n'est pas un circuit combinatoire.

### 3.2.5 Construction bottom-up de circuits combinatoires

Si vous considérez une expression booléenne comme un arbre (ce qu'elle est en fait, tout ce cirque avec les priorités et les parenthèses c'est pour écrire un arbre dans un langage qui se lit de gauche à droite), vous en déduisez immédiatement un circuit logique qui l'implémente, et qui est par définition un circuit combinatoire.

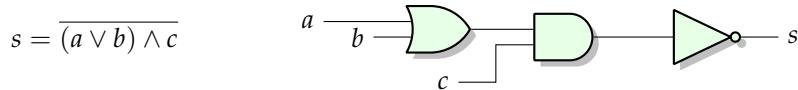


FIGURE 3.5 – Une expression donne trivialement un circuit combinatoire, si on sait la lire comme l'arbre qu'elle est.

La classe des circuits combinatoires est légèrement plus vaste que ceci, puisqu'elle permet les DAG (directed acyclic graph). Étant donné un ensemble de portes de bases (qui sont toutes des circuits combinatoires), on peut donner une définition récursive des circuits combinatoires bien formés (CCBF) comme suit :

In CCBF est

- une porte de base
- ou un signal
- ou la juxtaposition de deux CCBFs posés l'un à côté de l'autre
- ou un circuit obtenu en connectant les sorties d'un CCBF aux entrées d'un autre CCBF
- ou un circuit obtenu en connectant entre elles deux entrées d'un CCBF.

On peut montrer qu'on obtient un graphe sans cycle de portes de bases.

Ce qu'on s'interdit :

- faire des cycles, car cela permettrait des situations mal définies comme le circuit instable de la figure 3.6, ou le circuit à deux états stables de la figure 3.7 (dans quel état va-t-il être?) Les circuits sur ces deux figures appartiennent à la classe des "circuits potentiellement foireux" de la figure 3.1.
- connecter des sorties entre elles (que se passera-t-il si une sortie est 1 et l'autre 0?<sup>3</sup>)

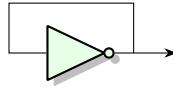


FIGURE 3.6 – *Un exemple simple de circuit pas combinatoire : le pas-stable*

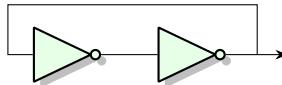


FIGURE 3.7 – *Un autre exemple simple de circuit pas combinatoire : le bistable*

Au chapitre suivant, on lèvera ces deux interdictions dans des circonstances bien maîtrisées, et on sortira du domaine des circuits combinatoires : avec des circuits à cycles, on pourra construire des oscillateurs (en mettant des retards dans notre pas-stable) et des mémoires (le bistable en est une, mais dans laquelle on ne peut pas entrer d'information à cause de la seconde interdiction, donc il faudra le bricoler un peu).

### 3.2.6 Construction top-down de circuits combinatoires

On va voir bientôt (en 3.4) comment construire un circuit combinatoire correspondant à une fonction booléenne donnée.

En gros on passe par un stade intermédiaire : on commence par déterminer une expression algébrique de la fonction, et comme illustré par la figure 3.5, il est ensuite facile de traduire cette expression sous forme de portes.

Malheureusement cette belle idée se heurte à un petit souci : il y a une infinité d'expressions booléennes correspondant à une fonction booléenne donnée, donc il faudra faire un choix. Commençons donc par nous donner les outils pour faire le bon choix.

### 3.2.7 Métriques d'un circuit combinatoire

Pour une fonction booléenne donnée, il y a une infinité dénombrable de manières de l'implémenter. Ce qui est intéressant, en architecture des ordinateurs, c'est de trouver la meilleure. Exemple : construire un ET à quatre entrées à partir de NON-OU.

Meilleure selon quel critère ? La technologie amène avec elle ses métriques de qualité : taille, vitesse, consommation d'énergie, niveau d'émission électromagnétique... Voyons les plus basiques.

#### Surface d'un circuit combinatoire

Le coût unitaire<sup>4</sup> d'un circuit VLSI en silicium est une fonction de sa surface<sup>5</sup>. La surface est elle-même en première approximation la somme des surfaces de chaque porte logique (l'approximation étant qu'on ignore la surface des fils). Bref, le nombre total de porte est une bonne mesure du coût unitaire d'un circuit.

Parenthèse technologique contingente : en technologie CMOS, si l'unité de coût est le transistor, alors l'inverseur coûte 2, les portes non-ET et non-OU à  $k$  entrées coûtent  $2k$ , et ce sont les principales portes logiques de base : un ET logique à 2 entrées est construit en assemblant un non-ET et un inverseur, pour un coût de 6 transistors. Si cela vous amuse ce sera détaillé Section 3.7

3. L'état est logiquement indéfini, et si cela vous arrive avec de l'électronique, il se passera un joli court-circuit, car rappelons que le zéro c'est la masse et le 1 c'est l'alimentation positive. Adieu les portes.

4. Le vrai coût d'un circuit est composé d'un coût fixe  $c_f$  (essentiellement le coût de conception) et d'un coût unitaire par pièce  $c_u$  : pour fabriquer  $n$  circuits il vous en coûtera  $c_f + n \times c_u$ . Pour des petites séries  $c_f$  est le coût prédominant, pour les grandes séries c'est  $c_u$ . La valeur de  $n$  pour laquelle le coût unitaire commence à dominer est de plus en plus grande avec les années.

5. Le coût unitaire est proportionnel à la surface pour les petits circuits. Pour les plus gros c'est plus compliqué : on imprime les circuits sur des galettes (wafer) de silicium d'environ 30cm de diamètre, 900cm<sup>2</sup> qu'on va découper en  $N$  circuits. Malgré tous les efforts de la communauté de recherche sur les aspirateurs de pointe, il tombe sur chaque galette statistiquement  $k$  poussières ( $k$  est petit, de l'ordre de la dizaine). Chaque poussière va ruiner le circuit sur lequel elle tombe. Si on met un énorme circuit par galette, il sera ruiné presque à coup sûr. Si par contre on met  $N >> k$  petits circuits, alors seuls  $k$  seront ruinés et on pourra vendre presque toute la galette. C'est cette équation qui est cachée derrière la formulation "une puce économiquement viable" dans ma définition de la loi de Moore.

### Délai d'un circuit combinatoire

Dans la notion la plus abstraite de circuit combinatoire, il n'y a pas plus de notion de temps que dans la notion d'expression booléenne. Par contre, toute réalisation physique d'un circuit combinatoire aura un certain *délai* de fonctionnement, noté  $\tau$ , et défini comme le temps maximum entre un changement des entrées (à l'instant  $t$ ) et la stabilisation des sorties dans l'état correspondant (à l'instant  $t + \tau$ ). Les sorties du CC physique peuvent passer par des états dits transitoires pendant l'intervalle de temps  $[t, t + \tau]$ . C'est le cas du circuit de la figure 3.8. On apprendra au chapitre 5 à construire des circuits dans lesquels on garantit que ces états transitoires sont sans conséquence.

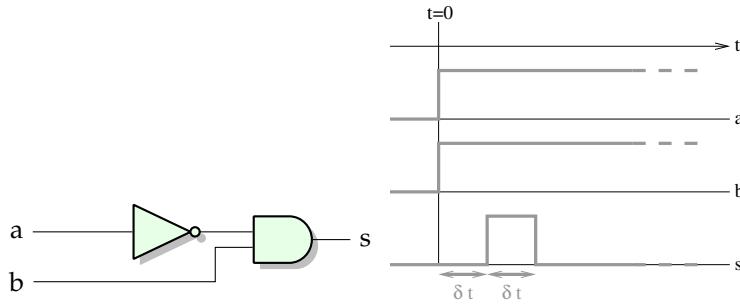


FIGURE 3.8 – Un circuit pas compliqué qui nous fait des misères transitoires

On peut définir (ou mesurer) le délai de chaque porte de base. Une fois ceci fait, on peut ramener le calcul du délai dans un CCBF à un problème de plus long chemin dans le graphe du CCBF. Ce plus long chemin est appelé *chemin critique* (en Australien : *critical path*). On aura une tendance malheureuse à identifier le chemin critique et le délai du chemin critique.

Dans la vraie vie, le délai d'une porte de base peut être différent suivant l'entrée et la sortie considérée. Dans le cas général, pour une porte de base à  $n$  entrées et  $m$  sorties, on aura une matrice  $n \times m$  de délais. Chaque entrée de la matrice est elle-même le max des délais pour toutes les configurations des autres entrées. De plus, le délai pourra être différent sur front montant et sur front descendant, on aura alors une matrice  $2n \times m$ . Naturellement le calcul du délai d'un CCBF devient un peu plus compliqué, mais cela reste parfaitement automatisable, et même vous sauriez faire.

Enfin, les authentiques gourous du circuit sauront retailler les transistors pour obtenir les délais qu'ils veulent. Ce sera survolé superficiellement à la section 3.7

### Consommation d'un circuit logique

Il est clair que la consommation d'énergie sera proportionnelle au nombre de portes.

En électronique, pour chaque porte, il y a

- une consommation statique (elle chauffe un peu juste en étant alimentée)
- et une consommation dynamique : à chaque passage de 0 à 1 puis de 1 à 0 un peu d'énergie est dissipée (un peu de courant passe de l'alimentation positive à la masse).

La consommation dynamique dépend de l'*activité* de la porte, qui est une mesure du nombre de cycles de 0 à 1 à 0 qu'elle fait par secondes. Pour les circuits synchrones (qu'on verra plus tard) on ramène l'activité à un nombre entre 0 et 1 qui donne la probabilité que la sortie de la porte réalise une transition dans un cycle d'horloge.

## 3.3 Quelques constructions algorithmiques de circuits combinatoires

### 3.3.1 Les portes de base à deux entrées

Exercice : énumérons toutes les fonctions logiques à 2 entrées et une sortie, et donnons leur à chacune un petit nom.

### 3.3.2 Le multiplexeur, abstraction du si-alors-sinon

Dans un train électrique, on a une brique de base appelée *aiguillage*, qui en fonction d'une information binaire de *direction* (entrée du bas sur la figure), transmet l'information (le train) soit selon un chemin, soit selon l'autre. Cette brique (non orientée dans le cas de l'aiguillage) est représentée figure 3.9

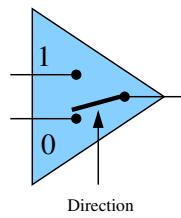


FIGURE 3.9 – L'aiguillage

### 3.3.3 Le multiplexeur d'ordre $k$

En assemblant  $1 + 2 + 4 + \dots + 2^{k-1} = 2^k - 1$  aiguillages, on sait construire une gare de triage complète (Figure 3.10). Observez que l'adresse de la case d'arrivée est donnée en binaire sur les fils de direction.

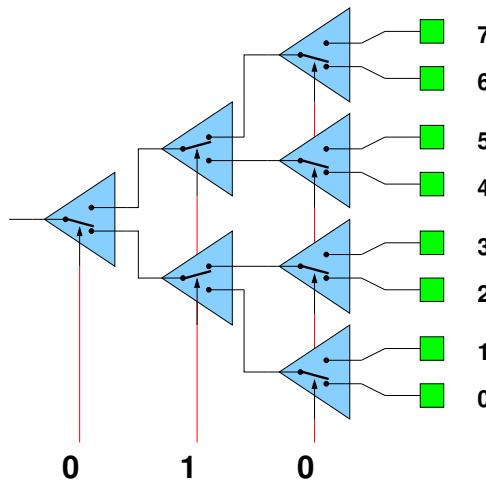


FIGURE 3.10 – Une gare de triage permet de construire une mémoire adressable

L'aiguillage, en technologie train électrique, marche dans les deux sens.

Si on parle de circuits logiques, on distinguera deux familles de circuit qui ont *la même construction sous forme d'arbre binaire* : le **multiplexeur** ( $2^n$  en 1) et le **démultiplexeur** (1 vers  $2^n$ ).

Chacun a une entrée de sélection d'adresse (en général appelée *select* ou *sel*) de  $n$  bits. De plus,

- le multiplexeur (ou MUX) a aussi  $2^n$  entrées, et une sortie. Il transmet vers sa sortie celle des  $2^n$  entrées dont le numéro est donné en binaire sur l'entrée *sel*.
- le démultiplexeur (ou DEMUX) a aussi une entrée et  $2^n$  sorties. Il transmet son entrée vers celle des  $2^n$  sorties dont le numéro est donné en binaire sur l'entrée *sel*. Les autres sorties valent 0.

La figure 3.11 vous donne le dessin standard d'un multiplexeur 2 vers 1, à partir duquel vous saurez construire tous les autres (en suivant la construction de la figure 3.10. Le démultiplexeur, c'est le même dessin en retournant certaines flèches.

Exercices :

- Dessinez le DEMUX 1 vers 2 à côté de la figure 3.11.
- Construisez le MUX avec des portes ET, OU et NON.
- Construisez dans la figure 3.12 le MUX 8 vers 1 à partir du MUX 2 vers 1.

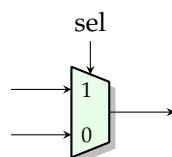


FIGURE 3.11 – Le multiplexeur 2 vers 1, une porte logique à 3 entrées et une sortie. Dessinez à côté le démultiplexeur 1 vers 2.

Avec des multiplexeurs on fabrique aussi le décaleur (*barrel shifter*).

La figure 3.12 propose une autre construction du démultiplexeur, plus naïve. Peste, mais ce n'est pas du tout la construction récursive précédente?!? C'est une illustration que pour chaque fonction booléenne, il existe une infinité d'expressions, donc une infinité de circuits, l'implémentant.

Demandons-nous si cette construction à base de portes AND est moins chère.

FIGURE 3.12 – Dessinez vous-même un multiplexeur 8 vers 1. Indice : la structure générale est celle de la figure 3.10

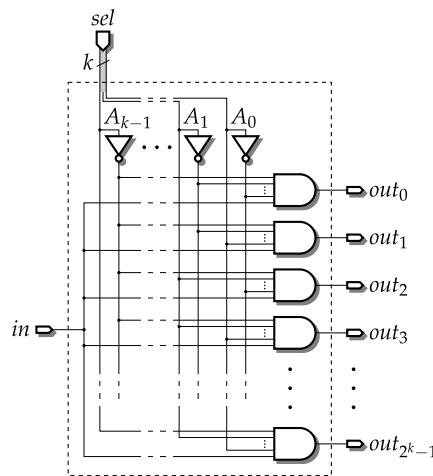


FIGURE 3.13 – Une autre architecture de démultiplexeur 1 vers  $2^k$ .

- Combien de portes AND ?  $2^k$ . L'arbre binaire de MUX compte au total  $2^k - 1$  MUX2 : il gagne.
  - Oui mais un AND à  $k + 1$  entrées est moins cher qu'un MUX2, pensez-vous ? Eh bien cela dépend sans doute de  $k$  : Si  $k$  est grand on va devoir l'implémenter lui-même comme un arbre binaire.
- Mais c'est bien d'avoir le choix entre plusieurs architectures pour pouvoir se poser ces questions.

### 3.3.4 Pour le calcul sur les entiers

La base du calcul sur les entiers c'est l'addition : construisons un additionneur.

On part de l'algo d'addition en décimal, qu'on généralise un peu, et on le formalise mathématiquement :

Soit  $\beta$  un entier (la base, qui vaut 10 pour votre petite sœur mais vaudra 2 en binaire). Soit la fonction *table d'addition* TA :

$$\begin{aligned} \{0,1\} \times \{0..\beta-1\} \times \{0..\beta-1\} &\rightarrow \{0..1\} \times \{0..\beta-1\} \\ (r, x, y) &\mapsto (r', s) \\ \text{tq } \beta r' + s &= r + x + y \end{aligned}$$

- cette fonction prend une "retenue entrante" (0 ou 1) et deux chiffres, et retourne leur somme ;
- Cette somme est comprise entre 0 et  $2\beta - 1$  (par exemple en décimal entre 0 et  $9 + 9 + 1 = 19$ ). Elle s'écrit donc en base  $\beta$  sur deux chiffres :
- Le chiffre de poids faible de la somme est appelé *somme modulo la base* ou *juste somme*

— Le chiffre de poids fort de la somme est appelé *retenue* et vaut 0 ou 1 quelle que soit la base. L'algorithme d'addition de deux nombres de  $n$  chiffres est alors

```

1:  $c_{-1} = 0$ 
2: for  $i = 0$  to  $n$  do
3:    $(c_i, s_i) = \text{TA}(c_{i-1}, x_i, y_i)$ 
4: end for

```

On peut le dérouler de plusieurs manière en matériel : algo séquentiel (utilisation du temps) ou parallèle (utilisation de l'espace). Cela aussi est une idée générale.

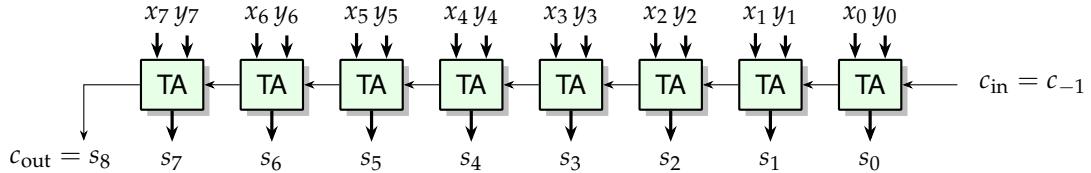


FIGURE 3.14 – L'algo d'addition déroulé dans l'espace. TA c'est "table d'addition".

Pour construire l'additionneur binaire il suffit de considérer  $\beta = 2$ . Mais alors les trois entrées deviennent symmétriques (toutes entre 0 et 1). La porte obtenue s'appelle le *full adder*<sup>6</sup> et sa table de vérité est donnée figure 3.15. La fonction booléenne correspondante à 3 entrées et deux sorties : elle écrit en binaire sur 2 bits la somme de ses trois entrées (qui vaut donc entre 0 et 3).

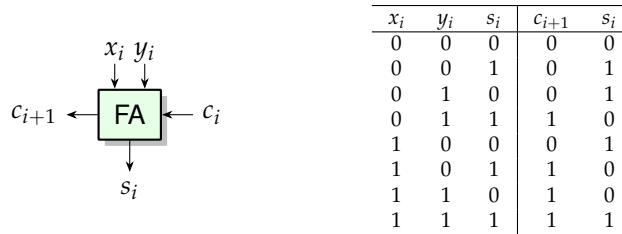


FIGURE 3.15 – Le "Full Adder". Exercice : construisez-le à partir de portes logiques.

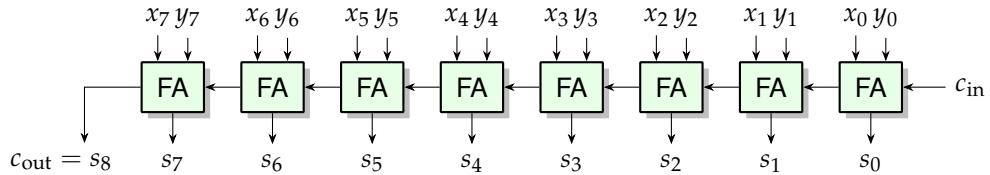


FIGURE 3.16 – Additionneur binaire à propagation de retenue.

## Complément à 2

Pour calculer une soustraction, on peut construire un soustracteur. Mais il ressemble à un additionneur : il propage des retenues, etc. On aimerait bien partager le même matériel entre addition et soustraction.

Pour cela, on utilise la notation en complément à 2, introduite à la section 2.2.4

Rappels : Pour calculer 745-169, on peut calculer 745+1000-169-1000. Le -1000 final, c'est juste d'ignorer la retenue sortante. Quant au +1000, on l'écrit +999+1, et notre soustraction devient :

745 + (999-169) +1. La soustraction 999-169 se fait chiffre à chiffre, sans propagation de retenue. Quant au 1 qu'on ajoute en plus de l'addition normale, c'est juste une retenue sur la première colonne, qui justement boudait de ne point en avoir.

Tout cela se transpose en binaire (en remplaçant 999 par 111, bien sûr), et l'on obtient l'architecture d'additionneur/soustracteur de la figure 3.17. Le complément, en binaire, est juste la négation (pour un booléen  $x$ ,  $\bar{x} = 1 - x$ ).

Re-rappel : par la magie du modulo, cet additionneur/soustracteur fonctionne pour des entiers positifs comme pour des entiers négatifs.

6. C'est idiot comme nom pour un bit d'additionneur, il est pas complet du tout comparé à un additionneur 64 bits, mais c'est comme ça. Le nom est pas opposition à une autre fonction appellée *half adder* qui fait la moitié du boulot puisqu'elle n'a que deux bits d'entrée.

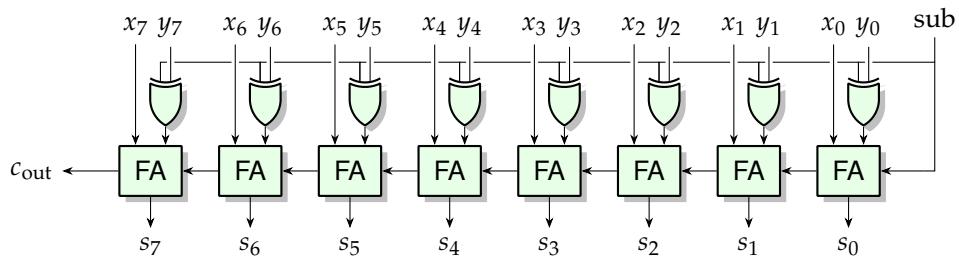


FIGURE 3.17 – Additionneur/soustracteur en complément à 2.

### Dépassement de capacité en complément à 2

Soient  $m = (c_m)_{\bar{2}}$  et  $n = (c_n)_{\bar{2}}$  en complément à 2 sur  $p$  bits.

On dit qu'il y a dépassement de capacité dans une opération en complément à 2 sur  $p$  bits lorsque le résultat de l'opération ne peut pas être représenté sous cette même forme (il est soit trop petit, soit trop grand).

- En l'absence de dépassement de capacité, le codage de  $m + n$  en complément à 2 sur  $p$  bits est le codage de l'entier naturel  $(c_m + c_n) \bmod 2^p$ . Si  $m$  et  $n$  sont de même signe, il y dépassement ssi le signe du résultat calculé diffère du signe des opérandes ; s'ils sont de signes opposés, aucun dépassement n'est possible.
- En l'absence de dépassement de capacité, le codage de  $-n$  en complément à 2 sur  $p$  bits est le même que celui de l'entier naturel  $(\bar{c}_{p-1}\bar{c}_{p-2}\dots\bar{c}_1\bar{c}_0)_2 + 1 \bmod 2^p$ . Il y dépassement ssi le signe du résultat calculé est le même que celui de l'opérande.

### 3.3.5 Conclusion

Les gares de triage et les additionneurs sont des circuits construits par une approche algorithmique. Pour les gros circuits, c'est l'approche à privilégier : il est toujours mieux de faire marcher son intelligence, de diviser pour régner, etc.

Mais il existe aussi une approche automatique que nous allons voir maintenant. Vous pouvez l'appliquer à la main, ou la programmer. On verra qu'elle ne s'applique en pratique que pour de "petites" fonctions booléennes (par exemple pour construire un *full adder* ou un multiplexeur) à partir de portes logiques.

## 3.4 D'une fonction booléenne à un circuit combinatoire

### 3.4.1 Par les formes canoniques

Expansion de Boole : soit  $f$  une fonction booléenne à  $n$  entrées, alors

$$f(x_1, x_2, \dots, x_n) = x_1 \cdot f(1, x_2, \dots, x_n) + \bar{x}_1 \cdot f(0, x_2, \dots, x_n)$$

Démonstration : considérer  $x_1 = 0$  puis  $x_1 = 1$ .

Exercice : donner l'expansion de Boole duale de celle-ci. J'ai perfidement utilisé la notation que j'ai dit que je n'aimais pas.

On appelle parfois un *monôme canonique* un ET de toutes les variables apparaissant chacune soit sous forme directe soit sous forme complémentée. On peut appeler *monal canonique* le dual du précédent.

Les deux variantes de l'expansion de Boole, appliquées récursivement, permettent d'obtenir automatiquement, à partir d'une fonction booléenne donnée par sa table de vérité, une expression booléenne de cette fonction sous forme dite canonique.

- Forme canonique disjonctive : OU (ou disjonction, ou somme mais je vous ai dit que je n'aimais pas) de monomes
- Forme canonique conjonctive : ET (ou conjonction) de monals (monaux?)<sup>7</sup>

Exemple : fonction majorité à trois entrées.

Quelle forme préférer ? Considérons la forme conjonctive. Naturellement, on simplifie tous les monômes correspondant à une valeur 0 de la fonction. Donc cette forme sera préférée pour une fonction ayant une majorité de 0 dans sa table de vérité. Inversement, s'il y a une majorité de 1, la forme disjonctive aura moins de monaux rescapés que la forme conjonctive de monômes.

7. C'est sans doute pour cela qu'on préfère habituellement la disjonctive...

Exemple : la fonction majorité à 3 entrées, qui renvoie 1 si au moins deux des trois entrées sont à 1 et 0 sinon.

**Simplification des formes canoniques** Supposons pour anticiper qu'on veut se ramener à un nombre le plus petit possible de NON-ET et de NON-OU à moins de 4 entrées.

Il faut parenthésier notre forme canonique. Il y a un très grand nombre de parenthésages possibles (un amateur de combinatoire me dira combien?).

Ensuite on peut appliquer des règles de réécriture. Lesquelles, et dans quel ordre ?

On peut privilégier la minimisation du nombre de portes (surface) ou la minimisation de la profondeur de l'arbre (délai). Dans les deux cas on sait comparer deux expressions pour choisir la meilleure. Par contre, pour obtenir la meilleure expression pour une fonction booléenne donnée, on ne sait pas faire tellement mieux que toutes les essayer (en commençant par essayer tous les parenthésages) : le problème est NP-fâcheux.

Heureusement il y a de bonnes heuristiques. Voyons maintenant l'une d'elles.

### 3.4.2 Par les arbres de décision binaire

Un *arbre de décision binaire* (en néozélandais BDD pour *binary decision diagram*) est simplement un arbre d'évaluation de la fonction en considérant la valeur de chacune de ses variables l'une après l'autre. Un exemple est donné par la figure 3.18. Pour connaître la valeur de  $f(x_1 x_2 x_3)$ , on part de la racine, et à chaque nœud de profondeur  $i$  on part dans le sous-arbre gauche si  $x_i = 0$ , et dans le sous-arbre droite si  $x_i = 1$ . Arrivé à une feuille, l'étiquette de la feuille donne la valeur de la fonction.

Si vous voulez, c'est comme un arbre de détermination pour reconnaître les papillons ou les fleurs des champs, si vous avez pratiqué ces choses-là.

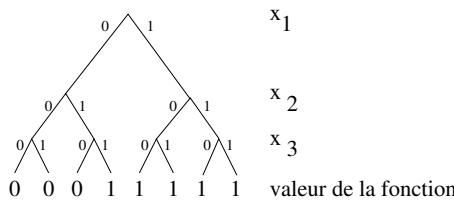


FIGURE 3.18 – Un arbre de décision binaire

La construction du BDD est triviale en partant de la table de vérité de la fonction. Faisons le au tableau.

Attention, l'arbre obtenu dépend de l'ordre des variables, c'est pourquoi on utilise souvent le terme OBDD (pour *ordered*).

L'OBDD a autant de feuilles que la table avait d'entrée, donc c'est une structure plutôt plus lourde que la table de départ. Son intérêt est de pouvoir ensuite être réduite pour obtenir une représentation plus compacte de la fonction. On applique pour cela les transformations suivantes :

- Si deux sous-arbres sont identiques, on les fusionne
  - Si un nœud a deux fils identiques, on le court-circuite et on le supprime.
- On obtient un arbre réduit (ROBDD), comme sur la figure 3.18, qu'il est ensuite facile de traduire en un circuit (exercice).

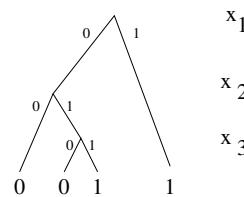


FIGURE 3.19 – Un arbre de décision binaire réduit

Tout ce processus est déterministe, et requiert au plus  $O(2^n)$  en temps et en espace, donc pas plus que ce que demande la construction de la table de vérité. Donc il est tractable pour des  $n$  assez grand (pour des fonctions jusqu'à 20 bits sans problème). C'est trop beau. Cela ne va pas durer.

En effet, il reste un problème : le facteur de réduction final dépend de l'ordre des variables choisi pour la construction de l'arbre. Or il y a pour  $n$  variables  $n!$  permutations de ces variables. Laquelle d'entre

elle va donner les simplifications les plus avantageuses ? Eh bien avant d'essayer, on ne sait pas. Nous voilà à nouveau contraints de bricoler des heuristiques, et tous les ans il y a des articles sur le sujet.

Remarque : l'OBDD, et même le ROBDD sont, pour un ordre des variables donné, des représentations canoniques. Elles sont aussi utilisées pour prouver l'équivalence de circuits (par exemple développé sur FPGA puis "recompilé" vers un ASIC).

## 3.5 Application : construction des circuits arithmétiques de base

### 3.5.1 Addition/soustraction binaire

Ah ben on a déjà tout vu en 3.3.4.

### 3.5.2 Multiplication binaire

En TD peut-être. Montrons juste qu'on sait poser la multiplication en binaire : cela suffit à définir les différentes fonctions booléennes nécessaires.

Evidemment c'est encore une construction algorithmique de circuit.

### 3.5.3 Division binaire

Même texte que pour la multiplication

## 3.6 Conclusion

On a une technique universelle pour transformer n'importe quelle fonction booléenne, donnée par sa table de vérité, en un circuit qui l'implémente. Cette technique s'adaptera facilement à une *bibliothèque de portes de base* pour une technologie donnée.

Toutefois, on se trouve confronté à des questions d'optimisation. La complexité des algorithmes d'optimisation, dans le cas général, est exponentielle en le nombre d'entrées de la fonction.

Pour cette raison, faire de l'algorithmique intelligemment donne en général de meilleurs résultats que la technique universelle. C'est ce que l'on a vu pour construire les multiplexeurs et les opérateurs de calcul.

## 3.7 Annexe technologique contingente : les circuits CMOS

On représente le 1 logique par une tension proche de la tension d'alimentation (ou Vdd, ou la borne + de la pile), et le 0 logique par une tension proche de la masse (ou Vss, ou la borne - de la pile). Dans la phrase précédente, "proche de" vous inquiète car ce n'est pas très précis. Il ne faut pas, on va utiliser une électronique construite pour travailler en mode "saturé", c'est à dire non-linéaire dans le sens qui pousse soit vers la masse, soit vers Vdd. Si vous êtes toujours inquiet, il y a un exemple ci-dessous, cf figure 3.21.

### 3.7.1 Transistors et processus de fabrication

Règles du jeu du CMOS :

- Un transistor est, en gros, un interrupteur, commandé par un fil de commande qui arrive sur sa "grille" (gate en Texan) : voir la figure 3.20 ci-dessous.
  - On a deux types de transistor : le transistor normal (qui s'appelle NMOS pour N-type Metal Oxide Silicium), qui est passant quand on met la grille à 1 et ouvert sinon, et le transistor-dessiné-avec-un-rond-sur-la-grille (qu'on appelle PMOS mais je refuse de m'en souvenir), que c'est le contraire.
  - Pour de sordides raisons électroniques (qui en plus font l'apologie du dopage alors je les traite par le mépris), le NMOS marche bien (i.e. conduit bien le courant) quand il est relié à la masse (ou à un autre NMOS relié à la masse), alors que le PMOS c'est le contraire : il aime bien être relié à l'alimentation positive.
  - du coup il faut les assembler de telle manière que les 1 logique produits par un circuit viennent à travers des PMOS, alors que les 0 logiques doivent être produits par des NMOS.
- Voyons quelques exemples de cette dernière règle.

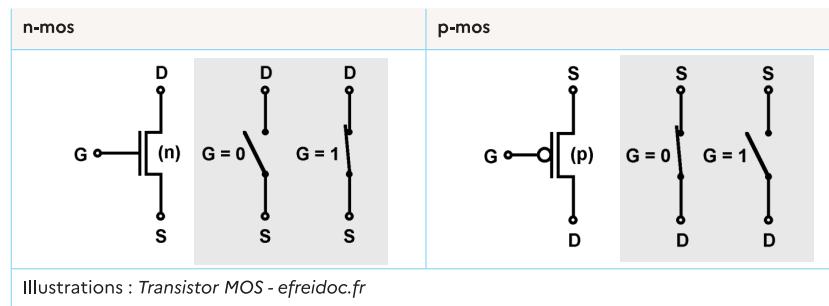


FIGURE 3.20 – Description fonctionnelle des transistors MOS sous forme d'interrupteurs

### 3.7.2 Portes de base

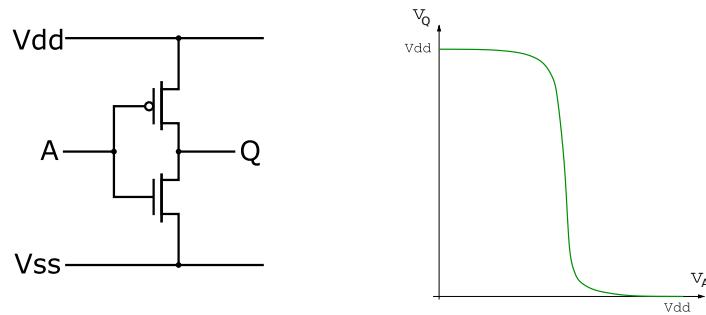


FIGURE 3.21 – A gauche un dessin d'inverseur pompé sur Wikipedia, et à droite sa caractéristique. Vdd cela veut dire +5V (pour une certaine valeur de 5) et Vss cela veut dire la masse.

La figure 3.21 montre comment on construit un inverseur CMOS avec deux transistors. La caractéristique de transfert en tension de l'inverseur montre qu'il "nettoie" un zéro de mauvaise qualité (loin de 0V) en un 1 de bonne qualité (proche de Vdd). C'est cela qui rend l'électronique numérique si confortable.

Sur la figure 3.22, on a une porte de base du CMOS, laquelle ? Observez le C pour "complémentaire" de CMOS : il y a autant de transistors P que de transistors N. Lorsque les P sont en série, les N sont en parallèle (et inversement). C'est beau comme du booléen, d'ailleurs la loi de De Morgan se cache dans cette construction.

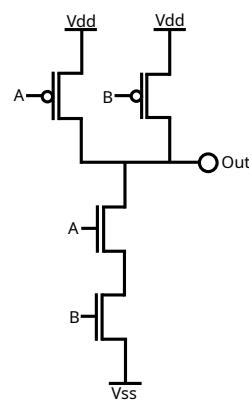


FIGURE 3.22 – A votre avis, c'est quoi comme porte ?

Et au passage, dans cette technologie, brancher deux sorties ensembles, c'est non seulement indéfini, c'est mal. En effet si les deux sorties branchées ensembles ne sont pas d'accord sur la valeur logique, on obtient un joli court-circuit de l'alimentation à la masse, qui fait assez vite fondre les transistors sur le chemin...

On peut réaliser directement en CMOS n'importe quelle négation de fonction croissante des entrées. Démonstration : on relie la sortie à la masse par un assemblage arbitraire série/parallèle de transistors. On lit la fonction en question comme la négation d'une formule à base de AND et OR des entrées. Ensuite on construit l'assemblage complémentaire pour relier la sortie à Vdd dans les autres cas.

Exemple de porte qui n'est pas une fonction croissante des entrées : le XOR.

### 3.7.3 Vitesse, surface et consommation

Sans entrer dans les détails, voici ce qui se passe dans un circuit composé de transistors branchés entre eux :

- La grille d'un transistor (gate) se comporte comme un condensateur : il faut charger et décharger ce condensateur pour le faire commuter. Les entrées d'une porte CMOS de base (inverseur, nand ou nor) sont toutes connectées à des grilles : pour faire basculer ces portes, il faut charger ces condensateurs.
- Par conséquent, lorsqu'on connecte une sortie à  $n$  entrées (fan-out de  $n$ ), on divise le courant qu'est capable de produire la sortie entre les  $n$  entrées, et on multiplie donc par  $n$  le temps qu'il faut pour charger ces  $n$  entrées (donc faire basculer les transistors correspondants, donc faire passer l'information). Typiquement, dans un circuit, on accepte un fanout de 4 maximum, au-delà c'est mal.
- La connexion entre la source et le drain du transistor se comporte comme une résistance, assez faible quand le transistor est passant, et élevée quand il est bloqué.
- Plus la grille est large, plus elle présente une capacité en entrée élevée, c'est-à-dire plus il faut du courant pour la faire basculer, mais plus elle peut laisser passer de courant entre la source et le drain.
- Autrement dit, un transistor plus large est plus rapide du point de vue de sa sortie (il peut passer plus de courant pour charger plus vite les transistors suivants) mais plus lent du point de vue de son entrée (il faut plus de temps pour le charger pour le faire basculer).

Tout cela se mord la queue. Heureusement, une chouette technique (Logical Effort) permet de modéliser la vitesse d'un circuit CMOS en fonction de la taille des transistors, et ensuite d'optimiser cette vitesse en dérivant ce qu'il faut pour chercher quand la dérivée s'annule.

# Chapitre 4

## Memoriser l'information

On peut dérouler la somme de  $n$  nombres dans l'espace (arbre binaire d'additionneurs) ou dans le temps, mais alors il faut une mémoire de la somme courante ("accumulateur").

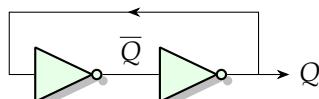
Cela vaut à toutes les échelles, par exemple dans une addition on a des "retenues" : ce qu'on retient. On peut en principe construire un additionneur  $n$  bits à partir d'un seul *full adder*, à condition

- de savoir mémoriser la retenue entre deux calculs, et
- de lui présenter dans le bon ordre les bits des nombres à additionner.

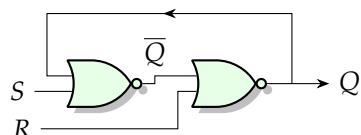
Ces deux exemples nous donnent l'intuition que, au delà de la mémorisation à long terme, **la capacité de mémoriser des bits est fondamentale pour construire des circuits fonctionnant dans le temps** (l'objet des deux chapitres suivants).

### 4.1 Mémoriser un bit

L'archétype de l'élément de mémorisation est le bistable que nous avons déjà vu :



Il possède deux états stables, mais... comment le faire passer d'un état à l'autre? La solution la plus simple est de remplacer les deux inverseurs par des portes non-ou :



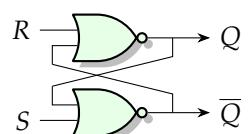
Si on étudie son fonctionnement,

- $(R, S) = (0, 0)$  : on retombe bien sur le bistable;
- $(R, S) = (0, 1)$  impose l'état stable dans lequel  $Q = 1$ . Si ensuite  $S$  redescend à 0 c'est cet état qui restera dans le bistable;
- $(R, S) = (1, 0)$  impose l'état stable dans lequel  $Q = 0$ . Si ensuite  $R$  redescend à 0 c'est cet état qui restera dans le bistable;
- $(R, S) = (1, 1)$  impose un état  $(Q, Q\bar{}) = (0, 0)$  qui n'est pas un des deux états stables puisque  $\bar{Q} = Q$ . Si on passe directement de cette entrée à l'entrée  $(0, 0)$  l'état du bistable sera indéterminé. A éviter donc.

Nous venons d'inventer la bascule RS (R pour *reset*, S pour *set*), qu'on trouve souvent sous la forme suivante : vérifiez que c'est bien une version plus jolie de la même chose. Au fait, il y a une variante avec deux portes non-et.

$R$	$S$	$Q$	$\bar{Q}$
1	0	0	1
0	1	1	0
1	1	0	0

(à éviter)



On va se dépêcher de l'oublier, parce qu'à l'usage cette interface set/reset n'est pas pratique du tout. Commençons par définir une interface plus simple.

### 4.1.1 Le registre 1-bit, ou flip-flop

L'interface minimale à un élément de mémorisation est le registre sur front, ou Flip-Flop<sup>1</sup> dessiné Figure 4.1. On verra l'explication de ce nom rigolo un peu plus bas (et aucun rapport avec les sandales de plage).

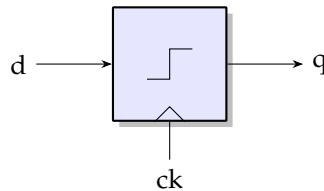


FIGURE 4.1 – Symbole pour un registre sur front montant

Il y a mémorisation de l'entrée d (pour donnée) au front montant (changement de valeur de 0 à 1) de l'entrée ck (pour clock). La dernière valeur mémorisée reste présentée sur la sortie q jusqu'au prochain front montant.

L'entrée ck sert à indiquer un instant, celui auquel on veut mémoriser. Toute mémorisation nécessite une telle entrée spécifiant un instant. Le triangle sur l'entrée ck indique que ce signal porte une information temporelle (un instant), pas une donnée.

*Une information d'instant sera toujours donnée par un changement de valeur, qu'on appelle un front (soit montant, soit descendant).*

C'est pour cela que j'ai dessiné le front sur le registre. On peut évidemment aussi construire des registres sur front descendant.

Et là, la question que tout le monde se pose : pourquoi la sortie s'appelle q ? C'est une tradition établie dans les années 50, et reliée au fait que les états dans les automates s'appellent aussi q (la sortie du registre donne son état). Les internets offrent trois réponses plausibles :

- Le registre stocke un quantum d'information, et déjà dans les années 50 c'était chic de mentionner le mot "Quantum".
- On voulait écrire O pour Output, mais on risquait de confondre avec 0 (zéro), donc on a pris un O avec une petite queue, donc Q.
- Les premiers registres étaient des bascules avec Set et Reset (S et R), donc en cherchant dans les lettres voisines quelqu'un a hésité entre Q et T.

### 4.1.2 Construction : du verrou (latch) au registre (Flip-Flop)

La figure 4.2 montre un petit circuit, construit à partir d'un multiplexeur, qui n'est absolument pas un circuit combinatoire bien formé (pourquoi?).

On l'appelle *verrou*<sup>2</sup>. Il recopie d (data) sur q lorsque keep vaut 0, et garde l'état précédent lorsque keep vaut 1. Remarquez que ce n'est pas encore mon registre : il ne mémorise que la moitié du temps.

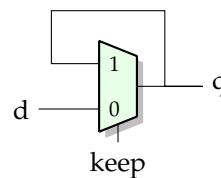


FIGURE 4.2 – Un verrou. Exercice : mettez des flèches sur les fils.

1. Dans la vraie vie (c'est-à-dire sur Wikipedia) vous trouverez parfois ce registre nommé "bascule D". Je n'aime pas parce que ce n'est pas terrible de nommer un objet à partir du nom donné à une des entrées, surtout quand tous les éléments de mémorisation sans exception ont une entrée qui s'appelle D. Et aussi parce que le mot "bascule" recouvre un peu n'importe quoi.

2. Dans la vraie vie, c'est à dire dans les vieux bouquins recopiés sur Wikipedia, vous trouverez des circuits qui sont des verrous mais qui sont appelés bascule RS et bascule JK, et des variantes de flip-flop qui sont appelés bascule D. C'est historique mais c'est confus. Pour les 0.5% d'entre vous qui auront besoin de se frotter à la vraie vie, comprenez bien la différence entre un verrou et un flip-flop, et tout ira bien.

Sur la figure 4.3, on colle l'un derrière l'autre deux verrous fonctionnant sur les états différents de keep. Sauf que du coup keep est renommé horloge (*clock* ou *ck*).

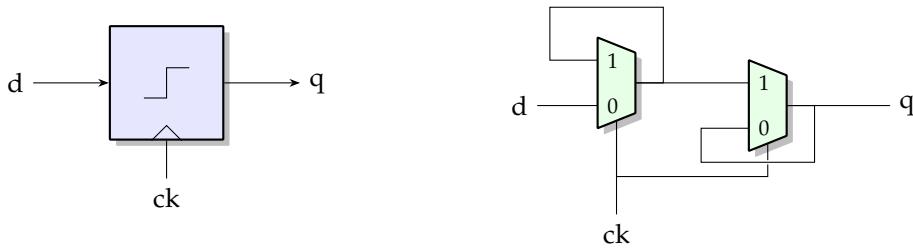


FIGURE 4.3 – Construction du flip-flop. Exercice : mettez des flèches sur les fils.

Convaincons-nous qu'on obtient le registre (ou Flip Flop) de la figure 4.1 p.44.

Le premier verrou est passant sur l'horloge basse, le second sur l'horloge haute. Ainsi, lorsque l'horloge vaut 0, la valeur de *d* est placée à l'entrée du second verrou (flip). Mais le second verrou ne la laisse pas passer. Jusqu'à ce que l'horloge passe de 0 à 1 (flop) : alors la valeur de *d* est mémorisée dans le premier verrou (et transférée à la sortie par le second verrou, qui est devenu passant). Lorsque l'horloge repasse à 0, c'est le second verrou qui mémorise à son tour la valeur (qui est toujours la valeur qu'avait *d* au front montant précédent). Le premier redevient passant mais on s'en fiche : on est revenu au début du paragraphe.

#### 4.1.3 Variantes : registre avec reset et / ou write enable

*L'utilité des deux variantes exposées ici n'est pas claire à première lecture. Promis, elles s'avéreront indispensables dans les chapitres qui suivent.*

On aura souvent besoin d'ajouter au registre une entrée *reset* qui définit son état initial. Ce sera un multiplexeur sur l'entrée *d*, comme dessiné sur la figure 4.4. On a ici un *reset* (remise à 0), synchrone car la remise à zéro attendra le prochain front d'horloge. Pour le même prix vous pouvez construire un registre avec "set", qui initialise à 1.

Il est possible aussi de construire un flip-flop avec reset asynchrone (qui n'attend pas le front montant) en ajoutant des portes sur les entrées 0 et 1 des mux de la figure 4.3 – je vous le laisse en exercice.

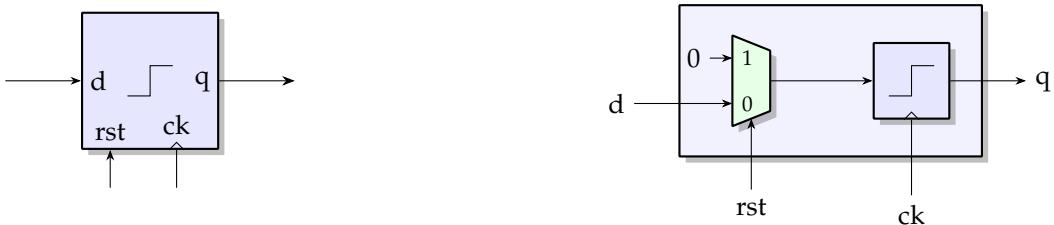
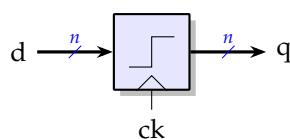


FIGURE 4.4 – Un registre avec reset (souvent abrégé *rst*) synchrone

On aura aussi souvent besoin d'y ajouter une entrée parfois nommée *write enable* (*we*), parfois *load*. Il arrive même qu'elle soit nommée *clock enable* mais c'est très maladroit : on verra au chapitre suivant qu'il est essentiel de s'interdire de mettre de la logique sur l'horloge. La raison principale est qu'on veut éviter le risque de *transitoires* sur l'horloge, qui ajouteraient des fronts montant, et donc des mémorisations non voulues. Donc on va juste ajouter un multiplexeur sur l'entrée *d*, comme sur la figure 4.5, qui soit choisi *d*, soit choisi de recycler la valeur précédemment stockée (et donc disponible en sortie).

#### 4.1.4 Le registre n-bit

Mettons 8 registres en parallèle, avec la même entrée d'horloge : on obtient un registre 8 bits :



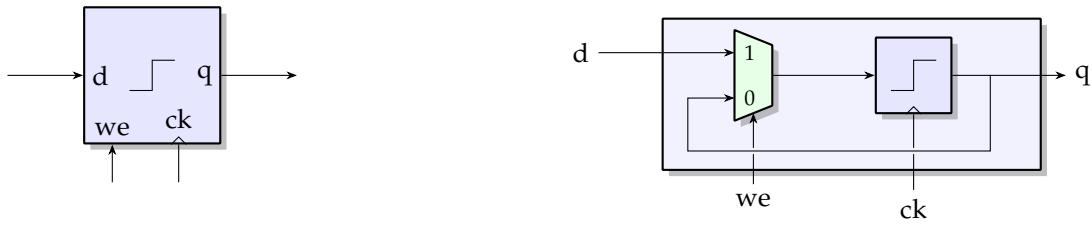


FIGURE 4.5 – Un registre avec “write enable” qui ne rajoute pas de logique sur l’entrée d’horloge

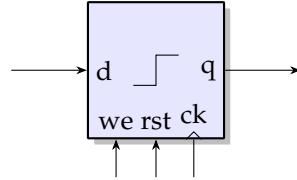


FIGURE 4.6 – Un registre avec “write enable” et reset. Exercice : construisez-le vous-même.

Par extension, les mémoires de travail dans les processeurs sont appelées des registres. Exemples de phrases utilisant cette acceptation du mot “registre” : *le PC (programme counter) est un registre à l’intérieur du processeur. Quand j’étais petit, les processeurs avaient 4 registres entiers et 8 registres flottants. On les a étendu au cours des années avec des registres plus grand, de nos jours ils ont plein de registres 512 bits.*

## 4.2 Mémoire adressable

Une mémoire adressable, c'est essentiellement plein de registres juxtaposés. Chaque registre a une adresse, et on peut sélectionner un registre particulier par son adresse. Cette sélection se fera par un genre de multiplexeur.

L'interface la plus simple d'une mémoire adressable est la figure 4.7.

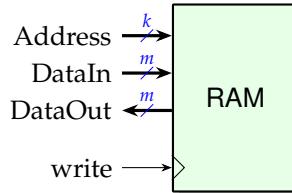


FIGURE 4.7 – Interface d’une mémoire adressable ou mémoire à accès aléatoire ou RAM de  $2^k \times m$  bits

Les opérations de base qu'on peut faire avec une mémoire adressable sont

- lire la donnée à l'adresse  $a$ . Dans ce cas il faut donner l'information d'adresse à la mémoire, et elle répond par l'information contenue dans la cellule visée.
- écrire une donnée  $d$  à l'adresse  $a$ . Dans ce cas on doit donner à la mémoire les informations d'adresse et de donnée, et un instant à partir duquel le contenu de la mémoire est changé.

Ici, la lecture d'une mémoire est une fonction combinatoire des entrées, alors que l'écriture demande un fil d'horloge (qui porte une information de temps : à quel moment précis le contenu de la mémoire va changer). C'est donc très différent.

On appelle aussi la mémoire adressable RAM (*random access memory*), parce que c'est plus court. Parenthèse culturelle : *random* veut dire aléatoire, ce qui parfois veut dire “au hasard”, mais pas ici. Ici *random* veut dire “au choix”. Pour comprendre ce terme il faut se souvenir qu'il est arrivé par opposition aux mémoires d'avant les RAM, qui étaient toutes des variantes de la bande magnétique ou du carton perforé : pour récupérer une donnée à l'autre bout de la bande, il fallait souvent attendre longtemps. Quand les RAM sont arrivées, cet accès aléatoire, c'était la fête.

### 4.2.1 Construction d'une mémoire adressable

Elle repose sur  $2^k$  registres/flip-flops de  $m$  bits (figure 4.8).

Côté lecture, c'est facile : un multiplexeur 2<sup>k</sup>-vers-1 contrôlé par l'adresse sélectionne la sortie de registre qui sera transmise sur DataOut.

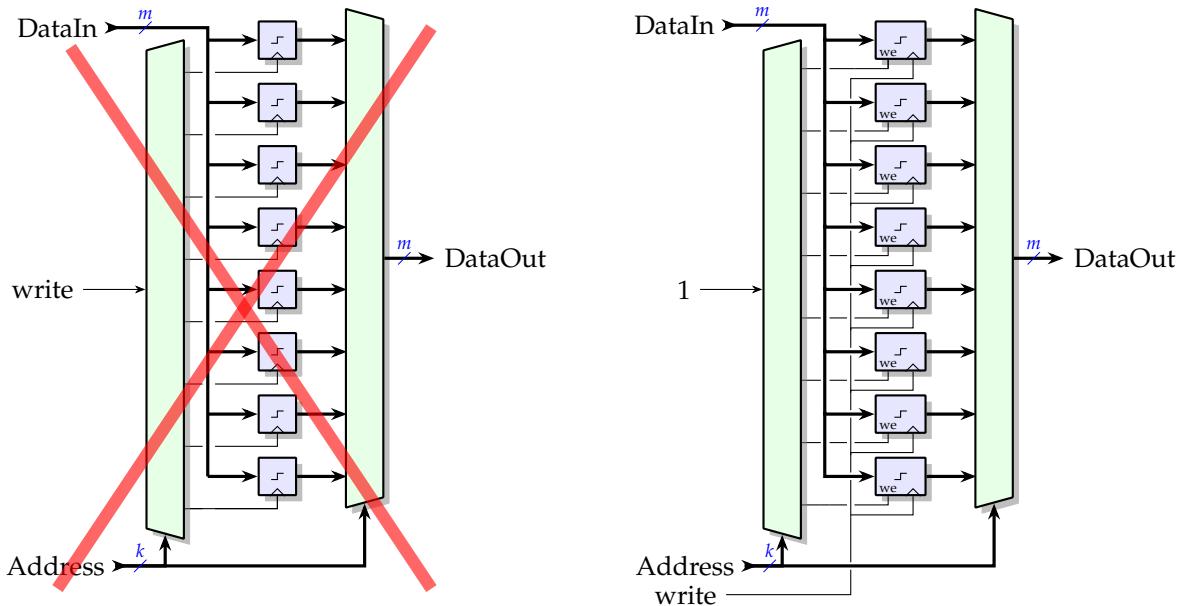


FIGURE 4.8 – Construction d'une RAM  $2^k \times m$  (ici  $k = 3$ ) : version douteuse à gauche, version correcte à droite.

Côté écriture, on présente DataIn en entrée de chacun des registres, et yaka avoir un démultiplexeur 1-vers- $2^k$  qui transmet le signal write sur toutes les entrées d'horloge des registres (figure 4.8, gauche).

Toutefois cette architecture pose un problème : si on fait changer l'adresse pendant que write vaut 1, cela peut créer des transitions de 0 à 1 sur les entrées d'horloge de plusieurs registres, donc plusieurs mémorisations...

Une solution est de relier write directement à l'entrée d'horloge de tous les registres. Pour choisir dans lequel on écrit, on utilise des registres avec *write enable*, et le démultiplexeur contrôlé par l'adresse va positionner à 1 une seule de ces entrées we. Convainquez-vous que cela résout le problème précédent :

- on n'écrit dans une cellule qu'à un front montant de write,
- et le démultiplexeur assure qu'une seule cellule est alors écrite.

*Les sections 4.2.2 et 4.2.4 sont contingentes à la technologie CMOS : considérez-les comme hors programme, ou moins fondamentales que tout ce qui précède. Vous pouvez sauter directement à la section 4.2.3. Ou pas, car ces trois sections sont fort intéressantes.*

## 4.2.2 Construction des mémoires utilisant des porte de transmission

Le multiplexeur de la figure 4.8 est  $m$  fois plus coûteux que le démultiplexeur puisqu'il aiguille des paquets de  $m$  bits. De plus, on a déjà l'information de la case à envoyer vers la sortie fournie par le démultiplexeur. N'est-il pas possible de l'exploiter ? La réponse est oui, mais cela demande de sortir de la stricte logique booléenne.

**Porte de transmission** La brique de base pour cela est la porte trois-états (aussi appelée porte de transmission). Elle agit comme un interrupteur : un fil peut désormais être dans trois états qui sont 0, 1, ou "débranché" (parfois noté Z pour "haute impédance"). La figure 4.9 donne son schéma et sa réalisation à l'aide d'un inverseur<sup>3</sup> et de deux transistors CMOS. Si *In* est relié à une valeur booléenne, alors *Out* est soit porté à la même valeur (0 ou 1) lorsque *Ctrl*=1 (les deux transistors sont passant), soit isolé de *In* lorsque *Ctrl*=0 (les deux transistors sont isolant).

Remarquez que cette "porte" est en fait bidirectionnelle. L'information peut la traverser dans les deux sens. Par contre elle n'amplifie pas (ne regénère pas) le signal qui la traverse.

3. Souvent l'inverseur est amorti car il est partagé par plein plein de portes... vérifiez sur les dessins qui suivent.

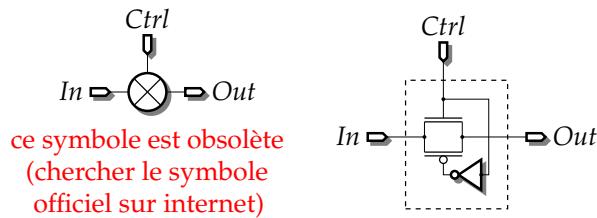


FIGURE 4.9 – La “porte de transmission”.

**Remplacement du multiplexeur de sortie par un bus trois états** Un petit dessin vaudrait mieux que ce long discours, il viendra un jour. En attendant je renvoie à la figure 4.10 qui vous donnera l’intuition.

Sur cette figure, vous reconnaissiez des petits bistables qui constituent chaque bit de mémoire. A gauche de chaque bistable, on a une porte de transmission qui permet de le brancher sur un fil commun vertical qui l’envoie sur DO. Ces portes de transmission sont contrôlées par les sorties ENi (comme Enable) du démultiplexeur d’adresse. Ceci assure mécaniquement qu’un seul bit de registre est connecté à DataOut. Et bien sûr, cette solution est moins chère que de mettre un multiplexeur par bit de DO.

### Une architecture mémoire statique utilisant des portes de transmission

La figure 4.10 montre aussi une solution bien plus crade, mais en bien moins de transistors que le flip-flop pour faire entrer une valeur dans un bistable.

Pour écrire une valeur, on la force dans le bistable au moyen des “gros” inverseurs/buffers du haut et des deux portes de transmission entourant chaque bistable. On va techniquement provoquer un petit court-circuit transitoire lorsque le bistable contenait un 0 et qu’on veut y écrire un 1. Mais cela ne dure pas assez longtemps pour abîmer quoi que ce soit, le bistable prend vite la nouvelle valeur.

Cecis permet de passer de 18 transistors (un flip-flop avec write enable) à 8 (2 par inverseur, 2 par porte de transmission) par bit de mémoire statique. Mais c’est navrant, on s’éloigne de plus en plus de la belle logique booléenne.

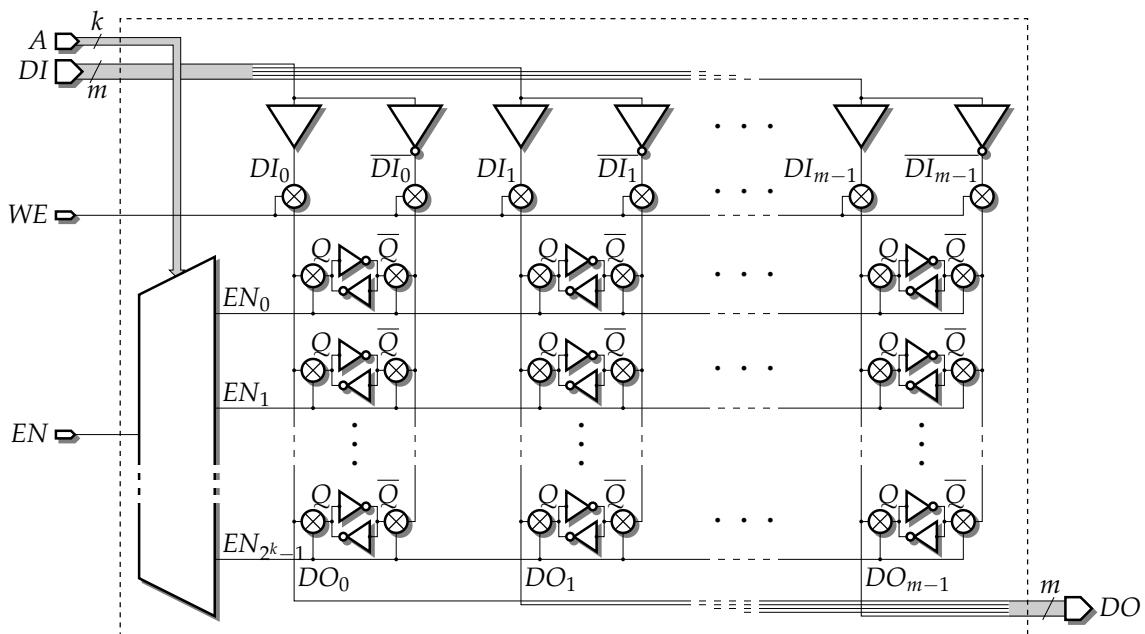


FIGURE 4.10 – Mémoire statique utilisant des portes de transmission

### Mémoire adressable à bus de donnée bidirectionnel

La porte de transmission nous permet aussi de faire passer DataIn et DataOut sur le même paquet de fils, qu’on va alors appeler bus de donnée, ou Data. Les données peuvent aller dans un sens ou dans l’autre sur ce bus (bus bidirectionnel). Pour cela, on définit (figure 4.11) un signal r/w qui donne la direction de l’information.

Au front montant de l'horloge,

- en cas de lecture (si  $r/\bar{w} = 1$ ) le contenu de la case adressée est transféré sur le bus de données,
- en cas d'écriture (si  $r/\bar{w} = 0$ ) le contenu du bus de données est transféré dans la case adressée.

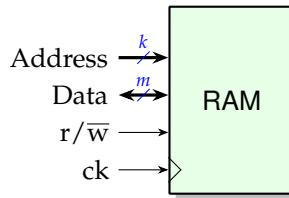


FIGURE 4.11 – Interface d'une RAM de  $2^k \times m$  bits qui nécessite un bus 3-états

Le processeur qui est relié à cette mémoire produit le signal  $r/\bar{w}$ , donc le connaît : à l'intérieur de ce processeur, on trouve également des portes de transmission qui relient le bus à des entrées de portes quand  $r/\bar{w} = 1$ , et à des sorties de portes lorsque  $r/\bar{w} = 0$ .

Bref, on peut faire du travail propre avec des portes de transmission à condition d'avoir mis en place un protocole qui garantit mécaniquement l'absence de court-circuit.

### 4.2.3 Et pour des puces plus carrées

Nos dessins vont s'étirer lorsque  $k$  devient grand. Pour obtenir un circuit plus carré et néanmoins régulier, on peut organiser nos  $2^k$  cellules mémoires en  $2^{k_1}$  lignes de  $2^{k_2}$  colonnes, avec  $k = k_1 + k_2$ . Pour retrouver une cellule, on découpe l'adresse en deux paquets de bits : les  $k_1$  bits de poids fort donnent l'adresse de la ligne, et les  $k_2$  bits de poids faible donnent l'adresse de la colonne dans la ligne. Le décodage se fait alors avec un démux vertical de 1 vers  $2^{k_1}$ , et un démux horizontal de 1 vers  $2^{k_2}$ .

### 4.2.4 Mémoire dynamique

Notre Flip-Flop est assez coûteux (une vingtaine de transistors dans la version propre, 8 transistors dans la figure 4.10).

Un bit de mémoire *dynamique* est une simple capacité isolée du reste du monde par une porte de transmission. Le contrôle de cette porte de transmission sera un unique démultiplexeur (celui de gauche de la figure 4.8). Tous les bits de mémoire peuvent alors être reliés à un bus partagé en lecture et en écriture.

Avantage : c'est plus petit.

Inconvénient : la lecture est destructrice (elle vide la capacité donc met le bit à 0 s'il était à 1) : il faut réécrire chaque bit juste après l'avoir lu.

De plus la capacité se décharge toute seule, lentement. Il faut donc la régénérer périodiquement en la lisant puis en la réécrivant à la même place. Pour ces deux raisons, une mémoire basée sur des capacités est traditionnellement nommée *dynamique*.

Par comparaison, tous les flip-flops dérivés du bistable sont de la mémoire *statique* : on a beau les lire, la valeur du bit est conservée. Et elle ne s'évapore pas avec le temps.

Les puces sur les barrettes de RAM DDR<sup>4</sup> contiennent un tableau de bits extrêmement dense puisqu'il n'y a plus qu'une capacité et une porte de transmission (voire une capacité et un transistor) par bit. Et autour on y trouve la mécanique qui rafraîchit périodiquement chaque ligne de mémoire (ne me demandez pas à quelle fréquence). C'est rentable puisque le bouzin qui assure le rafraîchissement périodique d'une ligne de mémoire a un coût proportionnel à la taille d'une ligne ( $2^{k_2}$  dans la notation précédente), et donc reste petit devant la taille de la mémoire ( $2^{k_1+k_2}$ ).

### 4.2.5 Mémoire volatile ou non volatile

Les technos vues jusqu'ici (le flip-flop, et la capacité isolée par une porte de transmission) sont des mémoires volatiles : quand on coupe l'alimentation, on perd l'info.

On sait aussi faire des mémoires *non volatiles* : par exemple la surface d'un disque dur, avec des particules qui peuvent être aimantées dans un sens ou dans l'autre. Dans le CDRW on avait deux états stables d'un alliage (amorphe ou cristallin).

4. Cherchez sur wikipedia lequel des D veut dire Dynamic.

Dans un SSD (solid-state disk<sup>5</sup>), on est en train de remplacer ces technologies par de la *mémoire flash* : on isole la grille d'un transistor, et on y piége des électrons par claquage. C'est une mémoire statique et non volatile, haute densité, à lecture non destructrice (puisque le transistor reste bloqué ou passant). Inconvénient : l'écriture est relativement lente (haute tension), et use la cellule (seulement de l'ordre de 100 000 cycles d'écriture).

Actuellement, il y a pas mal d'agitation pour trouver la mémoire "idéale" :

- petit point mémoire
- rapide
- non volatile, 10 ans de rétention d'information
- résistante à  $10^{15}$  cycles de lecture/écritures
- pas chère à construire

Aucune des solutions existantes (googlez MRAM, FeRAM, RRAM, etc.) ne réunit toutes ces qualités... mais il y a un gros gros marché pour le premier qui y arrive.

## 4.3 Mémoires à accès séquentiel : piles et files

### 4.3.1 File ou FIFO (*first in, first out*)

On appelle aussi la file : tampon ou *buffer*, parfois *queue* (qui marche dans les deux langues). C'est une primitive extrêmement importante dans un système informatique complexe : chaque fois que deux composants interagissent selon des rythmes différents, on met de l'huile dans leurs relations en introduisant un tampon qui introduit l'élasticité nécessaire.

Par exemple, lorsque votre carte réseau reçoit un paquet de données, elle les entasse dans un tampon en attendant que le bus vers la mémoire du PC se libère (tout en signalant qu'elle a besoin du bus). Quand le bus se libérera, elle pourra vider le tampon sur ce bus beaucoup plus vite qu'il se remplit. Il y a toute une branche de recherche sur le dimensionnement des files d'attentes.

En matériel un exemple d'interface de buffer est la figure 4.12. Notez les deux horloges, qui illustrent bien que le producteur et le consommateur ne vivent pas au même rythme. On peut imaginer que la lecture est combinatoire (on trouve sur DataOut la tête de file) et que CkOut se lit "Donnée suivante, SVP". C'est bien une horloge car elle change l'état interne du buffer. Cela dit on peut aussi avoir une seule horloge et deux *enable*.

### 4.3.2 Pile ou LIFO (*last in, first out*)

... ou *stack* en Australien.

L'interface est vraiment identique, comme quoi l'interface ne suffit pas à décrire le comportement d'un composant.

Cette primitive est très importante en programmation, mais on a en fait rarement besoin du composant matériel illustré par la figure 4.13.

### 4.3.3 Construction de piles et files

La construction de ces primitives est laissée comme un exercice pour le lecteur... mais quand il aura lu les deux chapitres suivants.

En effet on ne sait pas faire tellement plus intelligent que les construire autour d'une mémoire adressable.

Pour une pile, on ajoute un registre pointeur de pile, et un additionneur qui ajoute 1 (pour un push) ou -1 (pour un pop).

5. Bof. Les disques durs étaient déjà essentiellement composés de solides. Ici la notion importante est qu'il n'y a plus de pièce mobile.



FIGURE 4.12 – Exemple d'interface d'un buffer

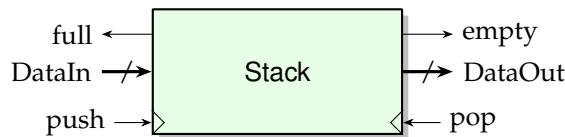


FIGURE 4.13 – Exemple d'interface d'une pile

Pour une file on a deux registres (un pointeur vers l'entrée et un vers la sortie) et deux additionneurs qui ajoutent 1 à l'un ou l'autre de ces registres. Les pointeurs se courrent après modulo  $2^n$  (où  $n$  est le nombre de bits d'adresse de la mémoire). On détecte les situations full et empty par des comparaisons des deux registres.

## 4.4 Mémoires adressables par le contenu

En anglais, *Content Addressable Memory* ou CAM.

Analogie avec le dictionnaire. On ne veut pas retrouver l'information qu'on connaît déjà (la clé, ici un mot) mais une autre information qui y est associée (la définition). En général, on a dans une mémoire adressable par le contenu des couples (clé, valeur).

Primitives :

- rechercher(clé)
- insérer(clé, valeur)
- arbitrage si clé est déjà présent
- supprimer(clé)

Il y a des variations suivant les stratégies de remplacement, et en général en fonction de la manière dont on les construit.

Normalement vous verrez l'implémentation de ces choses là en logiciel.

On discutera l'implémentations matérielle des CAM quand on en aura besoin (dans le chapitre sur la hiérarchie mémoire).

## 4.5 Disques (\*)

Cette section sera bientôt obsolète.

Piste, secteur.

Temps d'accès (qq ms), débit (qq Mo/s).

Les disques sont de petits objets sensibles plein de pièces mouvantes et fragiles. Solution : disques RAID (*redundant array of inexpensive disks*). Pour vous faire utiliser des codages redondants.

## 4.6 Une loi fondamentale de conservation des emmerdements

Comme l'illustre la table 4.1, une mémoire de grande capacité est lente ; une mémoire rapide est limitée en taille, et chère.

*Ce n'est pas juste une question de compromis technologique (par exemple statique contre dynamique), c'est aussi une question de complexité algorithmique :*

- plus votre mémoire a une grande capacité, plus elle est étendue dans l'espace, et donc plus déplacer certains de ses bits vers la sortie (même à la vitesse de la lumière) prendra du temps ;
- pour une mémoire de capacité  $n$ , le décodage d'adresse utilise  $\log_2 n$  étages d'aiguillages : plus votre mémoire a une grande capacité, plus l'accès sera lent.

Type	temps d'accès	capacité typique
Registre	0.1 ns	1 à 128 mots (de 1 à 8 octets)
Mémoire vive	10 - 100 ns	4 Goctets
Disque dur	10ms	100Goctets
Archivage	1mn	(illimité)

TABLE 4.1 – Hiérarchie mémoire simplifiée d'un ordinateur de 2010. Voir la table 13.1 page 117 pour la version complète.

Remarques :

- Cette table 4.1 est tout entière victime de la loi de Moore.
- Le coût de chaque étage est grossso modo équivalent. Le coût par octet est donc exponentiel par rapport à la vitesse d'accès.

On verra plus tard (au chapitre 13) comment construire des mécanismes qui donnent l'illusion que que toute la mémoire vive est aussi rapide que les registres (mémoire cache) et aussi grande que son disque (mémoire d'échange ou *swap*).

# Chapitre 5

## Circuits séquentiels synchrones

On a vu au chapitre 3 comment implémenter des fonctions booléennes sous forme de circuits combinatoires. Je ré-insiste lourdement sur le fait que ce sont des fonctions au sens mathématique ( $f : x \mapsto f(x)$ ), c'est-à-dire sans mémoire. Leur sortie ne dépend que de l'entrée, pas de ce qui s'est passé avant.

On a vu au chapitre 4 comment construire des mémoires.

Dans ce chapitre, nous allons fusionner tout cela pour construire des circuits plus généraux, mêlant calcul (combinatoire) et mémorisation. On appellera de tels circuits des circuits séquentiels.

### 5.1 Quelques exemples de circuits séquentiels

Voici des exemples de circuits qui sont séquentiels et non pas uniquement combinatoires :

- Un compteur, obtenu en reliant un additionneur et un registre de  $n$  bits, comme illustré sur la Figure 5.1. Dans le cycle de von Neumann, le PC avance grâce à un circuit similaire.

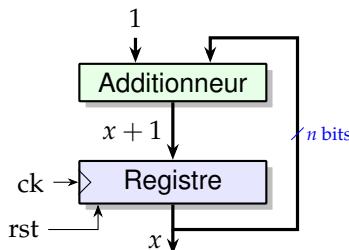


FIGURE 5.1 – Compteur automatique de moutons pour s'endormir plus vite (patent pending). Il a déjà compté jusqu'à  $x$  moutons et s'apprête à compter le  $x + 1$ -ième.

- Votre montre à quartz numérique<sup>1</sup> est pleine de compteurs : un pour les secondes, un pour les minutes, etc. C'est un circuit séquentiel.
  - Le circuit caché derrière votre digicode. Il doit se souvenir des chiffres déjà tapés ; il a une mémoire, c'est un circuit séquentiel.
  - la mémoire adressable du chapitre précédent a évidemment une mémoire
  - Votre téléphone, votre console de jeux, votre PC sont des circuits séquentiels.
- Bref, la classe des circuits séquentiels est la plus générale.

### 5.2 Restriction aux circuits séquentiels synchrones

Dans tout ce cours, on va faire une grosse restriction sur les circuits séquentiels : **ils ne devront avoir qu'une seule horloge**. Plus précisément, **les entrées d'horloge de tous les registres binaires devront utiliser le même signal d'horloge**. La classe de circuits obéissant à cette restriction est la classe des *circuits séquentiels synchrones*.

Remarque : il y a  $n$  registres binaires cachés dans la Figure 5.1, et ils utilisent bien tous le même signal d'horloge.

La raison de cette restriction est qu'elle rend les circuits infiniment plus facile à concevoir, comprendre, vérifier et fabriquer. La preuve : vous allez apprendre à faire tout cela dans les pages qui suivent.

1. Une vach'té de chouette idée...

Ce n'est pas une restriction pédagogique : l'écrasante majorité des circuits manipulés dans l'industrie sont des circuits synchrones. En effet, l'industrie aussi veut comprendre les circuits qu'elle fabrique, pour fabriquer des circuits qui marchent. Sans cette restriction, on arrive vite à des circuits séquentiels indébuggables, illisibles, imprévisibles, et en général mal élevés. La raison principale en est qu'on ne peut se permettre d'avoir des transitoires qui ajoutent des fronts sur l'horloge (on a parlé des transitoires p. 34).

On observe au passage que les circuits séquentiels contiennent des parties qui sont des circuits combinatoires (par exemple l'additionneur sur la Figure 5.1). On continue à s'interdire, dans ces parties combinatoires, ce qui était interdit jusque là (brancher deux sorties ensemble, faire un cycle). Par contre, on se permet des cycles dans les circuits séquentiels, **à condition qu'il y ait au moins un registre sur chaque cycle**. Remarquez que c'est le cas de la la Figure 5.1.

**Un circuit séquentiel qui n'est pas synchrone** *On peut construire un compteur uniquement en branchant entre eux  $n$  flip-flops. Je le dessinerais au tableau uniquement, pour être sûr que vous ne le retenez pas. Il n'est pas synchrone bien qu'il n'ait qu'une seule horloge en entrée, parce que les entrées d'horloge de certains registres n'utilisent pas cette horloge. Il est plus économique que le compteur de la Figure 5.1, mais il a aussi des défauts :*

- *Il n'est pas du tout évident que c'est un compteur en le regardant (alors que la Figure 5.1 se comprend vite).*
- *Comment le transformer en un compteur qui revient à 0 quand il atteint 60 (pour votre montre à quartz numérique) ? Sur la Figure 5.1, on ajoute un comparateur à 60 et un multiplexeur et c'est dans la poche.*

Désormais, on se permettra de dire "circuit synchrone" au lieu de "circuit séquentiel synchrone" qui est, du reste, redondant.

### 5.3 Correction et performance

Observons ce qui se passe lors d'une transition sur l'horloge dans un circuit synchrone, par exemple notre compteur.

Partons d'un front montant de l'horloge unique. Comme tous les registres utilisent ce signal d'horloge, toutes les sorties de registres basculent en même temps. Ces sorties de registres sont les entrées de la logique combinatoire qu'il y a entre les registres. À partir de cet instant, les informations font la course en se propageant dans ces parties combinatoires. Le chronogramme de la figure 5.3 décrit le déroulement de cette course. Pour la comprendre, il faut regarder la figure 5.2, copiée de la page 37, qui montre comment est construit l'additionneur.

Toutes les entrées de cet additionneur basculent au même instant (par exemple au premier front montant de  $clk$  sur la figure 5.3). Un peu plus tard, le bit  $s_0$  se stabilise, puis le bit  $s_1$  et ainsi de suite. En effet, chaque FA doit attendre que son entrée de retenue ( $c_i$ ) soit stabilisée, donc que le FA à droite soit stabilisé. On a donc une propagation de la retenue dans l'additionneur, avec des transitoires sur les sorties. Ces transitoires sont difficile à prévoir, et comme elles ne nous intéressent pas, on les dessine juste comme du gris sur la figure 5.3).

Mais au bout d'un certain temps, les sorties de l'additionneur sont toutes stabilisées. On est alors dans la situation décrite sur la figure 5.1 : il y a une valeur  $x$  en sortie des registres, et une valeur  $x + 1$  en entrée des registres. Au front suivant de l'horloge, tous les registres basculeront en même temps. La valeur  $x + 1$  sera alors stockée dans les registres, donc présentée à leur sortie, et un nouveau cycle pourra commencer.

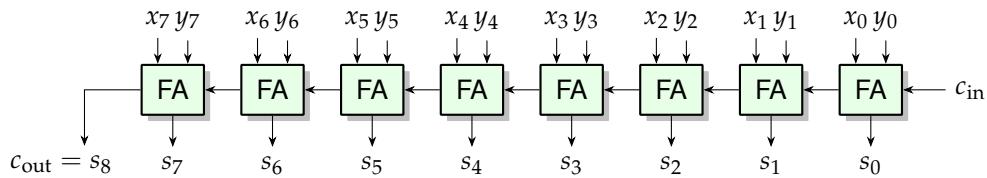


FIGURE 5.2 – Rappel : l'intérieur de la boîte "Additionneur" de la figure 5.1.

Il y a deux conditions pour que ce compteur fonctionne comme décrit.

- Il faut que tous les registres basculent en même temps.
- Il faut que le délai entre deux fronts montants de l'horloge soit plus grand que le temps de calcul au pire cas de notre additionneur.

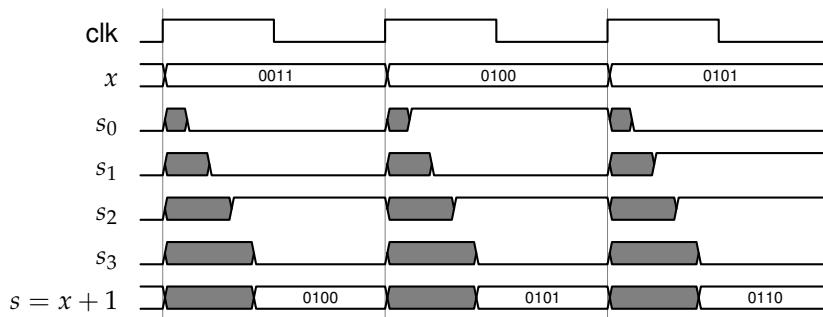


FIGURE 5.3 – Chronogramme détaillé de ce qui se passe dans le compteur. Le gris signifie “inconnu” ou encore “transitoires possibles”.

On a déjà parlé p. 34 du chemin critique dans un circuit combinatoire : c’était le chemin (à travers des portes et des fils) le plus long, c'est-à-dire qui va propager l'information le plus lentement. On généralise cela aux circuits synchrones : on appelle **chemin critique** d'un circuit synchrone le plus long chemin combinatoire, autrement dit le plus long chemin (toujours en terme de temps de propagation de l'information) entre une sortie de registre et une entrée de registre. Sur l'exemple du compteur, c'est le temps d'une propagation de retenue au pire cas, illustrée sur le chronogramme.

Le délai entre deux fronts montant d'horloge, c'est l'inverse de la fréquence d'horloge (le fameux 3GHz de votre PC). Au bout du compte, c'est donc le chemin critique qui définit la fréquence maximale à laquelle fonctionnera correctement le circuit.

**Oui mais la perfection n'existe pas** *Bien sur, en pratique on ne peut pas physiquement garantir que tous les registres basculent en même temps : on aura toujours un  $\delta t$  d'incertitude, due à des différences entre les longueurs des fils, la variabilité du processus de fabrication du circuit, etc. Qu'à cela ne tienne, on va borner cette incertitude, et il suffira d'ajouter ce  $\delta t$  au chemin critique : la condition pour que le circuit fonctionne devient “chemin critique +  $\delta t < \text{période d'horloge}”.$*

Maintenant, considérons un circuit séquentiel synchrone compliqué, par exemple la puce de votre pléستان dualscrine. Il a des entrées (les boutons de la console entre autres), des sorties (les pixels de l'écran entre autres), et plein de registres à l'intérieur. Et entre les registres, plein de portes branchées les unes aux autres. Eh bien le fonctionnement est similaire : tous les registres basculent en même temps. L'information fait la course dans les portes entre les registres. Au bout d'un certain temps elle est stabilisée. Ce temps s'appelle le délai du chemin critique. Alors on peut basculer l'horloge pour attaquer un nouveau cycle.

Si on prend un peu de recul, les conditions que nous avons imposées sur les circuits séquentiels synchrones permettent de cacher toutes les transitoires. Si on n'observe que les valeurs stockées dans les **registres**, ce qu'on voit est le **résultat des fonctions booléennes**, on a complètement caché leur implémentation. On peut ainsi raisonner sur un circuit, y compris séquentiel, en terme de fonctions booléennes. C'est bien, parce que c'est abstrait et que cela cache tous les sordides détails d'implémentation.

C'est comme cela qu'on arrive à construire des circuits séquentiels comportant des milliards de portes, et qui marchent quand même.

**Pourquoi on impose que tous les registres ont la même horloge** *Si on commence à se permettre de mettre des portes sur une entrée d'horloge d'un registre, on va devoir tenir compte de leur délai. On peut essayer de faire comme pour notre  $\delta t$  : calculer le chemin critique des portes qui sont sur l'entrée d'horloge, et l'ajouter séparément aux délais des chemins qui partent de ce registre, mais pas aux autres... C'est faisable mais vous voyez que cela devient vraiment compliqué. Et surtout, si on veut appliquer une certaine fonction booléenne sur un signal d'horloge, il faut absolument éviter la moindre possibilité de transitoire dans l'implémentation de cette fonction : une transitoire sur une horloge, en ajoutant un front, ruine tout. Il faudra être très convaincant qu'il y a quelque chose à gagner. Dans le doute, je vous invite à n'envisager que des circuits dans lesquels tous les registres ont la même horloge.*

On assure le synchronisme par la construction d'un *arbre d'horloge équilibré* : dès qu'on a de nombreuses entrées d'horloge à alimenter avec le même signal, il faut l'amplifier, par un arbre d'inverseurs. Un

inverseur typique est capable d'alimenter 4 entrées rapidement, au delà de 4 sa performance chute : on dit qu'un tel inverseur a un *fan out* de 4. Pour obtenir un arbre de distribution d'horloge équilibré à l'échelle d'une puce rectangulaire, on le construit en arbre quaternaire, comme sur le dessin ci-dessous.

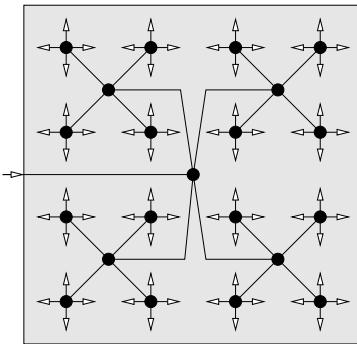


FIGURE 5.4 – Arbre d'horloge équilibré. Les ronds noirs sont des inverseurs

Révision : constatez qu'au chapitre précédent, la construction des registres avec *reset* et *write enable* respecte la règle : pas de logique sur l'horloge.

# Chapitre 6

## Automates

On avait le formalisme de la logique booléenne pour construire les fonctions combinatoires.

Nous allons présenter dans ce chapitre un formalisme, les automates<sup>1</sup>, qui permet de décrire, vérifier, et même construire automatiquement des circuits séquentiels.

Les automates sont une abstraction facile à dessiner du *comportement attendu* d'un circuit. Le comportement, c'est ce qu'on peut décrire de plus abstrait! De ce point de vue, l'automate est le pendant d'une fonction booléenne, qui décrit le comportement d'un circuit logique combinatoire.

### 6.1 Un exemple

Essayons de spécifier le comportement d'un circuit de commande d'un passage à niveau en Russie, entre les villages de Krasniy Kirpitchnik et Kamenki (emplacement choisi parce que c'est une voie unique où peuvent passer des trains dans les deux sens<sup>2</sup>, on n'a plus cela chez nous). On place des capteurs CG et CD assez loin à gauche et à droite du passage à niveau. Ce sont de simples interrupteurs déclenchés par le poids d'un train : ils renvoient un signal logique 0 quand la voie est libre, et 1 lorsqu'il y a un train dessus. On a par ailleurs un feu rouge, également abstrait par un booléen  $A$  (comme ampoule) :  $A = 0$  : éteint;  $A = 1$  : allumé.

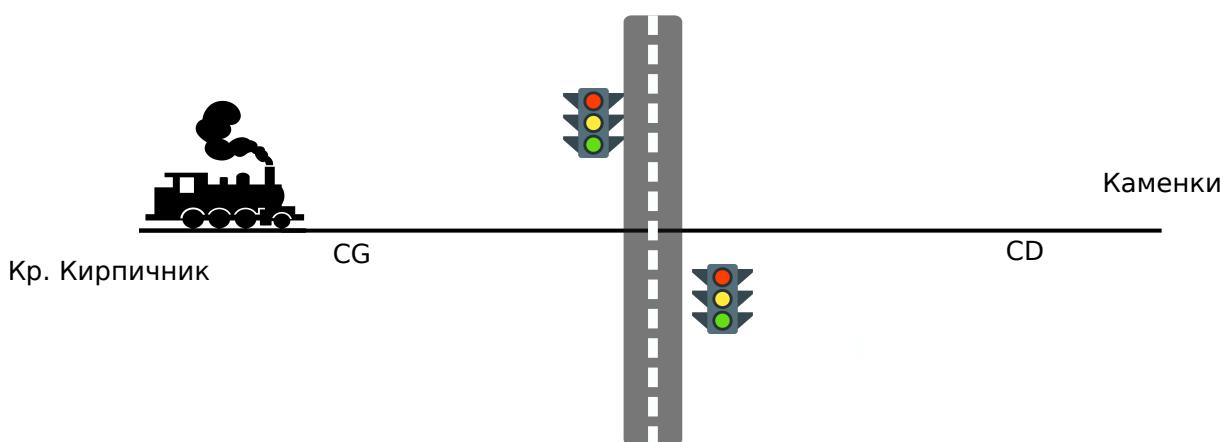


FIGURE 6.1 – Le communisme, c'est les soviets plus l'électrification des passages à niveaux.

1. Le vrai nom est *automate à nombre fini d'états*, parfois mal traduit en *automate d'états fini*, de l'anglais *finite state automaton* (c'est l'*automaton* qui est fini, pas le *state*), ou encore *finite state machine*. Tout ceci pour dire que vous verrez parfois les acronymes FSM ou FSA.

2. C'est là : <https://www.google.com/maps/dir/56.0402444,43.8462165/@56.039808,43.8463786,720m/data=!3m1!1e3>

On aimerait que le feu s'allume lorsque le train s'approche, et s'éteigne lorsque le train est passé. Convainquons nous qu'on ne s'en tirera pas par une simple fonction combinatoire des capteurs. Par exemple, parfois il passe de simples locomotives : il faut que notre circuit se souvienne de garder le feu allumé tant que la locomotive n'a pas quitté le second capteur.

La figure 6.2 décrit l'automate correspondant (seule une moitié a été complètement spécifiée, car c'est un poly interactif : vous devez gribouiller l'autre moitié pour vérifier que vous avez compris).

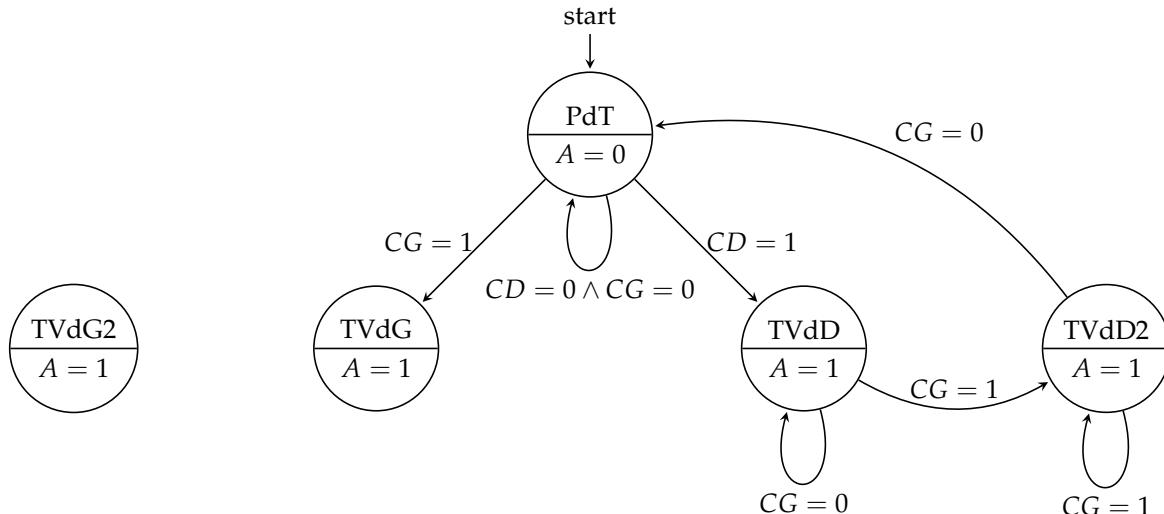


FIGURE 6.2 – *Mon premier automate. CG et CD sont les Capteurs Gauche et Droite. TVdG est l'abréviation de Train Vient de Gauche. Completez vous-même l'autre côté.*

Les patates sont des états. Un état a un nom, par exemple PdT pour *Pas de Train*. Il spécifie également les valeurs des sorties (ici on n'en a qu'une,  $A$ ), par exemple  $A = 0$  dans l'état PdT.

Entre les états, on a des flèches qui indiquent un changement d'état. On appelle ces flèches des transitions. Une transition est étiquetée par l'évènement qui déclenche le changement. Par exemple, on reste dans l'état *Pas de Train* tant que  $CG = CD = 0$ , et on passe dans l'état *Train Vient de Gauche* lorsque le capteur gauche passe à 1.

Par convention, une flèche sur laquelle une entrée n'apparaît pas signifie que la transition est prise quelle que soit la valeur de cette entrée.

Attention, il y a désormais dans ce poly deux sens techniques du mot *transition* : on connaissait une transition entre deux valeurs d'un signal (un front montant ou descendant), on a désormais aussi la notion de transition d'un état à un autre.

Notez au passage que, comme pour les variables dans un programme, cela ne fait pas de mal de donner des noms expressifs aux états, aux entrées et aux sorties.

Au tableau je montre l'automate du digicode, qui reconnaît un code.

## 6.2 Définition formelle d'un système de transitions

### 6.2.1 États, transitions

Allons-y donc pour le formalisme. Un automate de Moore, ou système de transition (c'est-à-dire tels qu'on les manipule en architecture) est un sextuplet  $(I, O, S, T, F, s_0)$  où

- $I$  est un ensemble fini des valeurs des entrées possibles,
- $O$  est un ensemble fini des valeurs des sorties possibles,
- $S$  est un ensemble fini d'états
- $T$  est une *fonction de transition* de  $S \times I \rightarrow S$ ,
- $F$  est une *fonction de sortie* de  $S \rightarrow O$ ,
- $s_0 \in S$  est un état spécial appelé *état initial*, car il en faut bien un.

Attention,  $I$  et  $O$  ne sont pas les ensembles des entrées/sorties du circuit, mais l'ensemble des combinaisons de valeurs qu'elles peuvent prendre. Par exemple si le circuit a  $n$  entrées binaires, alors  $I = \{0, 1\}^n$ .

On fait des dessins (dont la figure 6.2 est un glorieux exemple) avec

- des patates pour les états, qu'on étiquette avec le nom de l'état, et dessous la valeur de  $F$  correspondante

- des flèches d'un état  $s$  à un état  $s'$ , qu'on étiquette avec une valeur  $x$  des entrées, chaque fois que  $T(s, x) = s'$ .

Ces dessins s'appellent chez les gens sérieux des diagrammes états/transition, ou des diagrammes d'états.

### 6.2.2 Définition en extension des fonctions de transition et de sortie

Les deux fonctions sont des fonctions discrètes, c'est-à-dire avec un nombre fini d'entrées et de sorties possibles. On pourra donc les définir en extension, en listant leurs valeurs dans une table. C'est pourquoi on parle toujours d'automate fini. Par exemple :

$s$	$x=(CG, CD)$	$s'=T(s, x)$
PdT	00	PdT
PdT	01	TVdD
PdT	10	TVdG
PdT	11	XXX
TVdD	0X	TVdD
TVdD	1X	TVdD2
TVdD2	1X	TVdD2
TVdD2	0X	PdT
...	...	...

$s$	$y=F(s)$
PdT	0
TVdD	1
TVdD2	1
...	...

Que signifie le X dans ces tables ? Il signifie *don't care*, ce qui peut se traduire par "on s'en fout".

Mais attention, cela ne veut pas tout-à fait dire la même chose quand le X est en sortie de la table, et quand il est en entrée.

**En entrée**, on met un "X" sur une valeur d'entrée pour dire "quelle que soit la valeur de cette entrée". Dans notre exemple, pour chaque état, on devrait considérer les 4 combinaisons de valeurs possibles de nos deux capteurs. Mais dans l'état TVdD, le dessin à patates n'a que deux flèches. Il est donc incomplet. Dans la table de transition, on a précisé, par exemple qu'on prend la transition qui boucle de TVdD à lui-même tant que CG vaut 0, quelque soit la valeur de CD. En effet c'est bien ce qu'on veut : à partir du moment où un train vient de la droite, le feu est allumé, et ce qui nous intéresse est de savoir quand on l'éteindra. Et ce ne sera pas lorsque CD passe à 0, mais lorsque CG passera à 0. Mais avant CG doit passer à 1 : c'est cet événement qui déclenchera la seule transition sortant de TVdD. Peut-être qu'à ce moment CD sera déjà repassé à 0 (train court) ou pas encore (train long), mais on s'en fout.

En pratique, un *don't care* en entrée peut être remplacé par une énumération des transitions qu'il élide.

Par exemple, 

TVdD	0X	TVdD
------	----	------

 signifie exactement 

TVdD	00	TVdD
TVdD	01	TVdD

**En sortie** (des fonctions  $T$  ou  $F$ ), le "don't care" est plus fort... et plus dangereux. Il signifie "cette situation n'arrivera jamais, donc n'importe quelle valeur fera l'affaire". Par exemple, considérons la quatrième ligne de la table de  $T$  ci-dessus. On est dans l'état "pas de train" avec les deux capteurs à 0, et voilà que tout à coup, les deux capteurs passent à 1 en même temps. Cela signifie sans doute que deux trains sont en train de se précipiter l'un sur l'autre : il va bientôt avoir une épouvantable catastrophe ferroviaire. Et donc, on s'en fout. En effet, ce cas ne doit jamais arriver. Autre point de vue : l'automate peut bien se retrouver dans l'état qu'il veut, on aura d'autres chats à fouetter si deux trains se rentrent dedans.

En sortie, le *don't care* se lit donc plutôt "ce n'est pas mon problème", et vous voyez que pour l'ingénieur que vous êtes, cela peut être dangereux. Dans l'exemple du feu rouge, il y a peut-être plus intelligent à faire. Par exemple, aller dans un état où le feu est allumé, pour ne pas ajouter un petit accident au gros. Ou bien aller dans un état "erreur" qui va faire venir un technicien. Parce que l'allumage des deux capteurs vient peut-être aussi d'un faux contact quelquepart. Etc.

### 6.2.3 Correction et complétude d'un système de transitions

Revenons aux patates. On ne peut pas vérifier que votre automate répond bien aux spécifications fonctionnelles de votre circuit. Par contre, on peut vérifier un certain nombre de propriétés sans lesquelles il est forcément incorrect :

**déterminisme** : partant du même état, il n'y a pas deux transitions qui arrivent dans des états différents pour une même valeur des entrées.

**complétude** : dans chaque état,

- la valeur des sorties est définie.
- il y a une transition prévue pour chaque valeur des entrées (fût-ce à l'aide de *don't care*).

Ces deux propriétés sont vérifiables mécaniquement : il suffit d'essayer de construire complètement la table de transition et la table de sorties.

Dès que  $I$  grossit un peu (mettons  $I = \{0,1\}^m$  pour fixer les idées) on doit obligatoirement avoir  $|I| = 2^m$  transitions qui partent de chaque état pour que l'automate soit complet. Bonjour le plat de nouilles. Sur le dessin, il est alors naturel de regrouper ces transitions sur une seule flèche, étiquetée par un sous-ensemble de  $I$  qu'on va (tout aussi naturellement) exprimer par une expression booléenne des entrées du circuit qui capture ce sous-ensemble. C'est ce qu'on a fait implicitement sur la figure 6.2 : par exemple l'étiquette  $CG = 1$  capture 2 valeur de l'ensemble  $I$  des entrées possibles (lesquelles ?).

Mais attention, complétude et déterminisme ne seront pas triviales à vérifier sur le dessin si les flèches sont étiquetées par des fonctions compliquées. Ce n'est vraiment pas grave : dès qu'on se lance dans l'implémentation, toute incomplétude ou indéterminisme va nous sauter à la figure lors du remplissage de la table de vérité de  $T$  et  $F$ .

### 6.3 Synthèse d'un automate synchrone

Il ne reste plus qu'à coder les états par des bits. C'est toujours possible puisqu'ils sont en nombre fini. Par exemple, on peut numérotter les états, puis coder chaque état par la représentation binaire de son numéro. Alors, on peut remplacer  $s$  et  $s'$  par des vecteurs de bits dans les tables ci-dessus. Ainsi,  $T$  et  $F$  deviennent des fonctions booléennes. On sait donc les implémenter grâce au chapitre 3.

Et alors, on sait implémenter un automate par le circuit générique décrit par la figure 6.3.

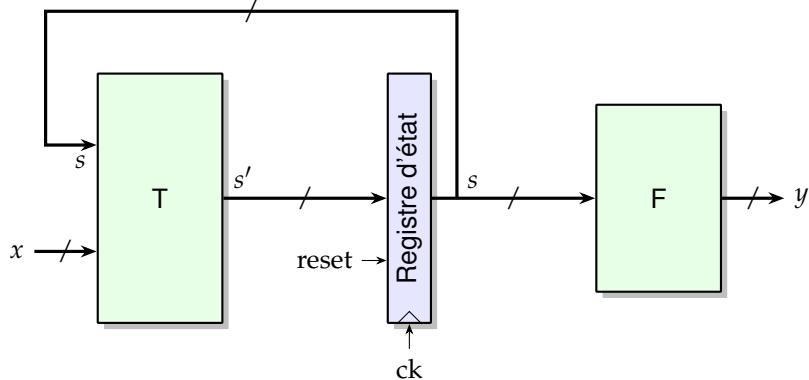


FIGURE 6.3 – *Implantation d'un automate synchrone.* Notations :  $s$  est l'état courant,  $s'$  est le prochain état,  $x$  est l'ensemble des entrées,  $y$  est l'ensemble des sorties. Notez que le reset n'est pas considéré comme une entrée de l'automate : il n'apparaît pas sur les transitions, il est la traduction en matériel de l'état initial  $s_0$ .

Explication de ce circuit : au front montant de l'horloge, l'ancien état  $s$  est remplacé par le nouvel état  $s'$ . Lequel  $s'$  est calculé par la fonction de transition  $T$  à partir de l'ancien état  $s$  et des entrées  $x$ . Par ailleurs, les sorties sont calculées à partir de l'état  $s$  par la fonction  $F$ . C'est bien ce qu'on a dessiné avec des patates.

Et on se ramène à des choses qu'on connaît bien : des flip-flops pour construire le registre d'état, et des fonctions combinatoires<sup>3</sup>.

Concrètement et bêtement, à partir d'un dessin de patates correct et complet,

- on choisit un codage des états sur  $k$  bits.
- on fait le dessin correspondant à la figure 6.3, en nommant les bits d'entrée, les bits de sortie, et les bits d'état.
- on trace le squelette de la table de vérité de  $T$  à partir de cela, en énumérant ses entrées possibles.
- on remplit les valeurs de  $T$  en suivant les patates du dessin de l'automate.

En fait il y a des tas d'outils automatiques qui feront cela pour vous.

3. Il nous faut des flip-flop avec reset pour pouvoir définir l'état initial de l'automate. C'est facile à construire, il suffit de mettre les portes qui vont bien devant l'entrée du flip-flop (reset synchrone).

### 6.3.1 L'approximation temporelle réalisée par l'automate synchrone

Oui mais on a décrit l'automate comme si il changeait d'état dès qu'un capteur passe de 1 à 0. Maintenant qu'on l'a construit, on constate que ce n'est pas tout-à-fait le cas. Il change d'état au front montant d'horloge suivant une transition des entrées. Ce n'est en général pas grave, il suffit d'avoir une fréquence d'horloge assez grande pour que ce petit délai soit négligeable. Par exemple, les trains russes roulant à 50km/h, notre circuit pourrait tourner à 100 Hz et ce serait bien assez. Or on sait le faire tourner à 1GHz.

Par contre, on comprend à présent l'intérêt de bien spécifier les transitions d'un état vers lui-même : dans notre exemple, ce seront les plus utilisées...

En pratique, il est souvent même utile d'ajouter des registres sur les entrées et sur les sorties. Cela permet de bien inclure tout  $F$  et tout  $T$  dans le calcul du chemin critique.

Pour avoir un circuit qui réagit vraiment instantanément à une transition sur une entrée, il faudrait envoyer cette entrée sur l'entrée d'horloge d'un registre d'état. Mais si on permet cela, ce n'est plus un circuit synchrone, et tout devient tout de suite beaucoup plus compliqué, comme vu au chapitre précédent. Donc je persiste à vous l'interdire.

**Les automates en UML** Vous avez peut-être déjà vu des diagrammes état-transition en UML – sinon cela vous pend au nez. Dans ces diagrammes, les transitions sont déclenchées par des événements de manière asynchrone. Elles peuvent être étiquetées par des actions, réalisées lors de la transition. Bref, c'est bien les mêmes concepts, mais les automates UML sont plus généraux et ne sont pas tous traduisibles mécaniquement en la figure 6.3. Donc ne pas mélanger.

Et pour finir, il subsiste un tout petit risque de *metastabilité* : si on a une transition d'une entrée trop près d'une transition d'horloge, certains registres peuvent se retrouver coincés entre le 0 et le 1. Cela passe tout seul, mais parfois en plusieurs cycles d'horloge<sup>4</sup>... Ce risque est bien isolé, et peut être minimisé par des solutions technologiques dans le détail desquelles je ne me risquerai pas.

### 6.3.2 Optimisation d'un automate synchrone

La question de la minimisation de la fonction combinatoire, qui était déjà un problème difficile, l'est encore plus ici, puisqu'on a un degré de liberté de plus : on peut choisir arbitrairement le codage des états de l'automate par des vecteurs de bits.

Toutefois, ici encore, la connaissance du problème permet souvent d'imposer un codage des états qui va minimiser la complexité de la fonction de transition. Typiquement, il s'agit de remplacer le gros automate par plein de petits sous-automates relativement indépendants (c'est à dire avec relativement peu de transitions entre eux).

Exemple : une montre à quartz, d'un point de vue formel, est un automate à autant d'états qu'elle peut afficher d'heures différentes (dans les  $24 \times 60^2$ , sans compter les jours et les mois). En pratique, il est construit algorithmiquement par composition de petits automates à moins d'états, de taille totale en gros  $24 + 60 + 60$ .

Chacun des sous-automates ayant moins d'états et moins d'entrées, la fonction de transition a moins d'entrées, et est donc plus facile à minimiser.

Pour finir, on peut envisager deux extrêmes pour le codage des états :

- les codages minimaux en termes de bits : on code  $n$  états par un vecteur de  $\log_2 n$  bits.
- le codage à jeton (*one-hot encoding*) : on code  $n$  états par un vecteur de  $n$  bits.

Le second peut aboutir à un circuit plus petit si la fonction de transition s'en trouve très simplifiée.

## 6.4 Comprendre les circuits séquentiels comme des automates

Nous sommes partis de l'automate-patates pour le construire sous forme de circuit synchrone. Mais on peut aussi faire le contraire. L'exemple de la montre à quartz vu comme un automate à  $24 \times 60^2$  états peut être généralisé : tout circuit séquentiel synchrone peut être vu comme un gros automate. Remarquez que notre compteur du chapitre précédent était un cas particulier de la figure 6.3.

Par contre le nombre d'états possible devient vite astronomique (et indessinable) : si vous avez  $n$  registres binaires dans votre circuit, alors vous avez  $2^n$  états possibles de ce circuit. Par exemple, votre Pentium contient des dizaines de milliers de registres binaires. Tournez les pages vers l'introduction de

4. Amis gamerz, désormais, quand votre petite sœur vous met une pâtee à KickFighter XVII, vous pourrez accuser la metastabilité dans votre manette de jeu chaque fois que vous serez en retard sur une action...

ce poly : tous les atomes de l'univers ne suffiraient pas à faire une feuille de papier assez grande pour dessiner  $2^{10000}$  patates.

Cette vue est cependant parfois utile, par exemple pour tester les circuits.

### 6.4.1 La norme JTAG

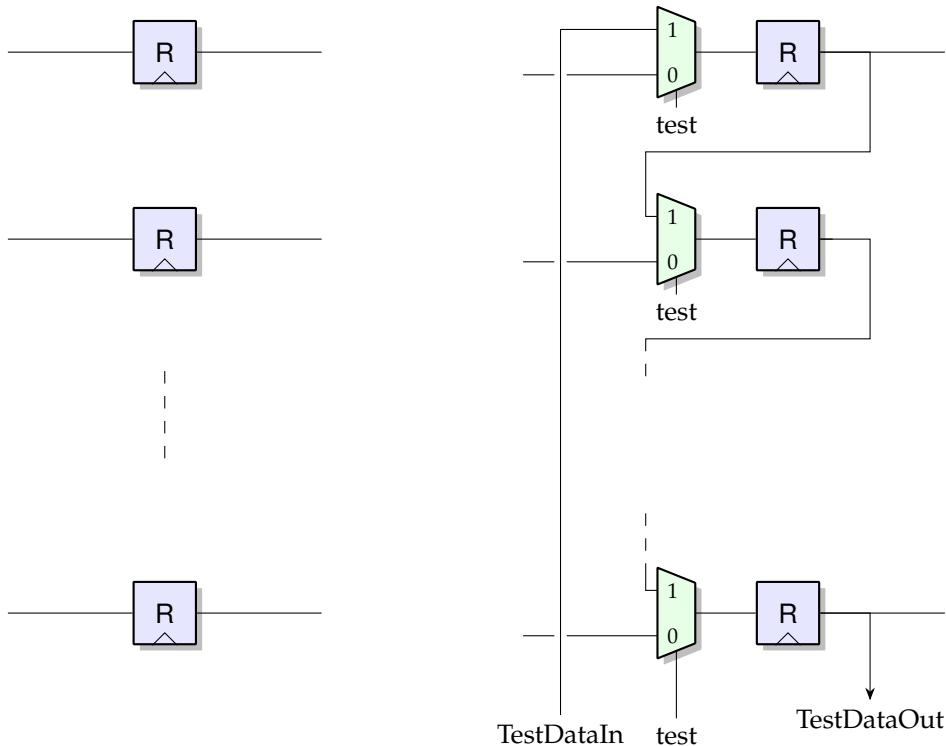
Comment teste-t-on un Pentium à  $2^{10000}$  états avant de le mettre en boîte ?

Bien sûr, on pourrait lui faire booter Linux puis Windows et jouer un peu à Quake dessus, et on se dirait qu'on a tout testé. Mais cela prendrait de longues minutes par puce, et le temps c'est de l'argent. Voici une technique qui permet de tester toute la puce en quelque centaines de milliers de cycles seulement (exercice si vous trouvez que c'est beaucoup : comptez combien de cycles à 4GHz il faut pour booter Linux en 20s).

On considère l'ensemble du Pentium comme un gros automate selon la figure 6.3. Dans cette figure, le registre d'état contient tous les registres de la boîte à registres, tous les registres de pipeline, et même tous les registres de l'automate qui contrôle le Pentium. L'état du processeur, quoi.

On ajoute un tout petit peu de circuitterie pour faire de l'ensemble des 10 000 registres binaires un unique immense registre à décalage. C'est une transformation automatique qui est décrite par la figure ci-dessous :

Le registre d'état de la figure 6.3, avant...      ... et après sa transformation en JTAG



Avec tout cela, on peut, en 10 000 cycles, mettre le circuit dans un état quelconque. On peut également, toujours en 10 000 cycles, lire l'état complet du processeur.

Une fois ceci en place, on teste le circuit comme ceci :

- On met *test* à 1, puis on pousse un état connu, pas forcément utile, dans le processeur.
- Puis on met *test* à 0, et on fait tourner le processeur pendant quelques centaines de cycles.
- Il fait sans doute n'importe quoi, mais ce n'est pas grave.
- On remet *test* à 1, et on sort l'état complet du processeur (tout en poussant un nouvel état à sa place).
- On compare l'état obtenu avec l'état dans lequel doit être le processeur si chacune de ses portes fonctionne correctement (obtenu par simulation).
- Et on recommence plusieurs fois, avec des états construits pour faire fonctionner tous les transistors de l'énorme fonction *T* – pas forcément des états dans lequel le processeur peut se trouver en fonctionnement normal.

Tout ceci est même normalisé par le Joint Test Action Group que vous avez peut-être croisé sous le nom de JTAG.

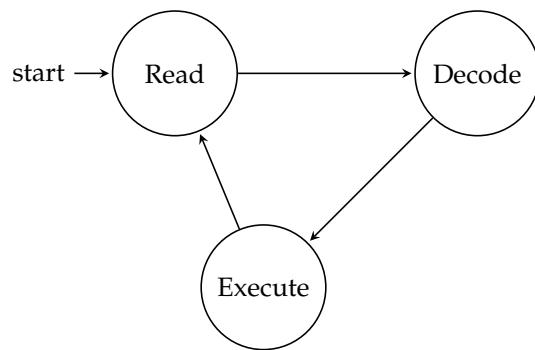


FIGURE 6.4 – Tiens, au fait, voici un automate simplifié pour le cycle de von Neumann. On le complexifiera dans la suite, par exemple parce que l'état "execute" sera décomposé en plusieurs sous-états, et le chemin pris dépendra de l'instruction à exécuter.

Mais ce n'est pas tout : une fois le circuit construit et emballé, on peut toujours utiliser ces nouvelles broches. En utilisant ce mécanisme, on peut observer ou changer la valeur de n'importe quel registre du processeur en quelque dizaines de milliers de cycle : on lit l'état, on change les bits qu'on veut, et on réécrit l'état modifié.

C'est comme cela que vous pouvez, en TP, observer dans un debugger ce qui se passe à l'intérieur d'un processeur ou d'un FPGA, à condition d'y avoir branché une "sonde JTAG".

#### 6.4.2 Equivalence de circuits

Un autre problème, plus ou moins théorique, qui bénéficie de la vue d'un circuit comme un automate est l'équivalence de circuits. On dit que deux automates sont équivalents si, vu comme des boîtes noires, ils sont indiscernables. Formellement, pour toute suite de transitions sur les entrées, les deux automates donnent la même suite de transition sur les sorties.

### 6.5 Conclusion : l'ingénierie des automates

En résumé, tout automate peut être traduit en un circuit synchrone d'une infinité de manière. Tout circuit synchrone peut aussi être vu comme un automate d'une infinité de manière.

Étant donnée la spécification d'un fonctionnement, il y a de bonnes et de mauvaises manières de dessiner un automate qui l'implémente. Un bon automate aura le moins d'états possibles, ou bien une fonction de transition simple. Il faut commencer par décomposer un fonctionnement compliqué en plein de petits automates simples, les entrées des uns étant les sorties des autres. Voir la montre à quartz.

L'encodage des états est important, et il faut y penser dès avant les patates. Par exemple on ne dessinera pas les  $2^n$  patates pour le compteur, on dessinera directement la boîte noire parce qu'on sait l'implémenter intelligemment.. Pour un automate plus aléatoire, on pourra chercher les encodages des états qui rendent la fonction de sortie  $F$  triviale.

Bien qu'on ait des outils presse-bouton qui transforment les patates en circuit, il faut donc quand même de bons ingénieurs pour bien les utiliser.

**Parenthèse culturelle : il n'y a pas des automates qu'en circuit et en UML** *Les grands théoriciens des automates n'ont pas été les architectes mais les théoriciens du langage. En effet, dans vos compilateurs, c'est un automate qui reconnaît un mot clé (il ressemble vraiment au digicode). C'est aussi un automate qui reconnaît un nombre entier, un nombre en virgule flottante, un nom de variable. C'est même un automate (d'un type un peu plus compliqué) qui valide la syntaxe de tout votre programme. En théorie des langages, les automates n'ont pas vraiment d'action à réaliser, mais ils ont des états finaux (ou acceptants) et leur seule entrée est le prochain caractère du langage à reconnaître. De plus ils ne sont pas synchrones : ils réagissent au caractère suivant de la séquence d'entrée, sans attendre un front d'horloge. En fait ils n'ont pas d'horloge.*

*Nos automates en archi sont différents surtout en ce qu'ils auront des actions à faire (des sorties), et qu'on ne distingue pas certains états comme étant finaux/acceptants : ils tournent à l'infini.*

*Vous verrez tout cela dans le cours "Grammaires et langages" en 4IF.*



# Chapitre 7

## Transmettre

### 7.1 Medium

Fil électrique, fréquence radio, fréquence lumineuse, rail avec des billes.

Notion de canal : vision logique du medium.

Notion de multiplexage : on envoie plusieurs canaux logiques sur un canal physique. Multiplexage temporel, multiplexage en fréquence...

Notion de débit (quantité d'information par unité de temps) et de bande passante d'un canal (quantité d'information pouvant passer en même temps dans le canal).

### 7.2 Liaison point a point

#### 7.2.1 Série ou parallèle

Utilisation du temps, ou utilisation de l'espace. En général on combine intelligemment les deux : sur une liaison parallèle on envoie tout de même les données en série !

Exemples : RS232, I2C, USB pour le pur série.

Jadis il y avait un "port centronics" à l'arrière des ordinateurs, il avait plein de broches parce que c'était un port parallèle qui transmettait 8 bits par cycle. De nos jours, on n'a plus que des bus série (*universal serial bus*) pour deux raisons toutes bêtes : 1/ il y qu'un fil de donnée, c'est moins cher et 2/ il n'y a pas à synchroniser plusieurs fils, c'est moins chiant.

Il reste des liaisons parallèles en fond de panier de votre PC (entre le processeur et la mémoire). La synchronisation de tous ces fils est toute une poésie.

#### 7.2.2 Protocoles

Si on n'a qu'un seul canal binaire pour transmettre des données entre un émetteur et un récepteur, tout ce qu'on sait faire c'est faire passer ce canal de 1 à 0 puis de 0 à 1. Si on enlève l'information temporelle, on n'observe que 101010101010101...

Une communication suppose donc une *synchronisation* (partage du temps en grec dans le texte).

Une solution est qu'émetteur et récepteur aient chacun une horloge suffisamment précise, et se soient mis d'accord à l'avance sur les instants auxquels telle et telle donnée sont transmises. Ceci s'appelle un protocole de communication.

Une solution plus simple est d'avoir deux canaux binaires (au moins) entre émetteur et récepteur. L'un pourra faire passer le tic-tac d'une horloge, par exemple. Avec un seul canal mais au moins ternaire, c'est aussi possible. Mais dans tous les cas il faudra un protocole qui définit comment les données sont emballées, comment commence un paquet de donnée, etc.

#### Maître et esclave

Notion de maître (celui qui donne les ordres) et d'esclave (celui qui obéit). En vrai, les deux sont

1. totalement esclaves du protocole (sinon la communication ne se fait pas), et
2. pour certains protocoles qu'on va voir, totalement libres de traîner aussi longtemps qu'ils veulent

Bref, le maître est juste celui qui est à l'initiative de la communication.

Le maître peut tourner, mais à un instant donné il vaut mieux qu'il n'y ait qu'un seul maître.

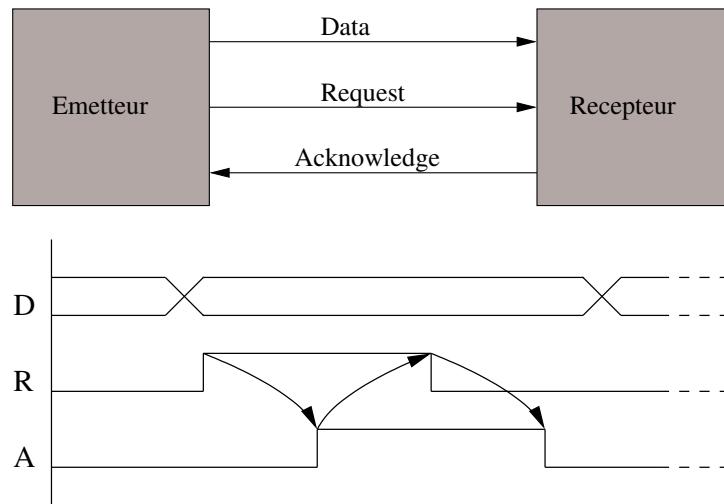


FIGURE 7.1 – Deux entités papotent en utilisant un protocole handshake. La donnée est une information d'état, alors que R et A portent des informations temporelles (changement d'état). Les flèches indiquent une causalité : elles sont fatidiquement de gauche à droite. En termes d'implémentations, Récepteur et Émetteur sont des automates : amusez vous à les dessiner.

**Remarque du 21ème siècle** Les protocoles maîtres / esclaves c'est super alors que l'esclavage c'est nul. Pour ne pas faire d'association positive à l'esclavage, ces derniers temps on essaye de remplacer les termes de maître et esclave par d'autres termes : il y a toute une poésie que vous irez lire sur wikipedia.

Et en attendant le poly reste sur une terminologie qui ne fait l'apologie de l'esclavage que lorsque ce sont des circuits logiques qui sont esclaves. Car rappelons-le, l'esclavage c'est mal.

## Protocoles synchrones

Utilisation du temps.

Exemple 1 : envoi d'une donnée de maître à esclave, au moyen de deux canaux. Le maître positionne la donnée sur l'un, puis positionne l'autre canal, que nous appellerons "donnée prêté" et qui valait 0, à 1 pendant 1 pataseconde, et puis de nouveau à zéro. Il sait que l'esclave lui obéira et aura lu la donnée au bout d'une pataseconde : c'est son esclave.

Exemple 2 : réception d'une donnée par le maître. Le maître positionne un signal "demande de donnée". Au bout de 1 pataseconde, il lit la donnée sur l'autre canal. Il sait que l'esclave aura obéi dans ce laps de temps.

Exemple 3 : deux fils dont un est un tic tac, envoi d'octets commençant par 10 (pour marquer le début d'un octet) puis 8 bits, puis parité pour la détection d'erreur.

Exemple 4 : Manchester encoding (ethernet) : 0 est codé par 0 pendant une pataseconde puis 1 pendant une pataseconde, 1 codé par 1 puis 0. Ainsi le recalibrage de l'horloge est inclus.

Ce type de protocole ne marche pas avec les circuits Insaliens. D'une part ils sont toujours en retard quand ils sont esclaves, d'autre part ils veulent tous être maître en même temps. C'est pourquoi on a mis au point des protocoles asynchrones.

## Protocoles asynchrones

Protocoles qui ne font pas d'hypothèse sur les durées de transmission des signaux.

Archétype : protocole *handshake* pour "secouons nous les mains" en Breton. Il faut trois canaux : un de données (D), un de demande (R comme *request*), un d'acquittement (A comme *acknowledge*) (voir figure 7.1).

Au repos, tout le monde est à 0.

Scénario 1 : une transmission pilotée par l'émetteur (c'est-à dire que l'émetteur est le maître). L'émetteur écrit sur R et D et lit sur A, le récepteur c'est le contraire.

- émetteur : "Voici une donnée" (positionne D, puis lève R, et le laisse à 1 jusqu'à nouvel ordre).
- récepteur : voit R levé, range la donnée ou il faut, puis dit "Bien reçu Chef" (lève A, et le laisse levé jusqu'à nouvel ordre).
- émetteur : voit A se lever, dit "Brave petit" (baisse R, puis attend la baisse de A avant de recommencer).
- récepteur : voit R se baisser, dit "ce fut un plaisir Chef" (baisse A et attend un nouvel ordre).

Attention, ici l'émetteur a l'initiative de la transmission, mais n'est pas maître de tout. Par exemple, il ne peut pas baisser R tant que le récepteur ne lui autorise pas en levant A.

On peut faire un handshake à l'initiative du récepteur (c'est-à-dire que c'est le récepteur le maître) : R signifie à présent "envoie une donnée SVP". C'est le récepteur qui écrit sur R et lit sur A (et toujours sur D), et l'émetteur écrit toujours sur D mais lit sur R et écrit sur A.

Au repos tout le monde est à zéro.

- Récepteur lève R.
- Émetteur positionne la donnée, puis lève A.
- Récepteur traite ou range la donnée, puis baisse R.
- Émetteur baisse A.
- Récepteur voit A descendre, et sait qu'il peut demander une nouvelle donnée.

Il y a quand même une hypothèse temporelle derrière ce protocole, c'est que quand on écrit "émetteur positionne la donnée puis lève A", cet ordre temporel sera respecté à l'arrivée au récepteur. Si le système physique respecte cette hypothèse, on peut d'ailleurs aussi bien envoyer les données par paquets en parallèle, ce qui réduit le coût de l'asynchronisme. C'est comme cela que le processeur parle avec la mémoire dans votre PC (la mémoire peut avoir des tas de bonnes raisons de ne pas répondre tout de suite quand le processeur lui demande une donnée, on verra lesquelles).

Voyons maintenant un protocole vraiment insensible au délai. L'idée est d'utiliser un code dit double-rail pour la donnée : un bit est transmis sur deux canaux, avec la signification suivante : 1 est codé par 01, 0 par 10, la situation de repos est codé par 00. Ainsi mettre une donnée sur le double-rail change obligatoirement au moins un bit, ce qui est détecté par le récepteur, et tient donc lieu soit de R (dans le protocole piloté par l'émetteur) soit d'A (dans le protocole piloté par le récepteur).

Loi de conservation des emmerdements : il faut dans ce cas vraiment deux canaux pour transmettre un seul bit.

Tout se complique si on suppose des erreurs possibles sur la ligne. Solution : encore plus de protocole. Vous avez déjà vu ce qu'il faut mettre dedans.

### 7.3 Bus trois états

On a déjà vu page 48 la porte 3 états pour construire des registres. Elle est en général utilisée pour partager un canal entre plusieurs émetteurs.

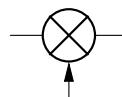


FIGURE 7.2 – Porte de transmission, ou porte "trois états"

Il faut alors un *arbitrage* qui garantit qu'à un instant donné un seul émetteur écrit sur le canal (sinon : situation logique indéterminée, qui se traduit souvent par une forte odeur de silicium fondu).

Prenons l'exemple de notre machine de von Neumann avec son bus de donnée bidirectionnel.

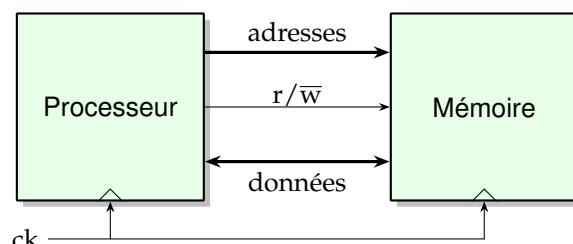


FIGURE 7.3 – On nous ment depuis le début, sans le fil  $r/\bar{w}$  et sans horloge cela ne pouvait pas marcher

En pratique ce n'est pas difficile d'assurer mécaniquement que lorsque P émet, M reçoit, et inversement : il suffit que leurs portes trois états respectives soient commandées par le même signal (enfin, des signaux complémentaires). Par exemple le signal  $r/\bar{w}$  de la figure 4.7 p. 46 (à droite) est certe utilisé par la RAM pour savoir si le bus Data va dans un sens ou dans l'autre, mais il est aussi utilisé en interne par le composant branché à l'autre bout de ce bus Data (au hasard, un processeur), composant qui est à l'origine de ce signal  $r/\bar{w}$ .

On en revient à cette subtilité psychologique dans la notion de maître/esclave : le maître est certes à l'origine du signal  $r/\bar{w}$ , mais il ne fait pas n'importe quoi ensuite, il reste complètement esclave du protocole.

En général, un bus est un moyen de transport partagé... ici c'est un fil. Exercice : dessinez un processeur relié à *deux* mémoires par un bus de données, et construisez la logique qui va arbitrer entre ces deux mémoires.

## 7.4 Réseaux en graphes (\*)

En supposant réglée la question de transmettre des données sur un canal, on veut à présent relier entre eux  $n$  circuits (ou ordinateurs) par des canaux. Pour simplifier, disons que tous ces canaux sont bidirectionnels et identiques (même débit etc). Le réseau a alors une structure de graphe, les canaux étant les arêtes et les circuits étant les nœuds.

### 7.4.1 Les topologies et leurs métriques

Considérons deux topologies possibles de réseaux pour relier  $n$  ordinateurs : l'anneau, représenté figure 7.4, et le tore bidimensionnel, représenté figure 7.6 qui est une extension de la grille bidimensionnelle (représentée figure 7.5) dans laquelle tous les nœuds ont le même nombre de canaux.

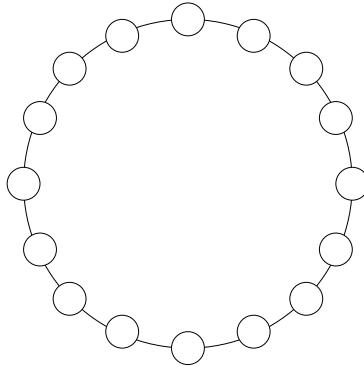


FIGURE 7.4 – Graphe anneau

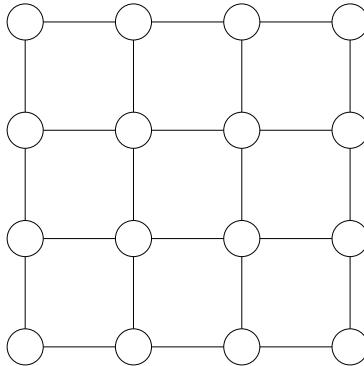


FIGURE 7.5 – Graphe grille 2D

On peut définir sur ces deux exemples les notions suivantes :

**Degré** Le degré d'un nœud est le nombre d'arêtes reliées à ce nœud. Par extension, le degré d'un graphe est le degré maximum des nœuds du graphe. Par exemple le degré est de 2 pour l'anneau, et de 4 pour la grille et le tore. A priori, plus le degré sera élevé, plus le nœud sera cher. Plus le degré est petit mieux c'est.

**Diamètre** Le diamètre d'un graphe est le maximum de la longueur du plus court chemin entre deux nœuds. Dans nos réseaux à  $n = 16$  nœuds, le diamètre de l'anneau est 8, le diamètre de la grille est 6, le diamètre du tore est 4. Le diamètre mesure la distance maximum à parcourir pour une information dans le réseau. Plus il est petit mieux c'est.

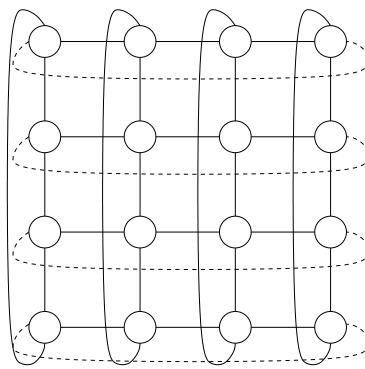


FIGURE 7.6 – Graphe tore 2D

**Bisection** La bisection est une mesure de la quantité d'information qui peut passer à travers le réseau à un instant donné : Plus elle est élevée, mieux c'est. Techniquement, on appelle bisection-arêtes (resp. bisection-sommets) le nombre minimum de liens (resp. de sommets) dont la destruction entraîne la déconnexion en deux moitiés de même taille (à un près).

Tore et anneau ont tous deux des degrés constants, mais le tore a une bisection et un diamètre en  $\sqrt{n}$ , ce qui semble mieux que l'anneau (bisection de 2 et diamètre en  $n/2$ ). On peut définir un tore tridimensionnel et plus (et au fait l'anneau est un tore unidimensionnel). Mais d'autres topologies offrent des compromis encore plus intéressants :

L'hypercube de degré  $d$  a  $2^d$  nœuds, un diamètre de  $d$ , un bisection de  $2^{d-1}$ . Son seul défaut est un degré qui peut devenir relativement élevé.

C'est pourquoi on a inventé le Cube Connectant des Cycles ou CCC, représenté figure 7.8 en dimension 3. On place des anneaux de taille  $d$  à chaque sommet d'un hypercube de degré  $d$ , et chaque nœud d'un anneau possède en plus un canal selon une des dimensions de l'hypercube. Le degré est 3 quelle que soit la dimension, et le diamètre est passé de  $d$  à  $d^2$  : c'est toujours logarithmique en le nombre de nœuds.

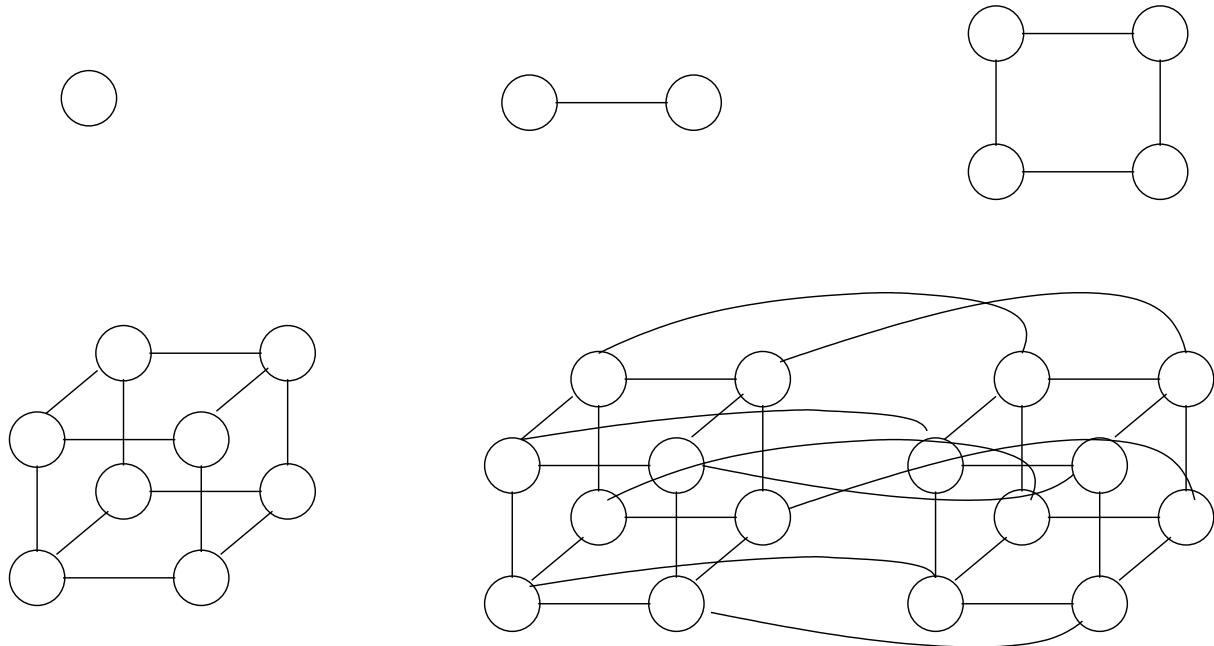
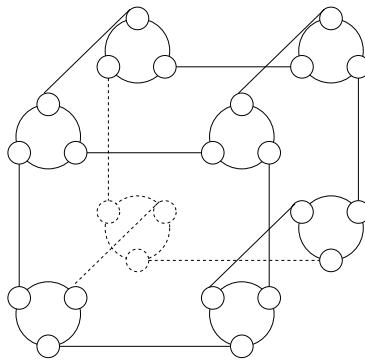


FIGURE 7.7 – Graphes hypercubes

#### 7.4.2 Routage

Le routage c'est le choix du chemin que va suivre l'information dans le graphe.

FIGURE 7.8 – *Graph cube-connected cycles*

Dans un circuit intégré le routage est surtout statique : on a tiré des fils d'un bloc logique A à un autre bloc logique B, et une transmission d'info de A à B passera par ces fils<sup>1</sup>.

Dans un grand réseau (téléphonique ou internet), par mesure d'économie, on n'a pas un fil par communication possible. On a tout de même plusieurs chemins possibles. Le routage d'une transmission d'information est calculé lorsqu'on en a besoin. On parle de routage dynamique. Bref, un routage dynamique, c'est trouver un chemin dans un graphe. Alors que les algos qui calculent le routage d'un circuit intégré peuvent se permettre de tourner des heures, pour un routage dynamique il faut un algo assez rapide.

C'est assez facile dès qu'on a un seul émetteur et un seul destinataire, on trouve alors toujours un algorithme glouton (exo : donnez le sur la grille, sur l'hypercube, sur le CCC).

Quand on a plein de paires (émetteur, destinataire) en même temps, le problème du routage devient difficile. Pourquoi ? Pourquoi c'est dur ?

- Il faut éviter que deux données se battent pour le même fil en même temps.
- Idéalement, chaque info va suivre le plus court chemin. Mais déjà, trouver le plus court chemin de A à B dans un graphe est coûteux.
- Quand le plus court chemin est impraticable (un lien est déjà occupé), il faut faire un détour. A force de détours on n'a plus un plus court chemin. Il faut alors montrer que l'on reste dans un facteur raisonnable du plus court chemin (ou du diamètre du graphe), sinon non seulement la communication est plus lente, mais en plus elle consomme encore plus de fils, ce qui aggrave le problème pour les suivants.

Dans ces domaines il est facile de bricoler des heuristiques qui ont l'air de marcher bien, mais il est très dur de montrer qu'elles offrent certaines garanties.

Les métriques du graphe deviennent importantes.

Il y a deux grandes familles de techniques de routage : la *commutation de ligne* (pensez au téléphone de l'URSS) et la *commutation de paquet* qui est ce qui se fait sur internet.

- Pour la commutation de ligne, il faut a priori un central qui supervise le réseau et réserve la ligne.
- Dans la commutation de paquet, un message est coupé en paquets de taille plus ou moins fixe, et ces derniers sont lancés dans le graphes et doivent y trouver de proche en proche leur chemin. Le gros avantage est que les algorithmes de routage deviennent décentralisés.

Remarque : on peut faire de la commutation de ligne virtuelle par dessus de la communication de paquets. Par exemple, on veut pouvoir réserver un canal virtuel de A à B, avec un certain débit garanti pour y faire passer de la vidéo en temps réel.

Deux techniques de commutation de paquets : store and forward, et wormhole qui est un genre de réservation de ligne. Les anneaux du corps du vers s'appellent des flit. Vous avez intérêt à avoir noté mes explications au tableau.

Il faut éviter les situations d'interblocage : la communication de A vers B attend pour se terminer que la communication de C vers D libère un certain fil. Mais la communication de C vers D attend elle-même pour libérer ce fil que la communication de A vers B libère un autre fil. Montrer qu'un algorithme de routage ne mènera jamais à un interblocage est en général très difficile.

Dans IP (internet protocol), technique sommaire qui garantit l'absence d'interblocage, et aussi évite qu'un paquet tourne à l'infini dans le réseau sans jamais arriver : chaque paquet part avec une durée de vie, qui est décrémentée à chaque passage dans un noeud. Quand la durée de vie arrive à zéro, le paquet

1. C'est vrai pour les petits circuits : un multiplicateur, un processeur. De nos jours on est capable de concevoir des circuits intégrants des dizaines de processeurs, de mémoires et d'unités de calcul. Si l'on relie tous ces blocs par du routage statique, on constate que la plupart de ces fils sont inactifs la plupart du temps, c'est du gâchis de surface. Il y a donc beaucoup de recherche actuellement sur les *Networks on Chip*, ou NoC, dans lesquels les fils seront partagés et le routage dynamique.

est détruit sans autre forme de procès. Eh, oh, et mes données ? Ben, tu les renverras... si le paquet qui te prévient de l'exécution de ton paquet t'arrive... Oui, les protocoles deviennent compliqués. On parlera des protocoles de l'internet quand on aura construit notre ordinateur.

En attendant, exemples de preuve que pas d'interblocage et pas de famine :

- anneau (protocole Token Ring). Par exemple, au fond de la PléStation3, il y a deux anneaux tournant en sens inverse.
- routage 2D Manhattan,
- routage hypercube.

### 7.4.3 Types de communication : point à point, diffusion, multicast

## 7.5 Exemples de topologies de réseau (\*)

### 7.5.1 Le téléphone à Papa

Commutation de ligne, protocole très centralisé ("central téléphonique").

### 7.5.2 L'internet

Couche physique : ethernet.

Architecture en graphe sans vraiment de structure. Hiérarchie à deux niveaux : *local area network* ou LAN, *wide area network* ou WAN.

Couches logiques : commutation de paquets, routage par store and forward, protocole décentralisé.

### 7.5.3 FPGAs

Grille 2D, commutation de ligne par *crossbar*, routage statique.

### 7.5.4 Le bus hypertransport

Allez lire la page Wikipedia, c'est votre collègue Nicolas Brunie qui l'a écrite, je me demande s'il me le pardonnera un jour.

### 7.5.5 Machines parallèles

On ne dit plus machines parallèles, on dit Ze Grid. Pardon, je me reprends. On ne dit plus Ze Grid, on dit cloud computing. Bref. On assemble actuellement les coeurs sur la puce par un bus, les puces sur une carte par un autre bus, les cartes entre elles par des réseaux simples (la dernière fois que j'ai regardé c'était des anneaux), et les armoires ainsi obtenues par l'Internet anarchique.

Pour les détails j'ai demandé à Loris Marchal qui a demandé à Jean-Yves l'Excellent : *La réponse plus complète, c'est que sur des machines parallèles du genre silicon graphics, il y a encore des topologies marrantes (hypercubes ? fat-tree,...). Sur des clusters, en général il y plutôt des crossbar quand c'est pas trop gros (par exemple : earth simulator : cross bar entre 640 noeuds) et quelque chose de hiérarchique pour les trucs plus récents et plus gros. D'autre part, il y a parfois plusieurs types de réseaux disponibles. Pour Blue Gene, il y a 3 réseaux : un pour les communications collectives synchrones (barrières et reductions), un pour les communications point à point et asynchrone, et un pour le monitoring et le système. Chacun utilise une topologie (et une technologie) différente.*

Si vous n'avez pas tout compris dans cette section c'est normal et c'est pas grave.



**Deuxième partie**

**Machines universelles**



# Chapitre 8

## Jeux d'instruction

### 8.1 Rappels

Une machine de von Neumann c'est un *processeur* relié à une *mémoire*.

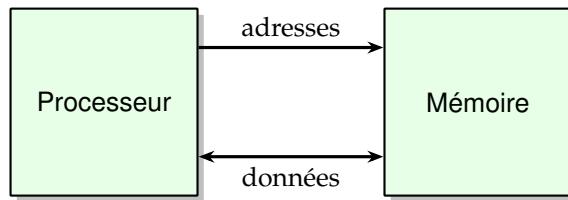


FIGURE 8.1 – *Mon premier ordinateur*

La mémoire est adressable par mot (de 8, 16, 32 ou 64 bits de nos jours), et grande (de nos jours, de l'ordre de  $10^9$ , ou  $2^{30}$  mots). La mémoire est non typée : chaque mot peut être interprété de multiples manières.

Le processeur exécute le **cycle de von Neumann**. Pour cela, il garde dans un coin une adresse qu'il appelle PC (*program counter*). Le cycle, réalisé de nos jours dans les  $10^9$  fois par seconde, est le suivant :

1. lire le contenu de la cellule à l'adresse PC
2. l'interpréter comme une instruction, et l'exécuter
3. Augmenter PC pour passer à l'instruction suivante, et recommencer

Une **instruction** est donc l'unité de travail que peut réaliser un processeur.

Définir un **jeu d'instruction** c'est définir la manière dont le processeur interprète  $n$  bits comme une instruction. C'est le travail du constructeur du processeur, et la courte histoire de l'informatique est pleine d'idées rigolotes dans ce domaine, pour  $n$  allant de 4 à 256.

Un bon jeu d'instruction est un jeu d'instruction

1. universel (ou Turing-complet, ou qui permet d'exprimer tout programme exprimable),
2. pour lequel il sera facile de construire le processeur qui l'exécute, et
3. avec lequel il sera facile d'écrire des programmes.

Le point 1 est facile à assurer, il suffit d'exhiber un programme qui simule la machine de Turing qui vous arrange. Par contre, il y a une Loi de Conservation des Emmerdements qui dit que les points 2 et 3 sont antagonistes.

### 8.2 Vocabulaire

Dans la suite on appelle les instructions du processeur **instruction machines**. Une instruction sera codée par un paquet de bits. Dans les processeurs RISC (*reduced instruction set computers*), toutes les instructions sont de la même taille, correspondant à un mot mémoire, ce qui simplifie leur décodage. Nous allons construire un tel processeur.

Dans certains processeurs dits CISC (*complex instruction set computers*), il peut y avoir des instructions de tailles (en bits) différentes. Le cycle de von Neumann est identique sur le fond. Nous allons observer

un tel processeur, le Pentium de mon portable. Je changerai de police pour évoquer les processeurs CISC, pour ne pas que leur complexité ne pollue trop mon propos.

Pour les besoins de l'interface avec les humains, chaque instruction (jusque là le contenu d'une case mémoire, donc un vecteur de bits) aura également une représentation textuelle non ambiguë, que l'on appelle **mnémonique**. L'ensemble des mnémoniques forme le **langage assembleur**. Le programme qui convertit le langage assembleur en séquence d'instructions binaires s'appelle aussi un **assembleur**.

## 8.3 Travaux pratiques

Les TP MSP430 sont là pour vous montrer tout cela pour un processeur RISC moderne (bien que petit). Voici deux autres exemples.

### 8.3.1 Le jeu d'instruction de votre PC

Tout ce qui suit est à essayer sur un PC sous Linux<sup>1</sup>. Prenez un programme C quelconque `toto.c` par exemple celui-ci :

```
main() {
    int i=17; /* une constante facile à retrouver dans l'assembleur */
    i=i+42; /* une autre */
    printf("%d\n", i);
}
```

Compilez-le avec

```
gcc -S toto.c
```

Vous obtiendrez un fichier `toto.s` en langage assembleur (il ne contient que des mnémoniques). Ici il s'agit de l'assembleur du Pentium. Voici les extraits de ce fichier qui sont lisibles à ce stade du cours.

```
(... de la paperasse administrative ...)
movl $17, -4(%rbp)
addl $42, -4(%rbp)
    (... des instructions pour passer les paramètres à printf )
call printf
    (... encore de la paperasse )
```

La première ligne met (*move*) la constante 17, considérée comme un entier long (suffixe `l`), quelquepart en mémoire. La seconde ligne ajoute 42 à la même case mémoire.

Il existe en fait un assembleur dans la suite `gcc` : vous pouvez donc vous amuser à modifier le fichier `toto.s` puis le transformer en un exécutable (un fichier dans lequel chaque instruction est codée en binaire) par `gcc toto.s`.

On a observé de l'assembleur dans `toto.s`. Le binaire correspondant c'est `a.out`. On peut tenter de l'ouvrir dans un éditeur de texte : il est illisible (sauf si vous utilisez un éditeur de texte de hackerz). Une manière de l'observer est la commande `objdump -d a.out` (le `-d` signifie *disassemble*, donc faire le travail inverse de l'assembleur). Vous obtenez des mnémoniques, et également le code binaire (enfin, hexadécimal) de chaque instruction.

Voici les lignes correspondant au code assembleur précédent :

Adresses	Instructions binaires	Assembleur
(...)		
40052c: 55		push %rbp
40052d: 48 89 e5		mov %rsp,%rbp
400530: 48 83 ec 10		sub \$0x10,%rsp
400534: c7 45 fc 11 00 00 00		movl \$0x11,-0x4(%rbp)
40053b: 83 45 fc 2a		addl \$0x2a,-0x4(%rbp)
40053f: 8b 45 fc		mov -0x4(%rbp),%eax
400542: 89 c6		mov %eax,%esi
% (...)		

On a des instructions de toutes les tailles. C'est vraiment n'importe quoi ce processeur, ils ne vont pas en vendre beaucoup.

1. Par la magie du Net on peut aussi désormais utiliser <http://gcc.godbolt.org/> qui permet de choisir le compilateur et le processeur. Pour observer le travail du compilateur c'est mieux (enfin à condition de fermer les yeux sur la gabegie énergétique d'envoyer la compilation au diable dans un nuage). Pour jouer directement avec l'assembleur c'est moins bien.

### 8.3.2 Le jeu d'instruction de votre téléphone portable

Pour le même programme, en utilisant des outils de *cross-compilation* vers ARM (par exemple le paquet `gcc-arm-linux-gnueabihf` sous Debian/Ubuntu). On tape à présent `arm-linux-gnueabihf-gcc -marm -S toto.c`. Vous obtiendrez un fichier `toto.s` contenant entre autres les mnémoniques suivantes :

```
(... de la paperasse administrative ...)
mov r3, #17
add r3, r3, #42
    (... des instructions pour passer les paramètres à printf )
bl printf
    (... encore de la paperasse )
```

Et voici les résultat de

```
arm-linux-gnueabihf-gcc -marm toto.c;
arm-linux-gnueabihf-objdump -d a.out
```

```
000083fc <main>:
83fc: e92d4800  push {fp, lr}
8400: e28db004  add fp, sp, #4
8404: e24dd008  sub sp, sp, #8
8408: e3a03011  mov r3, #17
840c: e50b3008  str r3, [fp, #-8]
8410: e51b3008  ldr r3, [fp, #-8]
8414: e283302a  add r3, r3, #42 ; 0x2a
8418: e50b3008  str r3, [fp, #-8]
841c: e3080488  movw r0, #33928 ; 0x8488
8420: e3400000  movt r0, #0
8424: e51b1008  ldr r1, [fp, #-8]
8428: ebffffae  bl 82e8 <_init+0x20>
842c: e1a00003  mov r0, r3
8430: e24bd004  sub sp, fp, #4
8434: e8bd8800  pop {fp, pc}
```

Vous observez que chaque instruction est encodée sur 32 bits tout pile. C'est quand même plus régulier : ce sera plus simple à construire et cela consommera moins.

*Bonus : plus tard, les gens qui faisaient l'ARM se sont rendu compte que ce code consommait beaucoup de mémoire, et ont défini un nouvel encodage des instructions dans lequel les instructions qu'on peut encoder en 16 bits sont encodées en 16 bits... C'est désormais l'encodage par défaut, et si vous ne donnez pas l'option `-marm`, c'est celui que vous observerez.*

## 8.4 Instruction set architecture

La définition du jeu d'instructions est décisive : ce sera l'interface entre le hard et le soft.

- Le programmeur (ou le compilateur) ne connaît que ces instructions machines et leur sémantique bien documentée dans de gros manuels de référence. Il ne sait pas comment elles sont implémentées, et s'en fiche pas mal. D'ailleurs cela change tout le temps.
- Le concepteur du pentium 12 doit construire un processeur qui implémente ce jeu d'instruction et sa sémantique. À part cela, il a toute liberté.

On parle souvent d'ISA pour *instruction-set architecture*. Des exemples d'ISA sont IA32 (implémenté par les vieux processeurs Pentium, Athlon, etc), son descendant 64-bit (celui de votre ordinateur) qu'Intel appelle EMT64 et Linux appelle AMD64 (car il a été proposé par AMD), IA64 (implémenté par les processeurs Itanium), Power (implémenté par les processeurs IBM Power très chers, et les processeurs PowerPC meilleur marché), ARM, MSP430, ... Il y a aussi une ISA open-source qui s'appelle RiscV (lire "risc five").

L'exemple d'EMT64/AMD64 montre combien l'ISA est déconnecté du processeur l'implémentant : il n'y a pas un transistor de commun entre les processeurs d'Intel et leurs concurrents d'AMD, et pourtant ils peuvent exécuter les mêmes programmes. On discutera plus bas la relation entre ISA et architecture.

En pratique, les ISA évoluent, et les grandes familles évoquées ci-dessus se déclinent en versions, ajoutant des instructions au jeu d'instructions de base. Exemples (plus ou moins connus grâce à un

marketing plus ou moins efficace) : les extension multimedia à IA32, SSE/SSE2/SSE3/... chez Intel<sup>2</sup>, 3DNow chez AMD, ou encore Thumb, Jazelle, NEON et VFP chez ARM, ...

Il est important de savoir bien distinguer les noms marketing des ISA de ceux des processeurs. Par exemple, ARMv7 est un ISA, alors que ARM7 est un processeur (qui n'implémente pas du tout ARMv7). AMD64 et EMT64 sont, à peu de choses près, des noms marketings différents pour la même ISA. Allez bouquiner sur Wikipedia les pages sur ARM et IA32, et faites cet exercice.

## 8.5 Que définit l'ISA

Un programme typique c'est une boucle qui va chercher des opérandes en mémoire, calcule dessus, puis les range en mémoire. On va définir des instructions pour tout cela.

### 8.5.1 Types de données natifs

Commençons par les opérations de calcul. L'ISA définit avant tout un certain nombre de **types de données** sur lesquels elle aura des opérateurs de calcul matériels. Les types typiques sont

- des adresses mémoire. On a besoin d'y lire et d'y écrire.
- des champs de bits, sur lequel on peut faire des ET, des OU, des XOR, des décalages...
- des entiers de différentes tailles (8, 16, 32, 64 bits pour l'IA32; 8 et 32 pour l'ARM; 32 pour l'Alpha, etc). Il y a deux sous-types, les entiers signés et non signés. Grâce au complément à 2, ils se distinguent uniquement par la gestion des débordements et des décalages. Autrement dit, la même instruction d'addition marche aussi bien pour des entiers signés et non signés (c'est l'addition modulo  $2^p$ ). Par contre il faut des instructions différentes pour les comparaisons de deux entiers : par exemple sur 4 bits, le test  $1111 > 0001$ ? renvoie vrai si on considère que ce sont des entiers non signés, et faux si on est en complément à 2.
- des nombres en virgule flottante : simple précision (32 bits) et double précision (64 bits). Ces formats sont imposés par un standard, appelé poétiquement IEEE 754, dont la version courante date de 2019. Ce standard définit également les précisions *quadruple* (128 bits) et *half* (16 bits). IA32 et IA64 ajoutent des formats à 80 bits, IA64 ajoute un format à 82 bits...

Les extensions multimedia, introduites dans les années 90, définissent des vecteurs de petits entiers (MMX) et des vecteurs de petits flottants (3DNow, SSE, AVX<sub>y</sub>). Pour l'ISA ARM, l'équivalent s'appelle NEON.

### 8.5.2 Instructions de calcul

L'ISA définit ensuite des **instructions de calcul** pour ces différents types de données : addition, multiplication, décalage, OU logique... Nous allons filer l'exemple de l'addition, pour les autres c'est très similaire.

Il faut, dans une instruction d'addition, pouvoir spécifier ce qu'on additionne (les opérandes) et où on le range (la destination). Idéalement, on aimerait donner trois adresses dans la mémoire. Une instruction serait alors "ajouter le contenu de la case mémoire numéro X au contenu de la case mémoire numéro Y et ranger le résultat dans la case numéro Z, les trois contenus étant interprétés comme des entiers 32 bits". Pour un processeur moderne qui peut adresser une mémoire de  $2^{32}$  mots, l'information totale des trois opérandes nécessiterait déjà  $3 \times 32 = 96$  bits. Si je veux coder chaque instruction dans un mot de 32 bits c'est mal parti.

C'est pourquoi les ISA définissent tous une mémoire de travail plus petite, interne au processeur, que l'on appelle les **registres**. Chez les RISC, ils sont tous identiques et forment une petite mémoire adressable interne. Si l'on a 16 registres et que les opérations de calcul ne peuvent travailler que sur le contenu de ces registres (cas de l'ARM), les trois opérandes d'une instruction peuvent se coder sur  $3 \times 4 = 12$  bits seulement. Un avantage supplémentaire est que ces registres seront lus et écrits plus vite que la mémoire principale, puisqu'ils sont intégrés sur la puce du processeur.

*C'est pourquoi les ISA définissent tous une mémoire de travail plus petite, interne au processeur, que l'on appelle les registres. Chez les RISC, ils sont tous identiques et forment une petite mémoire adressable interne. Si l'on a 16 registres et que les opérations de calcul ne peuvent travailler que sur le contenu de ces registres (cas de l'ARM), les trois opérandes d'une instruction peuvent se coder sur  $3 \times 4 = 12$  bits seulement. Un avantage supplémentaire est que ces registres seront lus et écrits plus vite que la mémoire principale, puisqu'ils sont intégrés sur la puce du processeur.*

*C'est pourquoi les ISA définissent tous une mémoire de travail plus petite, interne au processeur, que l'on appelle les registres. Chez les RISC, ils sont tous identiques et forment une petite mémoire adressable interne. Si l'on a 16 registres et que les opérations de calcul ne peuvent travailler que sur le contenu de ces registres (cas de l'ARM), les trois opérandes d'une instruction peuvent se coder sur  $3 \times 4 = 12$  bits seulement. Un avantage supplémentaire est que ces registres seront lus et écrits plus vite que la mémoire principale, puisqu'ils sont intégrés sur la puce du processeur.*

2. En 2007, à 15 jours d'intervalle, Intel a annoncé l'extension SSE4 qui introduit par exemple une instruction calculant  $a*b+c*d$  en flottant avec 3 arrondis, puis AMD a annoncé l'extension SSE5 ajoutant une instruction FMA :  $a*b+c$  avec un seul arrondi. Je ne sais pas comment ils se sont partagé ces effets de manche, mais j'ai parié trois polys que les processeurs qui sortiraient 5 ans plus tard auraient tous les deux extensions. Allez voir si j'ai gagné sur [http://en.wikipedia.org/wiki/FMA\\_instruction\\_set](http://en.wikipedia.org/wiki/FMA_instruction_set), parce que l'histoire est rigolote.

par des ISA datant de la préhistoire de la paléoinformatique, avant que les vitesses de la mémoire centrale et du processeur ne fassent le grand écart,

Nous pouvons à présent définir une instruction d'addition dans un processeur RISC. Son mnémonique sera par exemple `add R0, R1 -> R3`. La forme générale pour une opération binaire sera `op Rx, Ry -> Rd`, où `op` peut être `add, sub, mul, xor, ...` et `d, x` et `y` sont les numéros des registres destination et sources. Le codage de cette instruction sera très simple : un certain nombre de bits pour coder `op`, (mettons les 8 premiers, ce qui nous permet 256 instructions différentes), un certain nombre pour coder `d`, un certain nombre pour coder `x` et `y`. Il sera d'autant plus simple de construire l'architecture que ce code est simple : on mettra le code du registre destination toujours au même endroit, et idem pour les deux opérandes.

Remarquez que les "vrais" mnémoniques sont le plus souvent écrits `op Rd, Rx, Ry` – et parfois la convention est de mettre la destination en dernier (voir `toto.s`), et on trouve même les deux conventions qui cohabitent pour le même processeur suivant les systèmes d'exploitation ! La raison en est surtout historique : les premiers mnémoniques reprenaient les champs de bits dans l'ordre dans lequel ils étaient dans l'instruction, ce qui facilitait l'assemblage – je parle d'une époque où le programme assembleur était lui-même écrit en assembleur sur des cartes perforées. Notre convention, utilisant `->` est plus lisible, et comme le mnémonique n'est qu'une représentation textuelle on la garde : il faut se convaincre que, si le choix de l'ISA lui-même fait la différence entre un bon processeur et un mauvais, le choix de ses mnémoniques, par contre, n'est qu'une convention et n'a pas grande importance.

Nous avons défini une addition dite à **trois opérandes** puisque les trois opérandes sont exprimés. Il y a d'autres solutions :

- Instructions à **2 opérandes** : `op Rd, Rx -> Rd`. On écrase l'un des opérandes.
- Instructions à **1 opérande** : idem, mais `Rd` est fixé et implicite. On l'appelle en général l'accumulateur, en hommage aux temps héroïques où les processeurs n'avaient même pas de multiplicateurs, et où l'on accumulait des séquences d'additions.
- Instructions à **pile** (0 opérande) : le processeur dispose d'une pile (comme les calculettes HP), et une instruction `add` prend ses deux opérandes au sommet de la pile et y range le résultat.

Il y a là un compromis :

- l'instruction à trois opérandes est la plus puissante, et donc la plus compacte en termes de nombres d'instructions. Elle est préférée pour les processeurs RISC récents : leur unité d'information fait 32 bits, et on se donne donc 32 bits pour chaque instruction, ce qui permet d'y coder trois opérandes voire plus (voir ci-dessous)
- Avec des instructions à 0, une ou deux opérandes, il faut souvent faire des copies des registres écrasés (c'est une autre instruction), ou bien des `swap` et des `dup` sur la pile : un programme donné a besoin de plus d'instructions. Par contre, en terme de nombre total de bits, le programme est plus compact.

Le plus compact est le jeu d'instruction à pile, puisqu'il n'y a que l'instruction à coder, pas ses opérandes. C'est la raison pour laquelle ce type d'instruction est souvent choisi pour le *bytecode* des machines virtuelles pour les langages comme Caml ou Java.

Il faut noter que dans tous les ISA, l'un des registres peut souvent être remplacé par une constante. Le mnémonique est alors par exemple `add R12, 4 -> R1`. Techniquement, c'est une instruction différente, à laquelle on peut donner un mnémonique différent, par exemple `addi` pour *add immédiate constant*. On peut aussi préférer laisser le mnémonique `add`, puisqu'il n'y a pas d'ambiguïté.

Il est raisonnable de coder la constante en place du numéro de registre qu'elle remplace, ce qui signifie que si on n'a que 16 registres on ne peut coder que les constantes de 0 à 15. C'est déjà pas mal. Il y a plein d'astuces pour récupérer des bits de plus pour ces constantes immédiates.

Un cas particulier est l'instruction `move Rx -> Rd` qui réalise une copie. On aimerait que sa version constante, que nous appellerons `let`, puisse mettre une constante arbitraire de 32 bits dans un registre, mais ceci nécessiterait une instruction de plus de 32 bits... Définissons donc `LetHigh 123 -> R4 LetLow 456 -> R4`

qui placent la constante (16 bits) respectivement dans les moitiés hautes et basses de `Rd`. Cela permettra de charger un registre avec une constante arbitraire en deux instructions de 32 bits. Une option CISC (voir `toto.s`) est d'avoir une instruction de chargement de constante 32 bits qui fait plus de 32 bits.

*Dans `toto.s`, l'addition est une opération à deux opérandes, dont une constante et une en mémoire (sur laquelle nous reviendrons).*

### 8.5.3 Instructions d'accès mémoire

On a aussi besoin d'instructions pour l'accès à la mémoire.

### Lecture et écriture basique

Par exemple, une instruction qui demande à la mémoire le contenu de la case d'adresse le nombre contenu dans R2, et le place dans R5, s'écrira

Read  $[R2] \rightarrow R5$

et l'instruction d'écriture s'écrira

Write  $R3 \rightarrow [R6]$

### Adressage indirect

Dans une machine 3 opérandes, on n'utilise pas le troisième champ opérande. Comment l'utiliser intelligemment ? En inventant des *modes d'adressages* plus sophistiqués. Par exemple, si  $A$  est un tableau stocké en mémoire à partir de l'adresse  $a$ , l'accès à l'élément  $A[i]$  de ce tableau demande d'ajouter  $a$  et  $i$  pour obtenir l'adresse de la case. Il est naturel de proposer les instructions

Read  $[R2+R3] \rightarrow R5$

Write  $R3 \rightarrow [R6+R7]$

Remarquez que dans le write, on n'a aucun registre de destination, les trois registres sont lus. Cela signifie que cette instruction aura un vrai surcoût architectural par rapport aux instructions de calcul qui lisent deux registres et en écrivent un. Si on choisit d'offrir cette instruction, cela implique que notre architecture matérielle sera capable de sortir trois registres différents de la boîte à registres : on peut se demander alors s'il n'y a pas d'autres instructions (de calcul) qui ont besoin de lire trois registres aussi. On pensera au MAC (multiply-accumulate) qui réalise  $Ri+Rj \times Rk \rightarrow Rd$  (et à sa version flottante qui s'appelle FMA pour *fused multiply-add*, le mot *fused* faisant référence au fait qu'il n'y a qu'un seul arrondi à la fin de ces deux opérations). Ce sont des instructions à 4 opérandes.

Cette remarque est un peu prématuree, vous la comprendrez mieux à la révision, après être passé une fois sur le chapitre suivant. Ce qui est caché derrière, c'est qu'on réfléchit à l'architecture matérielle en concevant l'ISA : ici, on se souvient que lecture et écriture dans une mémoire utilisent deux ports de données différents (voir les figures pp 47 et suivantes – au fait, il faut se convaincre que pour lire plusieurs registres à la fois il suffit de séparer les ports d'adresse et de répliquer le multiplexeur de sortie sur la figure 4.8. Cela a un coût en surface, mais les accès en parallèle à plusieurs cases mémoires sont possibles).

Exemple d'accès à un tableau dans ARM : LD R1, [R3, R5 LSL#4] (si les cases du tableau sont de taille 16) et avec post-incrément : LD R1, [R3, R5 LSL#4] ! (le '!' se lit "bang" et met à jour R3 par  $R3 \leftarrow R3 + (R5 \ll 4)$ , ce qui permet de garder un pointeur sur la case courante).

Cette dernière instruction écrit dans deux registres, ce qui est en principe plus cher en matériel si les deux écritures sont en même temps : il faut arbitrer dans quel ordre se feront les écritures. Ici (pour le moment) pas de souci, une des écritures doit attendre le temps qu'une donnée revienne de la mémoire, alors que l'autre peut se faire plus vite. On utilise le port d'écriture des registres à un moment où il eut été inutilisé, ce qui n'a pas de surcoût matériel.

### Instructions de gestion de pile

Les langages de haut niveau utilisent beaucoup des piles (*last in, first out*). Par exemple, c'est comme cela que l'on gère les appels de procédures (`call` et `return` que l'on va voir bientôt).

- Une pile est stockée en mémoire.
- Un registre sert de pointeur de pile (on l'appelle alors SP pour *stack pointer*).
- L'instruction Push R1 va empiler la valeur de R1 sur la pile, c'est-à-dire réaliser la séquence d'actions Write  $R3 \rightarrow [SP]$  puis  $SP+1 \rightarrow SP$  (le tout en une instruction)
- L'instruction Pop R3 va faire le contraire, c'est-à-dire  $SP-1 \rightarrow SP$  puis  $[SP] \rightarrow R3$

La description précédente nous donne une pile montante (en mémoire) avec un pointeur vers la prochaine case vide. En fait il y a 4 variantes de pile :

- Si vous échangez SP-1 et SP+1 ci-dessus, vous obtenez une pile descendante : quand on empile, on descend dans la mémoire.
- Si vous échangez les deux actions de chaque séquence (par exemple Push devient  $SP+1 \rightarrow SP$  puis Write  $R3 \rightarrow [SP]$ ), vous obtenez une pile dont le pointeur pointe sur le dernier élément empilé.

Certains processeurs imposent l'une de ces 4 variantes (exemple : IA32), d'autres vous laissent choisir. Par exemple sur ARM n'importe quel registre peut servir de pointeur de pile, et le push et le pop sont des variantes des instructions mémoire classiques. Par exemple LD R1, [R3, #4] ! peut servir de pop sur une pile descendante qui pointe sur pile pleine, et dont le pointeur de pile est R3.

### 8.5.4 Instructions de contrôle de flot

Pour le moment notre processeur est capable d'exécuter un cycle de von Neumann sur carton perforé, mais pas de sauter des instructions ni de répéter en boucle un programme. Il reste à ajouter des instructions de *branchement*.

#### Sauts inconditionnels, relatifs et absolus

Il faut au moins deux instructions de *saut relatif* :

Jump +123 qui avance de 123 instructions

Jump -123 qui recule de 123 instructions.

Ces deux instructions sont en pratique des additions/soustractions sur le PC. Cela s'appelle un saut ou branchement<sup>3</sup> *relatif* (relatif au PC). La constante ajoutée/soustraite occupe tous les 24 bits restant (une fois qu'on a choisi de coder l'instruction elle-même sur 8 bits).

On ne peut donc pas sauter à plus de  $2^{24}$  cases de la position actuelle du PC. Si vous y tenez, le processeur vous offrira en général une instruction permettant de sauter à un emplacement arbitraire (saut ou branchement *absolu*). A nouveau, soit il faut plus de 32 bits pour coder une telle instruction, soit il faut bricoler. On peut simplement mettre des adresses de saut dans des registres, et le saut absolu est alors un MOV Rx → PC (il nécessite donc plus d'une instruction, pour remplir Rx avec l'adresse voulue avant le saut). Il y a aussi des mécanismes plus spécifiques. Par exemple, on a vu des sauts absolus indirects, motivés par le besoin d'avoir des appels systèmes partagés : l'ISA décrète que les  $2^{10}$  premières cases mémoires contiennent des adresses de saut (que vous remplirez avec les adresses des fonctions système de base comme malloc ou printf), et l'instruction de saut absolu n'a besoin que de 10 bits pour coder une telle adresse. Par contre, un saut absolu nécessite une lecture mémoire supplémentaire. Il y a toujours une LCDE sur les sauts absolus.

#### Sauts conditionnels – drapeaux

Pour revenir à une instruction de saut minimale, il faut tout de même pouvoir implémenter des boucles for et while. Les instructions de saut existent en version *conditionnelle*, par exemple :

GoForward 123 IfPositive

qui avance de 123 instructions si le résultat de l'instruction précédent cette instruction était positif. On aura un certain nombre de telles conditions (si le résultat était nul, non nul, s'il y a eu un débordement de capacité). Pour ces branchements conditionnels, la distance de saut occupe moins de bits, car il faut des bits pour coder la condition.

Le processeur doit donc garder en mémoire l'information correspondante (résultat positif, résultat nul, ...) d'une instruction à l'autre. Cette information est compacte (peu de bits) et est stockée dans un registre de *drapeaux*. Voici les 4 drapeaux (archiclassiques) de l'ISA ARM :

- Z (zero) qui vaut 1 si le résultat d'un calcul est nul
- C (carry) qui contient la retenue sortante d'une addition (dommage de la perdre)
- N (negative) qui contient une copie du bit de signe du résultat
- V (oVerflow) qui indique si une addition/soustraction, **considérée en complément à 2**, a débordé.

Techniquement, pour l'addition, si les deux opérandes ont des signes différents, le résultat ne déborde jamais. Si les deux opérandes sont du même signe, alors l'addition a débordé si le bit de signe du résultat est différent de celui des opérandes. Pour la soustraction, réfléchissez vous-même.

C'est à peu près le seul endroit où l'on s'inquiète du complément à 2, car (rappel) l'addition et la soustraction se fichent pas mal de savoir si leurs entrées et sorties sont signées ou pas, c'est une des beautés du complément à 2.

C'était le minimum syndical : des drapeaux, et une instruction de saut conditionnel. Mais plus la technologie évolue plus un saut conditionnel devient coûteux, essentiellement parce qu'il introduit de l'incertitude qui empêche le pipeline – on verra plus tard. Voyons donc quelques variations.

Une idée rendue possible par le passage aux instructions 32 bits (où on a plein de place, par rapport à 8 ou 16 bits), est de permettre que *toutes* les instructions soient conditionnelles : la condition sous laquelle une instruction va s'exécuter est codée dans un champ supplémentaire du mot d'instruction, et on trouve ce champ dans toutes les instructions. C'est le cas dans l'ARM et IA64.

Dans IA64, de plus, il n'y a pas un seul jeu de drapeaux, mais des *registre de prédicat* : chaque instruction de test peut choisir dans quel registre de prédicat elle stocke le résultat du test, et chaque instruction peut être *prédictée* (conditionnée) par un registre de prédicat.

Dans la variante récente dite Thumb2 du jeu d'instruction ARM, il y a une instruction IT pour If-Then qui rend conditionnelles les instructions suivantes (jusqu'à 4 : l'instruction suivante est conditionnelle, et

3. Branchement par rapport au tronc d'exécution si on faisait PC=PC+1...

on encode dans le IT si les 3 suivantes le sont, et même si elles font partie du *then* ou du *else*). Au final cela permet de rendre toutes les instructions conditionnelles pour des instructions encodées sur 16 bits.

Les étudiants de l'ENS 2016-2017 ont inventé une variante simplifiée, l'instruction `snif` pour `skip next if`, qui permet de se passer complètement des drapeaux : dans Thumb2 on encode encore une condition sur les drapeaux dans l'instruction IT. Dans un `snif` on encode directement la comparaison, par exemple `snif R1<R2`. Cette instruction existe (sous un autre nom) dans les microcontrôleurs PIC.

Dans l'ISA RISC-V, c'est un mélange : le saut conditionnel est une instruction qui compare deux registres (elle inclut donc une opération arithmétique). La comparaison est encodée sur 3 bits : il faut au moins  $=$ ,  $\neq$ , et  $>$  et  $\geq$  en version signées et non-signées. Si la comparaison est vraie on saute à une distance définie par une constante sur 12 bits. RISC-V profite donc de la largesse d'encodage pour se débarrasser complètement des drapeaux.

Remarque : dans tous les processeurs à drapeau, on a un drapeau C qui contient la retenue sortante d'une addition. On a aussi une instruction `add with carry` (en général ADC) qui est une addition qui prend comme retenue entrante la valeur du drapeau C. Cela permet de faire facilement des additions sur des entiers plus large que la taille native du processeur : par exemple sur un processeur 32 bits on fait de l'arithmétique 64 bits en un ADD puis un ADC 32 bits.

Et donc le RISC-V renonce aussi à l'ADC en renonçant aux drapeaux : l'émulation de l'ADC coûtera un saut conditionnel. L'idée est que ce n'est pas très grave vu que la largeur du mot natif est au moins 32 bits. L'ADC était très utile à l'époque des processeurs 8 bits, beaucoup moins en 32 bits, pas du tout en 64 bits.

### Appel de sous-routine

Enfin, il y a une poignée d'instructions qui servent à supporter les langages de haut niveau, en particulier un `call` et un `return` pour les appels de procédure. Sur le fond, `call printf` est un saut absolu à l'adresse de `printf`. Toutefois, avant de sauter, il doit sauvegarder quelquepart l'adresse de l'instruction qui suit le `call`. Ainsi, l'instruction `return` (à la fin du code de `printf`) peut copier cette adresse sauvegardée dans le PC, ce qui reprend l'exécution juste après le `call`.

Le plus souvent, l'adresse de retour (l'adresse de l'instruction qui suit le `call`) est empilée sur une pile, et `return` est essentiellement un `pop PC`. On va voir des variantes un peu plus loin.

On parle d'appel de sous-routine et pas encore d'appel de procédure, parce qu'une procédure sous-entend un certain nombre de paramètres avec leur types, etc. Notre `call` fait juste un saut, sans passer de paramètre. Et `return` correspond bien à "return;" en C, mais on ne sait pas encore comment faire "return value;". La manière dont le passage de paramètre est implémenté sera vue en détail au chapitre 12.

### 8.5.5 Les appels système, les interruptions, et l'atomicité

... seront traités au chapitre 10.

### 8.5.6 Quelques ISA

Voici pour finir quelques exemples représentatifs d'ISA, par ordre croissant de performance... et de consommation électrique.

- **MSP430** est un jeu d'instruction RISC 16 bits à 16 registres qu'on retrouve dans les applications très faible consommation (cartes à puces, tags RFID). Il y a des instructions à un et deux opérandes. Si ces opérandes sont des registres, l'instruction tient en un mot de 16 bits. Les opérandes peuvent aussi être une case mémoire, dont l'adresse (sur 16 bits) est codée dans un des mots suivant l'instruction. Ou des constantes immédiates sur 16 bits, idem. Dans ce cas les instructions font 32 ou 48 bits. Certains registres sont spéciaux : le PC est R0, les drapeaux sont dans le registre R1, et le registre R2 peut prendre différentes valeurs constantes utiles (0, 1, -1...). *Un mot de contexte : c'est un jeu d'instruction pour microcontrôleur très basse consommation. Un microcontrôleur embarque processeur, mémoire, quelques périphériques (timers, etc) et interfaces d'entrée/sortie dans la même puce. L'objectif de faible consommation limite la fréquence, tandis que l'objectif d'une puce tout-en-un limite la capacité mémoire. D'où le choix d'une architecture 16 bits seulement, et le choix d'avoir des instructions qui calculent directement sur la mémoire. Le choix de fréquence basse permet en effet de faire une lecture mémoire et un calcul par cycle. D'ailleurs, le nombre de cycles que prend chaque instruction est exactement égal au nombres d'accès mémoire qu'elle fait.*
- **ARM** est l'ISA qu'on trouve entre autre dans tous les téléphones portables. Cette ISA a commencé comme un jeu d'instruction RISC 32 bits à 16 registres. Dans cette ISA originale, toutes les instructions sont codées dans exactement un mot mémoire (32 bits). Les instructions sont à 4 opérandes :

la quatrième est un décalage qui peut s'appliquer au second opérande. Le second opérande, et le décalage peuvent être des constantes. Par exemple, `add R1, R0, R0, LSL #4` calcule dans R1 la multiplication de R0 par 17, sans utiliser le multiplicateur (lent, et parfois d'ailleurs absent). Il y a d'autres applications plus utiles, comme la manipulation d'octets, et la construction de grandes constantes.

Pour le marché de l'embarqué, ARM a ensuite introduit un mode du processeur qui utilise un second ISA, dit *thumb* (comme Tom Pouce) d'économie de mémoire. Toutes les instructions y ont 2 opérandes seulement et tiennent sur 16 bits. Le processeur expande en interne chaque instruction en l'instruction 32 bits correspondante. Cette expansion n'est pas coûteuse, par contre le même programme, s'il fait moins d'octets au final, demande plus de cycles. Ce mode est utilisé typiquement pour les parties non critiques de l'OS d'un téléphone portable.

Tous les processeurs ARM récents permettent un mode Thumb-2 dans lequel des instructions de 16 et 32 bits peuvent être mélangées. De nos jours, gcc produit du code Thumb-2 par défaut.

l'ISA ARM a été reprise en profondeur dans les versions 64 bits (ARM64 ou AArch64) : il y a désormais 32 registres de 64 bits, etc. Les instructions font à nouveau toutes 32 bits et la plupart des idées rigolotes des débuts ont été abandonnées au profit d'instructions flottantes, vectorielles, crypto, etc.

- **Power** est un autre jeu d'instruction 32 bits qu'on trouve de la Xbox aux serveurs de calculs d'IBM. Les instructions ont également 4 opérandes : l'opération arithmétique de base est le *fused multiply and add* ou FMA, qui fait  $Rd := Rx \times Ry + Rz$ , laquelle permet même de faire des additions et des multiplications.
- **SPARC** est ISA 32 bits à trois opérandes. Son originalité est d'implémenter à partir de SPARC V2 une *fenêtre glissante* de registres sur un ensemble de registres plus grands. Une instruction spéciale fait glisser la fenêtre. Ainsi on a peu de registres visibles à un instant donné (donc un mot d'instruction qui reste petit), mais beaucoup de registres architecturaux. Cela a plein d'avantages, comme lors du passage des paramètres à une fonction. Cela a enfin des avantages pour implémenter un pipeline logiciel. Cette idée est reprise par l'ISA IA64 dont il sera parlé en 11.3.
- **RISC-V** (prononcer risquefailleve) est un ISA open-source conçu par les gens qui ont inventé le RISC dans les années 90 et obtenu le prix Turing en 2017 pour cela. Il fait une synthèse de toutes les bonnes idées qu'on a eu depuis (en particulier dans les 3 ISA précédentes), la plupart du temps en les ignorant dans une démarche de retour aux sources du RISC qui place la simplicité avant tout. Toutefois, c'est un excellent exercice, à la fin de ce cours, de lire son document d'ISA (The RISC-V Instruction Set Manual) : il est relativement court, et plein de notes de bas de page qui justifient les choix faits en les mettant en regard avec les choix faits par d'autres processeurs et l'évolution de la technologie. Par ailleurs, c'est aussi l'ISA à la mode, et comme il est open-source son écosystème se développe à toute vitesse. Bref, si ce cours ne vous a pas donné toutes les billes pour lire cet ISA<sup>4</sup>, allez vous en plaindre au prof.
- **EMT64/AMD64** est une usine à gaz (introduite par AMD, donc nommée AMD64 sous linux, mais Intel l'appelle EMT64). Cet ISA possède plusieurs couches de registres ajoutées par les évolutions de l'ISA<sup>5</sup>, et la taille de ces registres a grandi aussi. Les instructions sont de tailles variables, comme on a vu. Il peut réaliser des opérations arithmétiques dont un opérande est directement en mémoire (voir notre `toto.s`), ce que ne font plus les RISC à cause de l'écart de performance entre accès à un registre et accès à la mémoire.
- J'en présenterai deux autres en 11.3.

## 8.6 Codage des instructions

Même avec de l'imagination, cela nous fait moins de 256 instructions différentes<sup>6</sup> : l'instruction elle-même sera codée sur 8 bits du mot d'instruction. Dans les 24 bits restant, on codera selon le cas trois registres, deux registres et une petite constante, deux registres (accès mémoire), un registre et une petite constante (décalage de bits), un registre et une grande constante.

On s'attachera à ce que le jeu d'instruction soit le plus *orthogonal* possible : les registres seront toujours codés au même endroit, ce qui facilitera la construction du circuit.

On constate que chaque instruction laisse plein de bits inutilisés. C'est du gâchis, et si on voulait

4. Avez-vous remarqué que tout ce bla bla ne vous apprend rien sur cet ISA ? C'est à croire que c'est exprès.

5. Il était une fois, à l'époque 8 bits, un registre de 8 bits nommé A. Au passage à 16 bits il fut renommé AX (pour A eXtended, et A désigne toujours les 8 bits de poids faible de AX). Au passage à 32 bits, il fut renommé EAX, pour Extended AX. Au passage à 64 bits, bien qu'il restât les lettres T, N et D à utiliser dans le mot "extended", il fut renommé RAX. Il contient toujours EAX, AX et A.

6. ...mais c'est qu'on manque d'imagination : il y en a plusieurs centaines dans AMD64.

construire un processeur compétitif, on coderait dedans des options supplémentaires qui rendent l'instruction plus puissante. Pour ceux que cela intéresse, cherchez "ARM assembly language" pour avoir une idée de telles options.

## 8.7 Adéquation ISA-architecture physique

En principe, le jeu d'instruction représente une bonne abstraction de la mécanique sous-jacente du processeur. Ainsi le programmeur a accès à toutes les ressources du processeur, et seulement à elles.

Par exemple, lorsque l'ISA offre une opération de division, cela veut en général dire que le processeur dispose d'un diviseur. Lorsque les opérandes d'une opération peuvent être des registres numérotés de R0 à R15, cela signifie en principe qu'il y a une boîte à registres (une petite mémoire adressable) à 16 entrées dans l'architecture.

Toutefois, cette belle harmonie, qui était de règle à l'époque héroïque des processeurs 8 bits, et qui est de règle pour les ISA récents, souffre trois grosses exceptions.

La première est le fait de jeux d'instruction historiques, tels l'IA32. Par exemple, cet ISA fut à l'origine défini avec 8 registres flottants, ce qui était à l'époque un bon compromis. De nos jours, 8 registres flottants ne sont plus suffisants pour la gestion efficace d'une unité *superscalaire* et *pipelinée*, c'est à dire capable de lancer à chaque cycle plusieurs (2 à 4) instructions flottantes dont le résultat n'arrivera que plusieurs cycles (3 à 5) plus tard. L'architecture physique comporte donc beaucoup plus de registres, et le processeur fait tout un travail de *renommage* des registres ISA en registres physiques. Ainsi, l'architecture physique ne correspond plus à l'ISA. Ceci a un surcoût matériel certain, mais aussi quelques avantages. Par exemple, un jeu d'instruction à 8 registres est plus compact qu'un jeu d'instruction à 128 (cas de l'IA64) : les programmes prendront moins de mémoire. Donc il y a un courant de pensée qui préconise des ISA ne correspondant pas au matériel, tant qu'on sait implémenter leur sémantique efficacement en matériel.

Le second cas de non-correspondance entre l'ISA et l'architecture date d'une époque où l'on programmait plus en assembleur qu'aujourd'hui. Il paraissait judicieux de mettre dans l'ISA des opérations rares et complexes, comme par exemple les fonctions élémentaires sinus, exponentielle, ou des opérations de traitement sur des chaînes de caractères, ou des modes d'adressages indirects réalisant en une instruction plusieurs accès mémoire... Ce sont des instructions IA32, qui au final sont exécutées par un petit programme stocké dans le processeur lui-même. On appelle cela la *microprogrammation*. C'est une caractéristique des ISA CISC, et pour le coup, tout le monde pense que c'était une mauvaise idée. En effet, par rapport à la même fonctionnalité implémentée en logiciel utilisant des instructions simples, cela rend le processeur

- plus coûteux (il faut stocker les microprogrammes),
- plus complexe (il faut un mini cycle de von Neumann à l'intérieur du grand),
- plus risqué (un bug dans le microcode nécessite de changer le processeur, alors qu'un bug dans du logiciel se corrige facilement)
- moins flexible : on ne peut pas adapter ces routines à un contexte donné. Par exemple, on aime avoir plusieurs versions d'une fonction élémentaire : l'une très rapide mais peu précise, une autre très précise mais plus lente, une troisième orientée débit, etc.

Petite histoire : en 1990, quelques années après que les instructions exponentielle et logarithme aient été microcodées pour le processeur 80386/80387 (ancêtre du pentium, en 1985), les PC (comme maintenant) étaient livrés avec des mémoires de plus en plus importantes (loi de Moore). P.T.P Tang proposa des algorithmes qui calculent ces fonctions beaucoup plus rapidement, grâce à l'utilisation de grosses tables de valeurs précalculées (grosse voulant dire : 1Ko). Depuis, tout le monde calcule ces fonctions en logiciel, et les instructions correspondantes sont inutilisées. Pourquoi ne pas implémenter les mêmes algorithmes en microcode, me direz-vous ? Parce que pour le coup, le coût en transistors de la table ne serait pas justifié.

Le troisième cas de non-correspondance entre ISA et architecture est d'ajouter à l'ISA des instructions qui ne sont pas implémentées dans les premières versions du processeur, mais dont tout porte à croire qu'elle le seront à l'avenir. Exemples : la précision quadruple (flottants de 128 bits) dans les SPARC, l'arithmétique flottante décimale dans certains processeurs IBM. Tant que le processeur n'implémente pas ces instructions en matériel, elles déclenchent un saut (techniquement, en utilisant le mécanisme des interruptions) vers une sous-routine logicielle qui traite l'instruction.

## 8.8 Un peu de poésie pour finir

IA32 a des instructions qui font de 1 à 17 octets. L'encodage est sans queue ni tête : il y a des préfixes, des suffixes, etc. Il n'y a que 8 registres entiers, et tous ne sont pas utilisables par toutes les instructions. Les opérations entières travaillent sur un modèle registre-registre ou registre-mémoire, et les opérations flottantes sur un modèle de pile (qui n'a jamais pu être implémenté comme initialement prévu, à savoir une pile virtuelle infinie dont la pile physique fonctionne comme un cache).

La bonne nouvelle est que les compilateurs arrivent très bien à vivre avec un sous ensemble très réduit de ce jeu d'instruction. Par exemple, certaines instructions traitent des chaînes de caractère (copie de chaîne etc), mais l'utilisation des instructions simples donne un code plus rapide.

Cherchez sur Internet le document qui spécifie le jeu d'instruction IA32, et comptez les centaines de pages.

L'extension d'IA32 à 64 bits est pire.



# Chapitre 9

# Construction d'un processeur RISC

Pour construire un processeur, on procède en principe en trois temps :

1. On définit le jeu d'instructions : voir le chapitre précédent.
2. On construit le *chemin de données* (ou *datapath* en néozélandais), dont un exemple est la figure 9.2 p. 89. Il est composé de blocs combinatoires (dont la boîte à opérations ou ALU pour *arithmetic and logic unit*), et de registres (la boîte à registres, les différents registres qui contiennent le PC ou les drapeaux, etc). On les relie par les fils qu'il faut, avec les multiplexeurs qu'il faut. A ce stade on laisse tranquillement débranchés un certain nombre de *signaux de commande*<sup>1</sup> – voir la figure 9.2 – car c'est l'étape 3 qui s'en préoccupera.
3. Enfin on construit un automate qui produit au bon moment tous les signaux de commande qu'on a laissé débranchés à l'étape 2, pour permettre à l'architecture d'implémenter un cycle de von Neumann.

En pratique on itère sur les étapes 1 et 2, selon le principe qu'un bon processeur est un processeur facile à construire, mais on n'attaque pas l'étape 3 avant d'avoir défini toutes les instructions et s'être assuré qu'on sait les implémenter dans le chemin de données. C'est en co-construisant le jeu d'instruction et son chemin de donnée que l'on se rend compte qu'une ISA orthogonale

- minimise le nombre de multiplexeurs nécessaires
- maximise la partie du décodage des instructions qui est dans le chemin de données, et minimise celle qui est dans l'automate, donc le nombre d'états de ce dernier, donc son coût.

Dans les détails on peut déplacer la frontière entre chemin de donnée et automate, cela reste un choix d'implémentation. Dans l'exemple qui suit, ce choix est dirigé par la simplicité de présentation, plus que par la performance du processeur obtenu.

## 9.1 Un jeu d'instruction RISC pas terrible mais à la mode

A titre d'exemple nous allons implémenter un jeu d'instructions 32 bits qui est un RISC-V très simplifié dont les caractéristiques sont les suivantes :

- Adresses d'octets sur 32 bits (on peut adresser  $2^{32}$  octets donc seulement  $2^{30}$  mots de 32 bits).
- 32 registres appelés  $R_0$  à  $R_{31}$ . Le registre  $R_0$  est spécial, il vaut toujours zéro, et écrire dedans c'est pisser dans un violon.
- Code 3 opérandes.
- Les bits 0 à 6 du mot d'instruction codent l'instruction (opcode). On a ainsi 127 instructions possibles.
- Le numéro  $d$  du registre destination  $R_d$  d'une instruction est toujours codé dans les bits 7 à 11.
- Le numéro  $i$  et  $j$  des deux registres opérandes  $R_i$  et  $R_j$  sont toujours codés dans les bits 15 à 19 et 20 à 24.
- pour les instructions ALU, le bit 5 détermine si le second opérande est une constante immédiate ou un registre.
- Lorsque le second opérande d'une instruction de calcul est une constante immédiate, la valeur de cette constante est codée dans les bits 20 à 31.
- Les autres bits servent à coder des trucs et des machins. Ils sont souvent inutilisés. Parfois ils définissent des sous-instructions pour un opcode donné. En particulier les 8 instructions ALU de bases (AND, OR, XOR, SLL, SRL, SRA, ADD, SUB) sont en fait entassées sur le même opcode

1. faux ami : en cockney c'est *control signals*, en français c'est signaux de commande.

(enfin sur 2 opcodes, la variante 2 registres et la variante à constante immédiate) et distinguées par les bits 12 à 14.

- Pas de drapeau, mais un saut relatif conditionnel qui compare deux registres ( $R_i$  et  $R_j$ ), la comparaison étant encodée dans les bits 12 à 14. Tous les autres bits sont récupérés pour faire un offset signé sur 12 bits, qui est multiplié par 4 parce que toutes les instructions font 4 octets.
- On a pour les accès mémoire :  $R_d \leftarrow [R_i + k]$  et  $R_j \rightarrow [R_i + k]$  où  $k$  est un offset signé de 12 bits construit différemment dans le load et le store (dans un cas en recyclant les bits de  $j$  comme pour les opérations ALU avec constante immédiate, dans l'autre en recyclant les bits de  $d$  à la place).<sup>2</sup>
- Pour les accès mémoire les bits 12 à 14 codent la taille de l'accès (un octet, 16 bits, 32 bits) et où ces bits vont dans  $R_d$  (pour 16 bits, dans la moitié du haut ou dans la moitié du bas, etc).

On a simplifié en ignorant plein d'instructions du vrai RISC-V pour que la figure 9.2 reste dessinable.

opération à 2 registres	24	20 19	15 14	12 11	7	6	0	
			<i>j</i>	<i>i</i>	<i>fct</i>	<i>d</i>	<i>opcode</i>	
opération à 1 registre et constante immédiate	31	20 19	15 14	12 11	7	6	0	
			<i>k</i>	<i>i</i>	<i>fct</i>	<i>d</i>	<i>opcode</i>	
lecture mémoire	31	20 19	15 14	12 11	7	6	0	
			<i>k</i>	<i>i</i>	<i>type</i>	<i>d</i>	<i>opcode</i>	
écriture mémoire	31	25 24	20 19	15 14	12 11	7	6	0
	<i>k<sub>H</sub></i>		<i>j</i>	<i>i</i>	<i>type</i>	<i>k<sub>L</sub></i>	<i>opcode</i>	
saut conditionnel	31	25 24	20 19	15 14	12 11	7	6	0
	<i>k<sub>H</sub></i>		<i>j</i>	<i>i</i>	<i>cmp</i>	<i>k<sub>L</sub></i>	<i>opcode</i>	

FIGURE 9.1 – Codage de certaines instructions du RISC-V.

## 9.2 Plan de masse

J'ai déjà dit tout le bien que j'en pensais de la notion d'orthogonalité.

L'architecture de notre processeur putatif est donnée par la figure 9.2. Par rapport à ce que j'ai pu dessiner en cours, il est possible que ce soit plus propre.

Comme déjà dit, à ce stade on laisse tranquillement débranchés un certain nombre de *signaux de commande* :

- les entrées de commande de certains multiplexeurs : par exemple celui qui contrôle le bus d'adresse, l'information peut venir du PC ou de l'ALU, et cela dépend de où on en est dans le cycle de von Neumann...
- les entrées *we* de tous les registres (pour *write enable*, parfois aussi appelée *ce* pour *clock enable*, mais c'est mal – voir la page 46). Sinon tous les registres ont implicitement la même horloge et on ne la représente pas. On nomme ces entrées, mais on les laisse également débranchées, car l'instant auquel on veut enregistrer une information (par exemple la nouvelle valeur du PC) dépend également de où on en est dans le cycle de von Neumann.

On a fait entrer tous les bits du mot d'instruction dans l'ALU et l'automate de commande, en fait elles n'ont besoin que des bits *opcode* et *fct* de la figure 9.1.

Il faut vous convaincre que vous savez déjà tout construire là dedans.

Quand on voit le résultat cela paraît compliqué, mais quand on le construit instruction par instruction c'est tout naturel (enfin il faut une bonne gomme quand même). Je vous invite donc à la redessiner sur une feuille A4 (vite fait en mode brouillon et au crayon à papier parce qu'on va gommer) en suivant l'ordre suivant.

1. Poser le registre du PC, et un additionneur qui ajoute toujours 4 devant. Le brancher sur la sortie *Addr*. Ne pas oublier son entrée *we*. Poser le registre d'instruction *IR*, et brancher son entrée sur l'entrée *DI* du processeur. Ne pas oublier l'entrée *we* du registre. Et voilà, on a un cycle de von Neumann qui ne fait rien.

2. Mais à quoi sert une instruction  $R_d \leftarrow [PC + k]$  ? Eh bien c'est très utile pour accéder à un attribut d'un objet. Le fait de ne garder que cette instruction illustre peut-être l'importance qu'ont pris les langages objets...

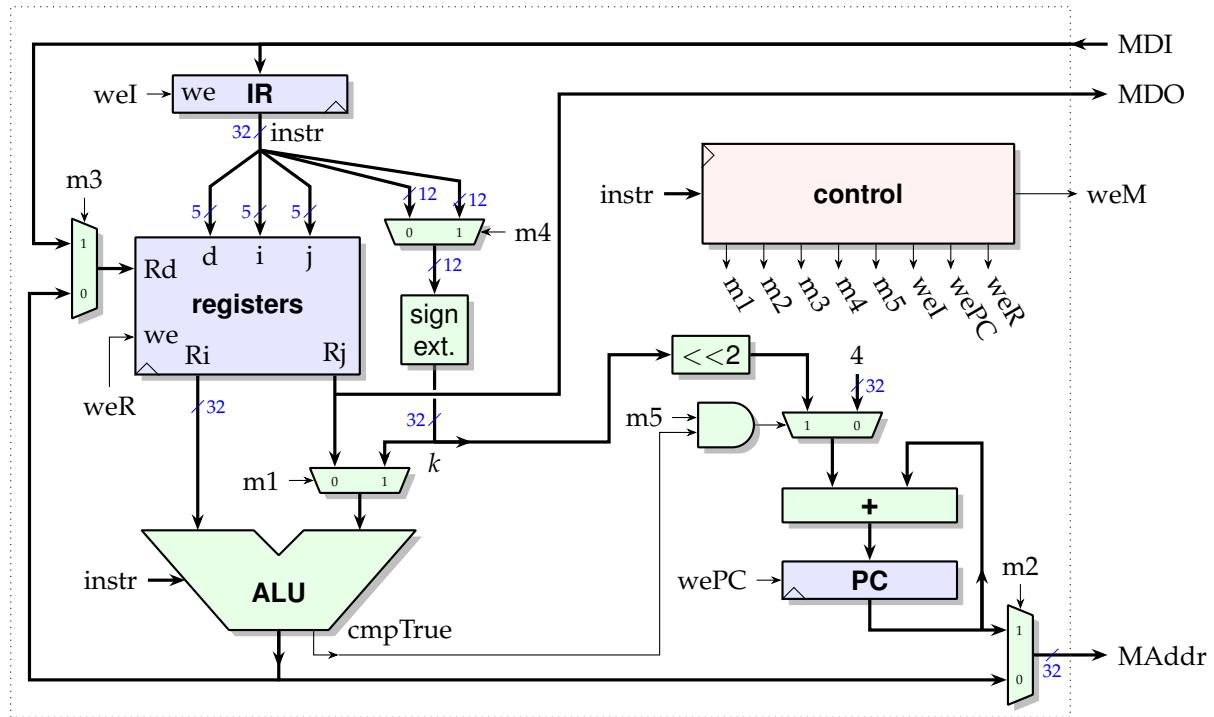


FIGURE 9.2 – S'il te plait, dessine-moi un RISC-V (en bleu les registres, en vert le combinatoire)

2. poser l'ALU et la boite à registres. Bien énumérer toutes leurs entrées et sorties. Par exemple, l'ALU a une sortie cmpTrue parce qu'elle va calculer une comparaison pour les sauts conditionnels. C'est normal et pas grave de l'oublier au début.
3. Pour faire marcher les instructions ALU à deux registres, tirer les fils entre Ri et Rj et les entrées de l'ALU, entre la sortie de l'ALU et Rd (sans mux pour le moment). Tirer aussi les fils entre le mot d'instruction et les entrées i, j, et d.
4. Pour faire marcher les instructions ALU à un registre et une constante immédiate, ajouter un multiplexeur devant une des entrées de l'ALU. Une entrée de ce MUX est Rj, l'autre est une constante qu'on apporte depuis le registre d'instruction.
5. Pour faire marcher les instructions mémoires, ajouter le mux avant la sortie Addr. Ajouter aussi un mux devant l'entrée Rd de la boite à registres. Tirer les fils.
6. Enfin, pour faire marcher le saut relatif, ajouter un mux devant l'entrée de l'additionneur du PC : des fois on ajoutera 4 (on passe à l'instruction suivante), des fois on ajoutera la constante de saut. Tirer les fils.
7. Pour que le saut relatif devienne conditionnel, on peut soit faire rentrer cmpTrue dans l'automate, qui positionnera m5 à 0 lorsque cmpTrue est à 1. Ou bien on peut choisir, comme sur la figure, que c'est le chemin de donnée qui décide : il faut ajouter 4k au PC si on est dans une instruction de saut relatif (m5 est vrai) et si cmpTrue est à 1. Sinon on ajoute 4 au PC. Remarque : en fait même m5 pourrait être considéré comme une fonction combinatoire du mot d'instruction... Mais autant cacher cette fonction dans la grosse fonction de sortie de l'automate de commande.
8. Enfin, lister les signaux de contrôle et poser l'automate qui les produit. Cet automate prend en entrée le mot d'instruction, et possiblement d'autres signaux qui viennent du datapath (mais pas ici).

### 9.3 Construction de l'automate de commande

Une fois ce dessin fait, il ne reste plus qu'à construire l'automate intitulé "control", et qui positionne tous les petits signaux de commande<sup>3</sup> en fonction du cycle de von Neumann de chaque instruction.

Par exemple, la figure 9.3 décrit un automate minimal qui permet d'exécuter les instructions ALU dans le chemin de donnée de la figure 9.2 :

3. Rappel : *control* en Elizabethain c'est commande en français, pas contrôle.

- au premier cycle, on positionne le multiplexeur en bas à droite ( $m2=1$ ) pour que le PC soit recopié sur le bus d'adresse MAddr. La mémoire réagit en plaçant le contenu de la case mémoire d'adresse PC sur le bus de donnée MDI. L'automate positionne aussi le *write enable* du registre IR à 1 (et laisse les autres we à 0 – par convention dans cette section, les signaux non précisés dans une patate sont à 0).
- ainsi au front montant qui termine ce cycle, la donnée présente sur MDI donc en entrée de RI est enregistrée dans RI. Dans tous les autres états du cycle de von Neumann on va laisser weI à 0, ce qui nous garantit que l'instruction reste stable le temps qu'il faut.
- Le cycle suivant s'appelle Decode, et il ne fait rien de particulier<sup>4</sup> sauf envoyer l'automate sur un chemin ou sur l'autre en fonction de l'instruction.
- Au cycle suivant, par exemple pour les instructions ALU, on fait deux choses en parallèle<sup>5</sup> :
  - on réalise le calcul dans l'ALU. La boîte à registres reçoit du mot d'instructions i, j, et d. Le mux m1 est aussi positionné en fonction du mot d'instruction (c'est son bit 5). Elle sort les contenus des registres
  - on augmente PC. Pour cela il faut positionner  $m5=0$  pour que l'additionneur ajoute 4, et wePC à 1 pour que le résultat de cette addition soit enregistré au prochain front montant.

Exercice : complétez en ajoutant un chemin pour les écritures mémoire (l'automate doit alors activer weMem au bon moment), puis un chemin pour les instructions de saut conditionnel (l'automate doit alors positionner m5 à 1).

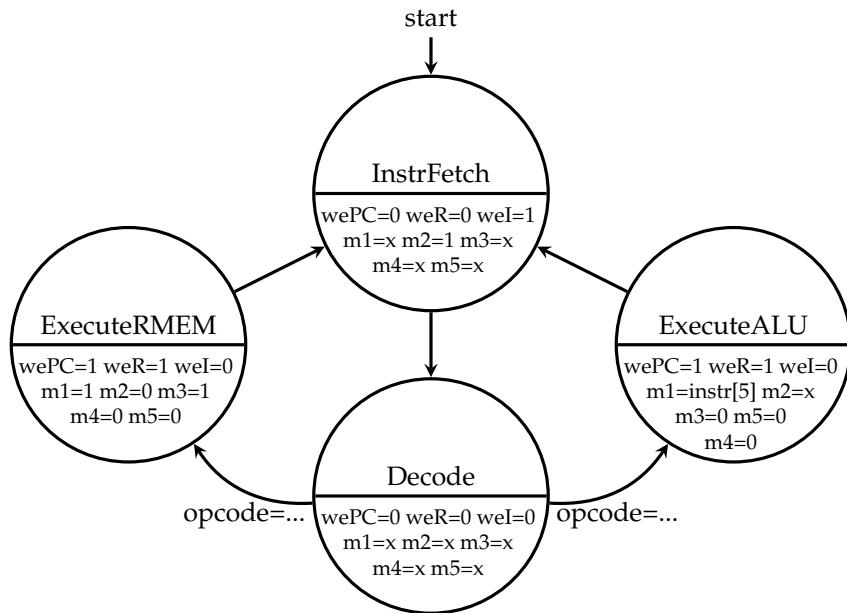


FIGURE 9.3 – Le cycle de von Neumann pour l'exécution des instructions ALU et lecture mémoire. Rappel : “x” veut dire “don't care”

Pour l'encodage des états, on pourra essayer de s'arranger pour que les bits codant les états soient directement des signaux de commande.

Enfin il ne reste qu'à implémenter l'automate comme vu au chapitre 6.

Et voilà un processeur universel qui fonctionne. C'est un objet de ce genre là qu'on vous demande de construire en TD. Au chapitre suivant nous explorons différentes manières de le rendre plus efficace.

4. Lorsqu'on pipelinera ce processeur, on trouvera du boulot à faire dans l'état Decode.

5. Ce serait plus pédagogique de décomposer ces deux actions en deux cycles, mais quand je le fais il y a toujours un petit malin pour me faire remarquer que cela peut se faire au même cycle.

# Chapitre 10

## Entrées/sorties, interruptions, exceptions, et appels systèmes

Rappels : `call/ret` versus `branch-and-link/mov PC`, R14 : cf 8.5.4 p. 82.

Dans cette section, nous allons montrer comment le mécanisme de `call` (ou `branch-and-link`), une fois en place, va se révéler incroyablement utile dans au moins trois situations : la gestion des interruptions, la gestion des exceptions, enfin les appels systèmes.

Mais d'abord, il nous faut discuter un peu comment la machine de von Neumann discute avec le Monde Extérieur.

### 10.1 Interfaces d'entrée/sorties

C'est vrai ça. Et le clavier ? Et l'écran ? Et la carte son ? Et la prise réseau ?

Tous ces gadgets sont appelés *périphériques* (en canadien de l'ouest *devices*).

Pour gérer un périphérique on met en place des *protocoles* partagés par l'ordinateur et le périphérique. Deux exemples (très simplifiés)

- Ce que vous tapez au clavier s'accumule dans une zone de mémoire appelée le *tampon* (ou *buffer*) clavier. L'ordinateur, lorsque cela lui chante, lit ces informations, puis informe le clavier (par une écriture dans une autre zone de mémoire) qu'il les a lues, et donc que la zone correspondante du tampon est à nouveau disponible.
- De même, l'ordinateur écrit, quand cela lui chante, dans une autre zone mémoire, des valeurs de couleurs. La carte vidéo, à la fréquence dictée par le moniteur, accède à ces mêmes cases mémoires et envoie l'information correspondante au moniteur qui les jette sur l'écran avec son canon à électrons, formant une image. Enfin du temps des moniteurs cathodiques c'était cela.

Plus généralement, les entrées/sorties s'appuient sur deux mécanismes :

1. la projection en mémoire des entrées/sorties (*memory-mapped IO*). Le périphérique a le droit d'écrire ou de lire des informations dans certaines zones bien précises de la mémoire, et ce dans le dos du processeur. Le matériel qui implémente cette interface s'appelle *device controller* dans le texte, parfois *coupleur* en québécois.

Remarque : il a jadis existé dans de vieux processeurs des *instructions d'entrées/sorties* qui permettaient au processeur de lire ou d'écrire une valeur sur un bus spécifique, distinct des bus Adresses et Données. La projection en mémoire est plus souple, mais aussi plus économique en broches d'entrées/sorties (les pattes de la puce).

2. un mécanisme d'*interruptions* qui permet à un signal externe d'interrompre temporairement le programme en cours pour que le processeur s'occupe de traiter l'événement. Ce sera détaillé en 10.2. En gros, lorsque survient un événement, l'instruction en cours est terminée<sup>1</sup>, le PC courant<sup>2</sup> est sauvegardé quelquepart, et sa valeur est remplacée par l'adresse d'une *routine de traitement d'interruption* ou (*interrupt handler*) ou *Interrupt Service Routine* (ISR) qui fait en général partie du système d'exploitation (tiens tiens, le voilà). Puis le cycle de von Neumann reprend, le processeur exécutant l'ISR qui gère ce qui a provoqué l'interruption. Ceci fait, le PC est restauré, et l'exécution du programme interrompu reprend là où elle avait été interrompue.

1. Dans le cas d'exécution superscalaire, les instructions en cours sont terminées au sens de terminator : on les re-lancera.  
2. Dans le cas d'exécution superscalaire, c'est toute une poésie de savoir quel est le PC courant.

On voit que le protocole qui régit la communication avec le périphérique est aussi en partie implémenté par du logiciel, qu'on appelle *device driver* c'est-à-dire *pilote de périphérique*. Ce sont les fameux *drivers* qu'il faut installer quand on ajoute un gadget à son PC.

**Parenthèse gamerz : l'invention des GPU** *L'accès à la mémoire video fonctionne à plein régime tout le temps. Pour un pauvre écran de  $1024 \times 800$  pixels codés sur 4 octets, rafraîchi à 60Hz (à moins, il fatigue les yeux), la carte video doit envoyer  $60 \times 4 \times 1024 \times 800 = 196608000$  octets/s au moniteur. Cela justifie que ces accès ne passent pas par le bus général de l'ordinateur, comme jadis<sup>a</sup>. Il y a désormais un bus dédié à la vidéo. Le problème reste que le processeur doit calculer et envoyer à la carte video une quantité similaire de données par seconde pour définir l'image, dans les contextes où elle bouge beaucoup (Lara Croft est toujours pressée). L'idée suivante est donc de faire réaliser les calculs graphiques (rotations, perspective et éclairage) par la carte video. L'avantage n'est pas uniquement que cela fait moins de boulot au processeur : cela fait aussi beaucoup moins de communications entre processeur et carte graphique, puisqu'une image est alors décrite par un ensemble de triangles et de textures, ce qui est beaucoup plus compact.*

a. Mon premier ordinateur, le ZX81, passait la moitié de sa vie à remplir son écran noir et blanc de  $256 \times 192$  pixels. Mais il était pas cher.

Une notion importante ici est que les entrées/sorties sont le plus souvent *asynchrones*, c'est-à-dire selon un minutage indépendant de celui du processeur (puisque le Reste du Monde n'est pas aux ordres de notre processeur).

Et le mécanisme pour ceci est celui des *interruptions*.

## 10.2 Interruptions

Une interruption c'est donc quand le Reste Du Monde interrompt le processeur. Exemples :

- Il est arrivé un paquet sur la carte réseau, et il faut s'en occuper.
- Un fâcheux a appuyé sur une touche du clavier. Pas moyen d'être tranquille.
- La carte son signale qu'elle n'a bientôt plus d'échantillons à jouer.
- Etc.

Une interruption est un événement qui peut être provoqué soit par une cause externe (il y a des broches pour cela sur la puce du processeur), soit par une cause interne, et on parle alors parfois d'*exceptions* (division par zéro, accès mémoire interdit, on en verra plein d'autres en 10.5).

Dans nos trois exemples, remarquez que les deux premiers sont des entrées du processeur (le monde extérieur doit interrompre le processeur, c'est naturel) mais le troisième, la carte son, c'est une sortie du processeur. C'est moins naturel. Cela illustre la chose suivante : une fois qu'on a un bon mécanisme d'interruption, on s'en sert partout.

Lorsque survient un tel événement, le processeur interrompt (temporairement) ce qu'il était en train de faire (son programme) pour sauter à la routine de traitement des interruptions.

C'est donc un mécanisme de *call/return*, avec une première grosse différence : l'adresse à laquelle le processeur saute est fixe (déterminée par la nature de l'interruption).

On va voir bientôt qu'il y a une autre grosse différence, qui est que le processeur change de mode.

En cas d'interruption urgente, du reste, il faut quand même que le processeur termine proprement le cycle de von Neumann de l'instruction en cours avant de sauter à la routine de traitement des interruptions. Ceci nous amène à la question suivante.

## 10.3 Instructions atomiques

Si l'on revient au *call* normal, en mode utilisateur, il fait deux actions en une instructions : 1/ sauvegarder l'adresse de retour, et 2/ changer le PC. Il est important que ces deux actions ne soient pas séparables (on parle d'*atomicité*) pour ne pas permettre qu'une interruption intervienne entre les deux. Plus précisément, ce qui doit être atomique est l'empilement de l'adresse de retour (*push*) : si on le fait en deux instructions (par exemple écrire en mémoire puis incrémenter le pointeur de pile), une interruption entre les deux va provoquer un second *push*, donc deux écritures à la même case mémoire, la première étant perdue.

Voici en détail deux mécanismes qui existent sur de vrais processeurs :

- *call atomique* : fait un empilement puis un saut de manière atomique. C'est la manière traditionnelle de faire. Une interruption déclenche également un *call* vers une adresse pré-déterminée. Vous pouvez vérifier que tout se passe bien.

- les designers de l'ARM ont trouvé que le mécanisme de call était trop complexe. Plus exactement il a trois actions à réaliser (une sur SP, une sur la mémoire, et une sur PC) : c'est trop. L'instruction ARM qui remplace le `call` est `BL` pour *branch and link* qui copie le PC dans R14 puis exécute le saut. C'est le destinataire du saut qui est chargée de sauvegarder R14 sur la pile (et parfois on peut l'économiser).

Mais alors, une interruption entre le BL et la sauvegarde de R14 ferait qu'on perdrait R14. Argh. Pour cette raison, une interruption, dans ces ISA, utilise un R14 et un SP qui sont spécifiques aux interruptions. Du coup pas besoin de les sauvegarder. De plus cela allège le temps de traitement d'une interruption urgente. Pour des raisons similaires, ARM a des instructions de sauvegarde atomique de tous les registres en mémoire.

Autrement dit, lorsque survient une interruption, le processeur change aussi de R14 et de SP. Certains processeurs font de même pour une partie de leurs registres standard Ri pour éviter les sauvegardes et restaurations en cas d'interruption urgente.

On voit que le processeur peut travailler dans deux modes : le mode normal, et le mode "traitement d'interruption", qui se distinguent par les registres physiques associés aux R14 et PC des instructions. Généralisons un peu.

## 10.4 Modes d'exécution

Donc on des instructions atomiques, donc non interruptibles. Mais que se passera-t-il si une seconde interruption survient alors que le processeur n'a pas fini de traiter la première (il est toujours dans l'ISR) ? On aimerait assurer une certaine atomicité (ininterruptibilité) à l'ISR.

La solution usuelle est qu'une interruption fait passer le processeur dans un mode ininterruptible.

C'est l'ISR qui choisira quand revenir au mode normal. Du coup il faut une instruction différente pour revenir d'une interruption vers le mode normal (RETI au lieu de RET).

C'est ainsi qu'un processeur qui se respecte peut fonctionner au moins dans deux modes : le mode *utilisateur* et le mode *superviseur* (système).

Certains processeurs ont toute une hiérarchie de modes intermédiaires, et une interruption fait passer le processeur dans un mode plus privilégié. Tout en bas, le mode utilisateur; tout en haut le mode superviseur; et au milieu un ou plusieurs niveaux d'interruptions avec des priorités entre elles. Par exemple des interruptions pas trop urgentes (niveau 1) et des interruptions à traiter en urgence (niveau 2). Les interruptions urgentes peuvent interrompre la routine de traitement des interruptions de niveau 1, mais le contraire n'étant pas vrai.

Remarquez que la montée en privilège lors du traitement d'une interruption assure que le processeur ne sera pas interrompu par une seconde interruption de même niveau tant qu'il n'aura pas terminé de traiter la première.

Avec tout cela, on comprend qu'une interruption n'est pas traitée instantanément. Il y a typiquement un mécanisme de handshake (request/acknowledge, vous vous souvenez) entre l'extérieur et le processeur. Et éventuellement une file d'attente à l'extérieur.

## 10.5 Exceptions

C'est un mécanisme qui permet de traiter des problèmes provoqués par le programme lui-même. Le Reste du Monde est ici innocent, mais le mécanisme mis en place va être exactement le même que pour traiter une interruption : montée en privilège, saut vers une ISR, puis retour.

Quelques exceptions avec des traitement utiles :

- invalid operation : l'ISR tue le processus fautif. Pour cela on voit bien qu'elle doit s'exécuter dans un mode très privilégié. Mais alors, que fait le RETI puisqu'on a tué le programme où l'on devait revenir ? Que mettre dans notre PC ? Eh bien votre Linux retourne au shell qui a lancé le processus fautif. C'est juste un autre processus en fait. Le mécanisme pour cela sera vu en 14.2.1.
- unavailable but valid operation, par exemple opération flottante alors qu'on n'a pas de matériel flottant : l'ISR l'émule en logiciel. Pas forcément besoin de privilège (et on accepte d'être interrompu).
- invalid operation in user mode. La sentence sera la mort.
- memory protection fault. Idem, mais ce mécanisme peut être détourné de plein de manières utiles.
- address translation fault, avec l'exemple de la pile.

## 10.6 Appel système

Supposons que le processeur dispose d'instructions craignos, comme celles qui servent à initialiser la protection de la mémoire (qu'on verra plus tard). Il faut qu'elles soient accessibles uniquement en mode superviseur. C'est ce qui va permettre d'isoler les processus les uns des autres (un processus ne pourra pas lire la mémoire d'un autre). Ce n'est pas juste pour se protéger des d'éventuels méchants (processus malveillants ou espions), c'est surtout pour se protéger des abrutis et des amateurs (c'est ce qui permet qu'un processus qui plante ne plante pas tout le système). C'est donc très utile<sup>3</sup>.

La seule manière de passer du mode utilisateur au mode superviseur est par un *appel système*. Techniquement, un appel système ressemble à un `call` ou `BL` : il sauvegarde une adresse de retour. Mais l'adresse de destination n'est pas un argument du `call` comme dans un `call` normal. C'est une adresse fixe, un point d'entrée du système, qui plus est inaccessible en écriture depuis le mode utilisateur (soit en ROM, soit en mémoire protégée).

Donc un appel système sauvegarde l'adresse de retour, change de mode, puis saute à une adresse système fixe. Le code système qui s'y trouve se termine par une autre instruction spéciale qui redescend en privilège (repasse en mode utilisateur) puis saute à l'adresse de retour sauvegardée.

Bref, appel système et interruption s'appuient sur les mêmes mécanismes. Réciproquement on peut voir l'appel système comme une interruption à la demande du logiciel. Par exemple, dans les ISA ARM, l'instruction d'appel système s'appelle `SWI` pour *SoftWare Interrupt*.

## 10.7 Instructions de synchronisation

Voici une autre classe d'instructions qui doivent être atomiques, et donc non interruptibles, et donc justifient des instructions spécifiques. Pour pouvoir programmer un processeur multicœur, il faut pouvoir implémenter des verrous qui permettent de synchroniser des cœurs. Les primitives de verrouillage doivent faire, de manière atomique, une lecture mémoire, un test sur la valeur lue, et éventuellement une écriture. Par la suite on a développé des techniques de synchronisation plus siouxes.

---

3. Car en sincère humaniste je veux croire qu'il y a sur terre plus d'ahuris que d'authentiques malfaisants.

# Chapitre 11

# Exécution parallèle

## 11.1 Pipeline d'exécution

## Les bases

L'idée est simple : c'est celle du travail à la chaîne.

Faisons pour commencer l'hypothèse simplificatrice que notre processeur ne sait faire que des additions et des soustractions. On suppose donc qu'on n'a qu'une unité d'exécution, et on oublie momentanément les accès mémoire.

Alors, pendant que le processeur charge l'instruction  $n$  du programme (*instruction fetch* ou *IF*), il peut décoder l'instruction  $n - 1$  (*decode* ou *D*), lire les registres pour l'instruction  $n - 2$  (*operand load* ou *OL*), exécuter l'instruction  $n - 3$  (*execute* ou *EX*), et écrire le résultat de l'instruction  $n - 4$  dans le registre destination (*writeback* ou *WB*).

Exemple : On a du matériel qui fait l'IF en 1ns, une lecture et une écriture des registres en 1ns, et toutes les opérations en 1ns. Dessiner le diagramme d'exécution (en  $x$  le temps, en  $y$  les étages du pipeline) pour le programme suivant :

```
1: R1 <- add R2, R3
2: R4 <- sub R5, 1
3: R6 <- add R7, R8
```

On définit ainsi les étages du pipeline. Techniquement jusque là c'est très simple, il suffit d'avoir un paquet de registres pour chaque étage, qui transmet l'instruction en cours d'un étage à l'autre. Remarquez qu'il faut transmettre le long du pipeline toutes les infos utiles à la vie future de l'instruction, par exemple son registre de destination.

## Les accès mémoire

Maintenant parlons des accès mémoire. On va déjà supposer qu'on a un accès mémoire séparé programme/donnée, pour pouvoir EX une instruction mémoire en même temps qu'on IF une instruction subséquente. En pratique, on a effectivement deux accès mémoire distincts lorsqu'on a des mémoires caches séparées pour les instructions et pour les données (ICache et DCache). On verra en 13.1 que cela a plein d'autres avantages. Admettons pour le moment.

Ensuite on peut considérer qu'une instruction mémoire est une instruction comme les autres, qui s'exécute dans l'étage EX. Cela pose deux petits problèmes : le premier est que c'est vraiment une mécanique différente des autres qui s'occupe des accès mémoire. Le second est : adieu les modes d'adressages compliqués.

Donc on va plutot dire que toutes les instructions passent par un étage supplémentaire, dit MEM, après l'étage EX. Ciel, cela ajoute un cycle au temps d'exécution de chaque instruction. Oui mais on est en

pipeline : si on considère le temps d'exécuter tout le programme de 100000 cycles, cela ne lui ajoute qu'un cycle aussi. Dessinez des chronogrammes de pipelines jusqu'à avoir bien compris la phrase précédente.

## Les aléas de la vie

Considérons le programme suivant :

```

1           R0 <- xxxx    // adresse du vecteur X
2           R1 <- yyyy    // adresse du vecteur Y
3           R2 <- nnnn    // taille des vecteurs
4           R3 <- 0

5 .boucle    R10 <- [R0]
6           R11 <- [R1]
7           R12 <- R10 * R11
8           R3 <- R3 + R12
9           R0 <- R0 + 4
10          R1 <- R1 + 4
11          R2 <- R2 - 1
12          ifNZ B boucle

```

On a plein de *dépendances* dans ce programme. On parle de dépendance lorsqu'une instruction dépend pour s'exécuter du résultat d'une instruction précédente. Une autre manière de voir est que s'il n'y a pas de dépendance entre deux instructions, on peut les exécuter dans n'importe quel ordre. Ou en même temps (en parallèle).

On distingue

**dépendance de donnée** qui sont de trois types :

- *lecture après écriture* ou *RAW*. Par exemple l'instruction 7 doit attendre pour lire la valeur de R11 que l'instruction 6 ait écrit cette valeur dans R11.
- *écriture après lecture* ou *WAR*. Par exemple l'instruction 9 ne peut pas modifier R0 tant que l'instruction 5 n'a pas fini de la lire.
- *écriture après écriture* ou *WAW*. On n'en a pas ici, mais si on avait deux lignes commençant par  $R1 <- \dots$ , on ne pourrait pas les échanger au risque de changer la sémantique du programme. Seule la première (*RAW*) est une vraie dépendance, c'est à dire une dépendance d'un producteur vers un consommateur. Les deux autres sont dites "fausses dépendances", car elles ne proviennent que du fait qu'on a un nombre de registres limités. On pourrait toujours les supprimer en utilisant un autre registre la seconde fois, si on avait un nombre infini de registres. Par exemple, la dépendance *WAR* de 5 à 10 se résoudrait en utilisant R100 à la ligne 10 (et en renommant R100 toutes les occurrences suivantes de R0 dans le programme). Cette notion est importante à comprendre, car on va avoir plein de techniques matérielles qui reviennent à renommer les registres pendant l'exécution du programme pour ne pas être gêné par les fausses dépendances.

Attention, il peut y avoir des dépendances de données entre deux instructions à travers d'une case mémoire. C'est plus difficile à détecter puisque cela ne se voit pas forcément à la compilation du programme.

**dépendance de contrôle** typiquement pour savoir quelle branche on prend en cas de saut conditionnel. Par exemple on ne peut pas décider ligne 12 du saut tant que la ligne 11 n'a pas produit la valeur du drapeau Z.

**dépendance de ressources** (encore une fausse dépendance qui pourrait être supprimée par plus de matériel) : si par exemple une lecture mémoire fait deux cycles et n'est pas pipelinée, l'exécution de l'instruction 6 dépend de ce que l'instruction 5 ait libéré la ressource.

Et pourquoi je vous embête avec cette notion dans le chapitre sur le pipeline ? D'une part parce que c'est une notion centrale en informatique (elle parle de la sémantique d'un programme). D'autre part parce que dans le pipeline, le W arrive longtemps après le R. Dans le cas de deux instructions consécutives avec une dépendance *RAW*, si on ne fait pas attention, le W de la première pourrait se dérouler après le R de la seconde, qui lirait alors n'importe quoi.

On gère les dépendances

1. d'abord en les détectant. Convainquez-vous que c'est relativement facile, bien que coûteux en matériel. En fait, c'est à cela que sert l'étage *ID* (*Instruction Decode*) entre *IF* et *EX* (pasque sinon, dans notre processeur RISC orthogonal, le décodage d'instruction était quasi trivial). Sur sur la figure 11.1, la boîte "Analyse de dépendances" compare les deux registres opérandes de l'instruction à exécuter avec les registres destination de toutes les instructions en cours dans le pipeline. . Le plus

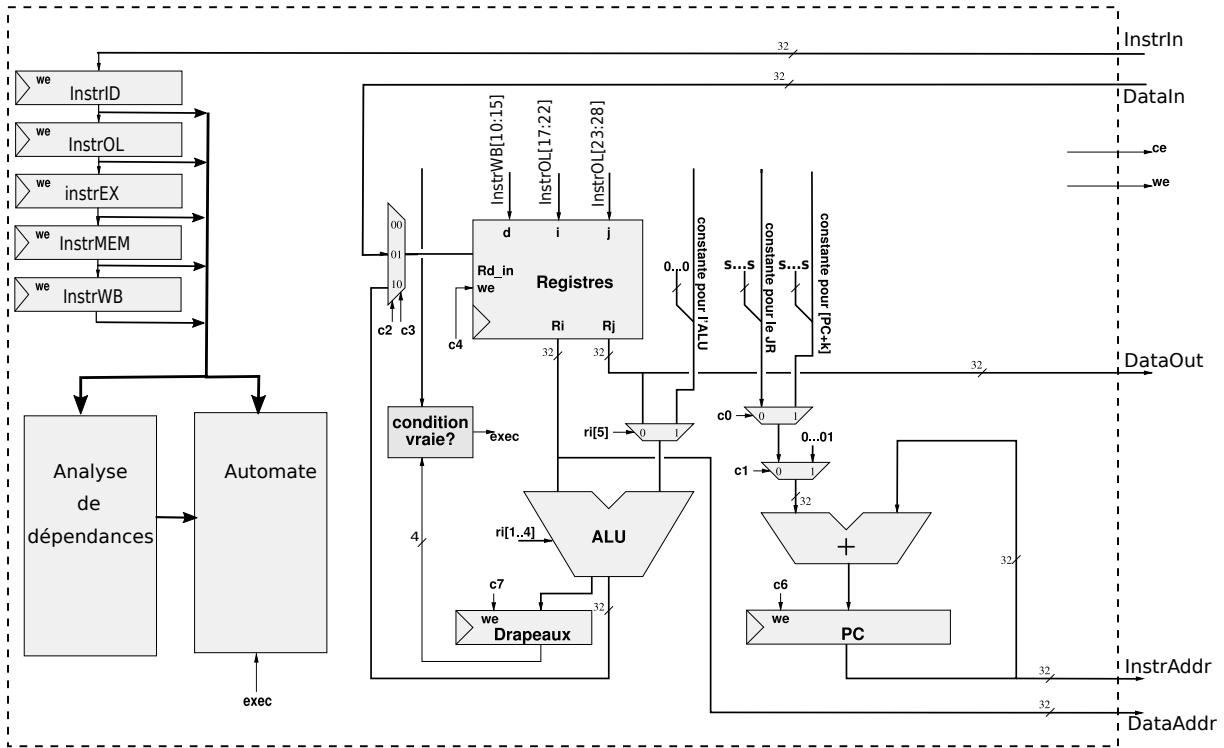


FIGURE 11.1 – Modifications au processeur de la figure 9.2 pour le pipeliner (brouillon incomplet). À gauche, le registre à décalage qui contient les instructions dans les différents étages. Pour le reste, on a juste un automate plus compliqué qui va activer les différents “write enable” plus souvent que dans la version non pipelinée. Exercice : indiquez vous-même sur les fils “constante” de quel étage ils proviennent.

coûteux serait de gérer les dépendances de données à travers la mémoire, mais pour le moment le problème ne se pose pas car lecture et écriture mémoire se font dans le même étage, donc dans l'ordre du programme).

2. ensuite en bloquant le pipeline (plus d'IF) en cas de dépendance jusqu'à ce que l'instruction bloquante ait avancé suffisamment pour que la dépendance soit résolue. On parle aussi d'insérer une bulle dans le pipeline (est-ce une métaphore ou une allégorie ?).
3. enfin en ajoutant des *court-circuits* à notre pipeline. Les deux principaux sont
  - un court-circuit des données : on avait un signal qui remontait le pipeline pour aller stocker dans la boîte à registre le résultat des opérations durant WB. Ce signal est désormais espionné par l'étage EX, qui y repérera les couples (numéro de registre, valeur) qui l'intéressent. Ainsi, une dépendance de données ne bloque plus une instruction dans l'étage OL mais dans l'étage EX. En fait, on va même court-circuiter l'étage MEM pour les instructions qui ne font pas d'accès mémoire, mais attention alors aux conflits possibles entre un couple qui viendrait de l'étage EX et un couple qui viendrait de l'étage MEM au même cycle. Les court-circuits se voient bien sur le diagramme temporel.
  - Un court-circuit des drapeaux vers l'étage IF.
4. Une dernière technique pour gérer les dépendances de contrôle est l'*exécution spéculative*. On prédit le résultat d'un branchement, on charge le pipeline en fonction de la prédiction, et quand on s'est trompé on doit jeter des instructions en cours d'exécution (fastoche, c'est le fil *reset* des registres du pipeline, et aucune instruction n'est définitive tant qu'elle n'a pas écrit dans les registres). En terme de cycles perdus, c'est comme si on avait bullé sur la dépendance de contrôle, mais on ne paye ce prix que lorsqu'on s'est trompé dans la prédiction.

Et comment qu'on prédit ?

- Heuristique simple : on prédit qu'un branchement est toujours pris. C'est (presque) vrai pour les branchements en arrière à cause des boucles, et vrai une fois sur deux pour les branchements en avant. Si on a toutes les instructions conditionnelles (ARM, IA64), les branchements en avant pour cause de `if` sont rares, et la prédiction est bonne à plus de 90% (elle sera mauvaise seulement une fois sur `nnnn` dans notre programme ci-dessus).
- On utilise un prédicteur de branchement : on associe à chaque adresse de branchement, par une fonction de hachage simple (typiquement les 8-10 derniers bits de l'adresse de l'instruction de

brancement), un historique des branchements à cette adresse (enfin, à ce hash). Le processeur utilise cet historique pour spéculer sur la branche à suivre pour le *instruction fetch*.

La première idée est juste de se souvenir si un branchements a été pris ou pas par un bit, et on prédit donc que le prochain passage du processeur à cette adresse prendra la même décision. Mézalors regardons ce qui se passe dans une boucle for qui est elle-même dans une boucle : la première prédiction sera au hasard, ensuite pendant qu'on reste dans la boucle interne, on prédit qu'on saute en arrière et on a raison, sauf à la fin de la boucle ou on se trompe. Ensuite quand la boucle externe fait à nouveau rentrer dans la boucle interne, on est sûr de se tromper une seconde fois, puisqu'on se souvient que le dernier branchements était de sortir de la boucle...

Une solution à peine plus chère est l'automate à 4 états (codés sur 2 bits) de la figure 11.2. Ce prédicteur doit se tromper deux fois pour changer la prédiction, si bien qu'il ne se trompe qu'une fois par boucle.

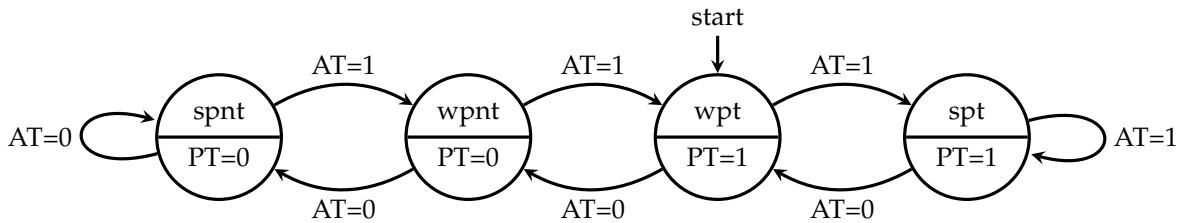


FIGURE 11.2 – L'automate d'un prédicteur de branchements. Un tel automate est associé à chaque adresse d'instruction de branchements. Il change d'état chaque fois que le processeur exécute une instruction de saut à cette adresse. L'entrée AT (actually taken) de l'automate vaut 1 si le saut a été pris. La sortie PT est la prédiction de l'automate : 1 si le branchements est predicted taken. Les noms wpt et spnt se lisent respectivement weak predict taken et strong predict not taken (inférez les autres).

On voit que la gestion physique des dépendances est variable : par exemple, certaines dépendances vont bloquer les instructions à l'étage OL, d'autres à l'étage EX. Quand on va dans les détails, ces techniques sont très complexes. Il faut se convaincre qu'on peut partir du plus simple (bullen) et aller vers le plus complexe (court-circuits et exécution spéculative).

En résumé, la notion de dépendance est une notion abstraite importante. Elle est gérée aussi en amont par le compilateur qui va construire le graphe de dépendance d'un programme pour minimiser leur impact. Reprenons notre programme : un compilateur optimiseur saura intercaler les instructions de gestion de boucle entre les instructions d'accès mémoire et de calcul qui présentent de vraies dépendance.

**Fenêtres de registres** Les systèmes de fenêtres de registres permettent un renommage de chaque registre à chaque itération en une seule instruction (et dans IA64 c'est même dans l'instruction de branchements) : c'est la technique du pipeline logiciel qui permet (dans notre exemple) de charger un registre à l'itération  $n$ , de faire la multiplication et l'addition à l'itération  $n + 1$ , et de ranger le résultat à l'itération  $n + 2$ . Ainsi on a éloigné les instructions dépendantes suffisamment pour qu'elles aient chacune le temps de traverser tranquillement le pipeline.

Il y a un surcoût : l'initialisation et la fin de la boucle sont plus compliquées, car il faut amorcer puis vider le pipeline logiciel. Donc le code est plus gros.

**Aléas** Enfin, il y a des aléas qui se traitent comme des dépendances. Par exemple, une division par zéro doit provoquer une interruption du programme (ou exception). Concrètement, en découvrant que le dividende vaut zéro (dans l'étage EX), le processeur doit sauter à une routine d'erreur. Du coup, toutes les instructions qui suivaient la division, et qui avaient été lancées dans le pipeline, deviennent caduques : on doit vider le pipeline avant de traiter une exception.

Il se passe la même chose en cas de traitement d'interruption en général : si on veut pouvoir la traiter en urgence on ne peut pas attendre que toutes les instructions lancées se terminent, on doit les annuler.

Un autre cas important est le cache miss (on verra ce que c'est en 13.1, vous comprendrez donc cette phrase seulement à la seconde lecture du poly). Un cache miss va bloquer à l'étage MEM (on parle d'aléa mémoire).

Comme le mécanisme de traitement des dépendances et celui de traitement des aléas sont similaires, les dépendances sont parfois aussi appelées aléas (en Cockney, hazard). C'est le cas dans le Patterson et Hennessy, mais je n'aime pas du tout. Un aléa, pour moi, c'est aléatoire.

### La techno et les pipeline

- Construire un pipeline simple c'est juste ajouter des registres. Pour qu'il soit efficace il faut
- qu'on ait bien équilibré la tranche de travail de chaque étage. Ce n'est pas si difficile, grâce à la notion de *retiming*.
  - que le surcoût des registres (en terme de temps de traversée) soit négligeable. En 2002 il y a eu deux papiers dans la conf ISCA convergeant vers l'idée que la profondeur optimale était de 8 à 10 portes de base dans chaque étage. Le pentium 4 était en plein dedans. Malheureusement ces papiers ont négligé les nouveaux petits soucis des techno submicroniques, et le P4 n'a pas pu être poussé jusqu'aux fréquences prévues.

## 11.2 Exploitation du parallélisme d'instruction : exécution superscalaire

Si on creuse, on constate que notre pipeline peut encore être amélioré.

- On a fait comme si l'étage EX était monolithique, mais en fait on a dedans un additionneur/subtracteur (1ns), un multiplicateur (5ns, pipeliné), les mêmes en flottant (8ns), une unité logique (1ns), etc. Tous ces opérateurs pourraient fonctionner en parallèle, avec des profondeurs de pipelines différentes.
- Mais alors, les instructions vont se terminer dans le désordre. Par exemple si l'instruction  $n$  est une multiplication et l'instruction  $n + 1$  est une addition, on aura le résultat de l'addition avant celui de la multiplication. Remarque : le multiplicateur en 5 cycles peut s'insérer dans notre pipeline de la section d'avant, il insérera 4 bulles, et voilà tout.
- Tant qu'à faire, vu que les instructions se terminent dans le désordre, on peut les lancer dans le désordre.
- Tant qu'à vivre dans le désordre, on peut aussi lancer plusieurs instructions à la fois.

Un processeur qui fait tout cela, *tout en conservant une sémantique séquentielle au jeu d'instructions* est appelé superscalaire. Nous allons détailler les mécanismes qui rendent cela possible ci-dessous.

**Sémantique d'un programme assembleur** *La sémantique de l'assembleur est (normalement) définie par l'ISA.*

- *La sémantique d'une instruction est une fonction qui prend l'ordinateur dans un certain état, et le rend dans un autre état. L'état est constitué des valeurs des registres visibles de l'ISA (le PC, ceux de la boîte à registre, les drapeaux) ainsi que des valeurs de la mémoire. Les autres registres (ceux ajoutés pour le pipeline, etc) ne doivent pas intervenir dans la sémantique.*
- *La sémantique d'un programme est construite par composition des sémantiques de ses instructions, avec un opérateur de point fixe pour les boucles. C'est cette sémantique que l'exécution dans le désordre doit préserver.*

Par opposition il existe des processeurs dont le jeu d'instruction a une sémantique exposant explicitement le parallélisme. L'extrême dans cette direction est un processeur *VLIW* (*very large instruction word*) dans lequel chaque instruction est un mot de (par exemple) 128 bits, et est composée de 4 sous-instructions de 32 bits destinées à être lancées sur autant d'unités d'exécution. On aura typiquement deux unité entière, un FMA flottant, et une unité mémoire. Remarquez qu'il n'y a pas lieu d'imposer l'orthogonalité du codage des instructions de chaque sous-classe. Les différents unités d'exécution sont toutes pipelinées, mais avec des profondeurs différentes. Dans la version extrême de cette idée, c'est le compilateur qui se charge de gérer toutes les dépendances de données : si le FMA fait 5 cycles, alors il ne produira pas de code dans lequel deux FMA dépendants sont séparés par moins de 5 instructions. Les bulles dans le pipeline sont ici des *Nop* (*No Operation*) insérés par le compilateur.

L'inconvénient du VLIW est d'avoir du code plus gros, car plein de Nops. Cela se soigne en ajoutant une petite couche de compression/décompression de code. Il faut se convaincre que cela sera moins coûteux que l'architecture superscalaire qui suit...

### 11.2.1 Architecture superscalaire (\*)

Comment construit-on une architecture superscalaire ? Pour détecter les instructions qu'il peut lancer en parallèle, le processeur doit analyser les *dépendances* sur une fenêtre d'instruction en entrées. Il doit ensuite renommer les registres, d'une manière ou d'une autre (voir plus bas) pour faire disparaître les fausses dépendances. Il peut alors sélectionner dans sa fenêtre des instructions indépendantes et les

lancer. Une instruction se termine au bout d'un nombre de cycles variables, qui dépend non seulement de la latence de l'opération à effectuer, mais aussi des dépendances de données et même des aléas mémoire, etc. Lorsque le calcul est terminé, le processeur envoie d'abord le résultat à toutes les instructions qui l'attendent (à travers éventuellement les registres de renommage), puis il doit recopier le résultat (le registre de renommage) dans les "vrais" registres (ceux du jeu d'instruction), en assurant la sémantique séquentielle. Ce processus est encore compliqué par le fait qu'on doit en plus maintenir la cohérence séquentielle en cas d'exception.

On va voir un exemple d'implémentation, mais encore une fois dans ce domaine l'imagination est reine. L'idée centrale est celle du renommage des registres, en gardant la correspondance avec les registres du code dans des tables. On va avoir plusieurs tables partiellement redondantes : c'est pour que leur accès soit rapide. Le mot clé à chercher pour avoir des explications plus précises sur les Internets est : *Tomasulo algorithm*, on trouve de bonnes vidéos.

### Analyse des dépendances et renommage

Les instructions qui arrivent sont lancées dans l'ordre ou le désordre (plusieurs par cycle). La seule condition à ce niveau pour lancer une instruction est qu'il y ait une unité libre pour la recevoir.

En fait l'instruction lancée atterrit dans une petite table de *stations de réservation* située en amont de l'unité d'exécution. Elle va y attendre la valeur de ses opérandes.

Chaque station de réservation (SR) contient une instruction en attente d'exécution, avec ses registres sources et destination, mais aussi des champs destinés à recevoir les opérandes. Ces champs peuvent contenir :

- soit la valeur de l'opérande, si elle était déjà disponible dans la boîte à registres au lancement de l'instruction,
- soit un pointeur vers la station de réservation qui contient l'instruction qui va produire cet opérande, en cas de dépendance de donnée.

On a donc au niveau du lancement une table qui dit "tel registre sera produit par telle SR". C'est cette table, consultée avant le lancement des instructions suivantes, qui assure le renommage des registres opérandes et supprime ainsi les fausses dépendances. Elle peut aussi tout simplement indiquer "tel registre est disponible dans la boîte à registres", c'est d'ailleurs son état au démarrage.

En résumé, les registres destination sont éventuellement renommés en SR au lancement. Ainsi, lorsqu'on a deux instructions proches qui écrivent dans R1, les deux R1 ne sont pas renommés dans la même SR. Bien sûr, au lancement d'une instruction dont la destination est R1, la table précédente est mise à jour pour dire "désormais R1 sera produit par telle SR".

### Exécution dans le désordre

Il y a un bus commun des résultats (il est très large, typiquement plusieurs centaines de bits, car il peut y passer plusieurs résultats par cycles). L'info qui passe est de type (registre destination, SR qui était responsable de cette opération, valeur du résultat). Chaque station surveille ce bus pour y attraper les opérandes qui manquent à ses SR (en comparant les SR de ses opérandes en attente aux SR qui passe sur le bus).

Enfin, à chaque cycle, chaque unité de calcul choisit, parmi ses SR, une entrée pour laquelle tous les opérandes sont disponibles, et la fournit au pipeline de calcul. La SR correspondante est alors libérée.

Au fait, le nombre de SR par unité de calcul est un paramètre dont la détermination est compliquée. Ce sera typiquement entre 1 et 4.

### Terminaison dans l'ordre

C'est pas tout, mais il faut écrire dans la mémoire et dans les registres ISA dans l'ordre. Pour cela, il y a une unité de "terminaison dans l'ordre" en aval des unités d'exécutions, et qui surveille également le bus commun. Cette unité contient un tampon circulaire des instructions lancées et pas terminées (*reordering buffer, ROB*). Chaque entrée est créée au lancement, et contient une instruction avec sa SR. Lorsqu'une instruction a fini de s'exécuter, son résultat passe sur le bus résultat suscité. Le ROB l'attrape et la marque comme terminée. À chaque cycle, le ROB prend un paquet d'instructions terminées consécutives, et les termine effectivement, en écrivant leur résultats dans les registres ou en mémoire.

Si vous avez suivi, une SR peut servir plusieurs fois à des intervalles rapprochés. Pour assurer qu'on pointe bien vers la bonne instruction, les pointeurs vers des SR évoqués plus haut sont en fait des couples (SR, entrée de ROB).

La ROB gère les mauvaises prédictions de branchement (et autres exceptions) : en cas de mauvaise prédition, on vide la ROB de toutes les instructions qui suivent le branchement mal prédit. Ces instructions vont finir de s'exécuter, mais leurs résultats n'arriveront jamais ni dans un registre ni dans la mémoire. Il faut également purger de la table "tel registre sera produit par telle SR" les informations mises par les instructions effacées de la ROB (cherchez comment, je ne suis pas sûr de moi). Puis on repart sur le code de la branche correcte.

Reste à assurer la cohérence séquentielle de la mémoire. C'est pour cela qu'on a besoin d'unités mémoire parmi nos unités de calcul. En lecture, elles doivent s'assurer avant de retourner la valeur lue qu'aucune des instructions précédemment lancée ne va modifier l'adresse à lire. Toute l'info est là, entre la ROB et les SR des unités d'écriture mémoire. En écriture, il y a un tampon qui calcule les adresses mais n'écrit qu'au feu vert de la ROB, donc dans l'ordre.

Au final, un vrai processeur va pouvoir lancer jusqu'à 4 à 6 instructions par cycles, en retirer autant, et en avoir bien plus en vol à un instant donné. Mais en moyenne, on aura un nombre d'instruction par cycles (IPC - l'inverse du CPI qui mesure la qualité du pipeline) qui dépassera rarement 2.

### 11.2.2 VLIW ou superscalaire (\*)

Avant tout, faisons un peu de statistiques sur du vrai code. On constatera avec amertume que le parallélisme d'instruction moyen que l'on arrive à extraire (en tenant compte uniquement des vraies dépendances, pas des fausses) est de l'ordre de 3-4. Autrement dit, avec un processeur superscalaire idéal, on arriverait en moyenne à lancer 3-4 instructions en parallèle par cycle. C'est une question compliquée, d'une part parce qu'il existe des code intrinsèquement parallèles (mon éternel produit de matrices l'est presque), d'autre part parce que cela dépend du jeu d'instruction : avec des instructions plus complexes, donc plus puissantes, on réduit le parallélisme d'instruction... Disons qu'il y a un consensus actuel disant qu'il ne sera pas rentable de construire des processeurs ayant plus d'une dizaine d'unités d'exécutions : on ne saura pas les remplir avec du code séquentiel (et pour le code parallèle, voir plus bas, 11.4.4).

La différence fondamentale entre VLIW et superscalaire est que la gestion du parallélisme d'instruction est faite dans un cas par le compilateur, dans l'autre par du matériel. Voyns les avantages et les inconvénients.

- Le gros avantage du VLIW est la simplicité de son matériel : on a vu que la détection des dépendances et l'exécution dans le désordre sont coûteuses. De plus, elles consomment de l'énergie pour une tâche purement administrative, pas pour le calcul. Les VLIW consomment beaucoup moins.
  - Le VLIW a typiquement du code plus gros (plein de Nop), ce qui finit par avoir un impact sur la conso et la perf (besoin de caches plus grands par exemple). Le superscalaire a du code plus compact.
  - Le code VLIW n'est pas portable d'une archi à la suivante : il faut tout recompiler si la profondeur du pipeline FMA passe de 3 à 5 par exemple. À l'opposé, votre Pentium 4 extrait les dépendances de vieux code DOS aussi bien que du code plus récent.
  - Le compilateur voit une fenêtre de code beaucoup plus grande, et est capable d'en extraire presque parfaitement tout le parallélisme d'instruction pour le donner au VLIW. Le matériel ne pourra considérer qu'une fenêtre plus petite du code en cours d'exécution.
  - Par contre, le compilateur fait un travail statique, alors que le matériel a l'avantage de ne considérer que les dépendances sur la branche de code effectivement prise. Sur du code plein de ifs, la détection des dépendances peut être beaucoup plus fine si faite en matériel.
- Considérant tout ceci, les processeurs VLIW ont du succès dans le domaine embarqué, où
- la basse consommation est importante ;
  - les traitements qui ont besoin de beaucoup de puissance de calcul sont très parallèles (multimedia, compressions de données) ;
  - le code est compilé une fois pour toutes (donc la non-portabilité n'est pas vraiment un problème) ;

## 11.3 Deux jeux d'instructions VLIW récents

### 11.3.1 IA64

Le jeu d'instruction IA64, conçu pour le calcul haute performance, est un peu bâtarde. Il considère les avantages et inconvénients ci-dessus, et essaye de prendre le meilleur des deux mondes en terme de performance. En principe, tout le travail que le compilateur peut faire pour un VLIW, il peut aussi le faire pour un superscalaire. Mais il faut pouvoir l'exprimer ensuite pour le jeu d'instruction.

Donc le jeu d'instruction IA64 est un VLIW bizarre (ils appellent cela EPIC pour *explicitely parallel instruction computing*) : paquets de 128 bits pour 3 instructions, mais le parallélisme est exprimé surtout

par des bits qui délimitent les ensembles d'instructions qu'on peut lancer en parallèle (et ces ensembles peuvent faire de une à beaucoup beaucoup de sous-instructions). Les sous-instructions ne sont pas directement en face du matériel correspondant, il y a une certaine liberté. Et il y a tout de même du matériel de détection des dépendances, pour que le code soit portable.

Au final, IA64 récupère aussi le pire des deux mondes (LCDE) : des compilateurs compliqués et inefficaces, et beaucoup de matériel consacré à l'administration.

Les processeur Itanium (IA64 est le nom de leur ISA) ont été un flop commercial, essentiellement à cause du retard de plusieurs années qu'a pris leur production. Intel et HP ont arrêté de les vendre en 2021.

### 11.3.2 Kalray K1

Le processeur Kalray est un VLIW pur, car sa conception a été orientée vers la basse consommation. Une instruction VLIW (lancée chaque cycle) se compose de

- deux instructions entières 32 bits (ou une instruction 64 bits)
- une instruction de multiplication et/ou de calcul flottant, ou une troisième instruction entière.
- une instruction mémoire
- une instruction de branchement/contrôle.

Ce processeur n'est pas aussi orthogonal que mon ARM ou mon RISC-V. En effet, la complexité matérielle du manque d'orthogonalité, tant qu'elle se limite à une petite couche de logique combinatoire pour le décodage des instructions, a un coût minime, donc on se l'autorise. La complexité en terme de conception est prise en charge par des outils avancés générant le processeur+compilateur.

Les registres sont complètement unifiés. Un registre 64 bits est obtenu en appariant deux registre 32 bits consécutifs.

L'exécution est complètement dans l'ordre, bien que l'exécution des instructions aient des latences variables : 1 cycle pour les instructions entières, 2 cycles pour les multiplications entières, 3 cycles pour les instructions flottantes.

Une originalité est l'instruction complexe (mais pas CISC) qui calcule  $a \times b + c \times d$  en flottant.

Mais le point fort du processeur Kalray est d'intégrer 256 coeurs K1 sur une puce. Ce qui nous amène à la section suivante.

## 11.4 Exploitation du parallélisme de processus

Il est facile de dupliquer un cœur de processeur pour en avoir deux. La difficulté est de les faire collaborer à la résolution d'un même problème, ce qui suppose a/ échanges d'informations et b/ synchronisation (b/ étant du reste nécessaire à a/).

### 11.4.1 Échange de messages versus partage de mémoire

Quand deux étudiants veulent communiquer, ils s'envoient des messages par un canal quelconque. Pour faire communiquer deux circuits/programmes/coeurs/processus/calculateurs<sup>1</sup>, on peut de même mettre en place des mécanismes d'échange de message (*message passing*). On a aussi une possibilité supplémentaire : le partage de mémoire (*shared memory*) (inaccessible aux étudiants, et c'est sans doute tant mieux). Conceptuellement, on a deux processeurs (ou plus) branchés à la même mémoire à travers un bus partagé.

En terme de modèle de programmation, le partage de mémoire est plus confortable. En terme d'architecture, l'échange de messages est plus simple à construire. Il est facile de se convaincre qu'on peut émuler l'un par l'autre et réciproquement. L'architecture sera typiquement du partage de mémoire au niveau d'une puce (processeur multicœur), et de l'échange de message entre deux PC/blades/cartes mères.

Mais comme l'un peut émuler l'autre, commençons par abstraire le modèle de programmation des mécanismes architecturaux sous-jacents.

### 11.4.2 Modèle de programmation : thread versus process

Ici on parle donc de modèle de programmation (ou bien de l'abstraction offerte par l'OS). On peut donner une tâche soit à un processus, soit à un *thread* (que Molière appelait "fil d'exécution" et Victor Hugo "processus léger"). La différence fondamentale est la suivante :

1. A ce stade on pourrait s'inquiéter d'une certaine confusion mais après une page ou deux tout sera clair.

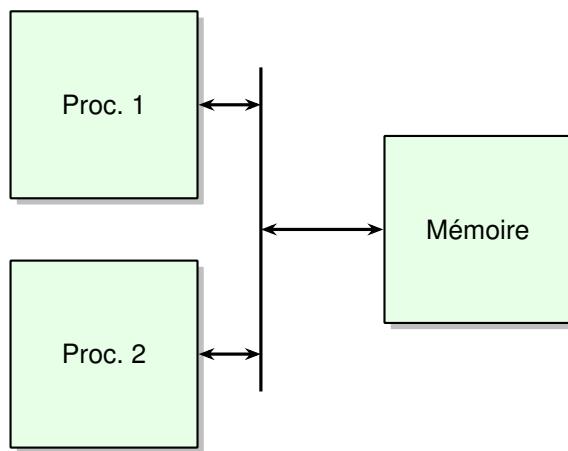


FIGURE 11.3 – Vue abstraite d'une architecture à mémoire partagée

- Deux processus vivent dans des espaces mémoires isolés. Un processus ne peut pas accéder à la mémoire d'un autre processus. Ce n'est pas juste qu'il n'a pas le droit, c'est qu'il ne peut même pas exprimer les adresses qui lui donneraient accès à la mémoire d'un autre processus. Cette isolation est garantie par des mécanismes qu'on verra au chapitre 13.  
Par conséquent, quand on veut que deux processus communiquent, cela doit passer par des messages, par exemple abstraits dans les tuyaux (pipes) d'Unix.
- Deux *threads* peuvent partager la mémoire... à condition qu'ils appartiennent au même processus.

On donnera donc à des processus distincts des tâches bien distinctes (par exemple un processus pour le traitement de texte, un autre pour gérer l'imprimante, un troisième qui transcode un DVD). Et quand on voudra paralléliser une tâche de calcul lourde dans un processus (le transcodage), on pourra essayer de la distribuer sur plusieurs threads au sein d'un même processus.

Ce qui est "léger" dans les threads, c'est que leur création ne demande pas de construire toute l'isolation mémoire.

### 11.4.3 Grandeur et misère de la mémoire partagée

Le fonctionnement en parallèle de deux threads partageant de la mémoire amène un **indéterminisme** intrinsèque qu'on n'avait pas jusque là. En effet, que ce passe-t-il lorsque les deux veulent écrire deux valeurs différentes à la même adresse ? On saura construire du matériel qui tranche dans ce cas, mais il est difficile de donner une sémantique propre à ce genre de situations. Il faut donc les éviter par des mécanismes de verrous, mais... c'est à la charge du programmeur : on renonce à donner une sémantique à tous les programmes. Au passage, tous les mécanismes de synchronisation à base de verrous reposent sur l'écriture par plusieurs threads dans une case mémoire partagée : la question "que se passe-t-il dans ce cas là" ne peut pas être glissée sous le tapis : on en a vraiment besoin.

Et donc, en fait, comme modèle de programmation, les threads sont une horreur. Le moindre programme de 4 lignes peut cacher des bugs horriblement subtils. Cela ne nous regarde pas vraiment dans ce cours, mais quand même, si vous pensez que les threads sont une bonne idée, googlez les deux petits articles suivants :

- *The problem with threads* par Edward Lee.
- *The trouble with multicore* par David Patterson.

Mais vous allez voir que c'est pas cher à implémenter, alors on va quand même en mettre partout.

### 11.4.4 Architecture : Multifilature (*multithreading*)

L'idée est que de construire un processeur capable de gérer plusieurs contextes, qui exécuteront chacun un processus légers (threads). On duplique essentiellement la boîte à registres et le PC, et on augmente un peu la taille de certaines autres structures de données comme les stations de réservation. Mais on ne duplique pas les unités d'exécution (les ALU, les FPU, etc) qui ne bossaient de toute manière pas à tous les cycles.

Quand un thread doit attendre (accès mémoire lent, ou même dépendance de donnée ou branchement conditionnel), le processeur occupe les unités de calcul avec les instructions des autres threads qui ne

sont pas bloqués. Bref, un thread remplit les bulles de l'autre et réciproquement. C'est une manière relativement simple et bon marché d'optimiser l'utilisation des ressources de calcul du processeur.

Il a été proposé<sup>2</sup> des calculateurs sans cache, mais avec 128 threads pour recouvrir la latence de la mémoire (*barrel processor*, ce qui se traduit sans doute par processeur à bâillet). Il faut les trouver, les 128 threads, dans votre Tetris. Oracle vend actuellement des processeurs SPARC très multithreadés (8 threads matériels par cœur) pour les serveurs web et autres, qui peuvent traiter des centaines de transactions à un instant donné.

On peut faire basculer un cœur SMT d'un thread à l'autre en un cycle, mais on ne peut pas le faire basculer d'un *processus* à l'autre en un cycle : l'isolation mémoire est coûteuse à mettre en place (on verra pourquoi au chapitre 13) et on a besoin de la rentabiliser sur plein de cycles.

Quand vous faites `cat /proc/cpuinfo` sur un processeur Intel 4 coeurs, vous en voyez 8. Ce que cette commande compte, c'est le nombre de PC (program counter) qui peuvent vivre leur vie en parallèle à un instant donné<sup>3</sup>.

#### 11.4.5 Architecture : processeurs multicoeurs

C'est facile, on duplique le processeur (voir les photos de puces multicoeurs). C'est pour cela que tout le monde en fait actuellement.

En vrai il y a un problème de taille : la mémoire était déjà trop lente pour un cœur, cela ne va pas s'arranger si 8 coeurs se la partagent. La solution est d'avoir des caches locaux à un cœur : les caches de niveau 1 (L1) sont privés à chaque cœur, le cache de niveau 2 est typiquement partagé à plusieurs coeurs sur une puce, et les caches de niveau 3 peuvent être partagés entre plusieurs puces. Typiquement.

Mézalors il faut assurer la *cohérence* de ces caches : si la mémoire est partagée, alors une adresse peut se retrouver dupliquée dans plusieurs caches, et il ne faudrait pas qu'il y ait des valeurs différentes à cette même adresse. On verra cela en 13.1.8.

### 11.5 Mécanismes architecturaux pour l'exclusion mutuelle

Voir le sujet d'exam "les instructions de synchronisation à travers les âges".

### 11.6 Conclusion : parallélisme partout

En résumé, le matériel de votre PC est plein de parallélisme :

- *bit-level parallelism* : par exemple pour faire un additionneur  $n$  bits en temps  $\log n$ .
- *word-level parallelism*, aka unités vectorielles ou SIMD (Single Instruction, Multiple Data).
- *instruction-level parallelism*, aka ILP, aka superscalaire ou VLIW.
- *thread-level parallelism*, aka multithreading,
- *core-level parallelism*, aka multicore, multiprocessor, etc. Il y a plusieurs étages dans ce dernier niveau : puce multicoeur, carte multipuce, boîtier (ou armoire) multicarte, centre de calcul multi-armoires, et Ze Grid multi tout. Mais au final on compte les coeurs.

Du point de vue du programmeur, les deux derniers niveaux pourraient tous deux s'appeler *process-level parallelism* ou *program-level parallelism* : dans les deux cas on a plusieurs cycles de von Neumann indépendants<sup>4</sup>.

Donc la distinction entre les deux derniers niveaux est surtout justifiée du point de vue de l'architecture : deux threads dans un processeur multithreadé partagent leurs ressources d'exécution (ALU, FPU etc) et donc ont un gros automate en commun pour gérer les deux cycles de von Neumann. Pour le pauvre architecte qui construit cet automate, c'est une manière profondément différente d'exploiter le parallélisme de processus.

Attention, du point de vue du programmeur/système, là où il y a une différence fondamentale, c'est entre thread (deux threads partagent la mémoire) et processus (deux processus vivent dans des espaces mémoire isolés). Mais cette remarque est prématurée car l'isolation mémoire sera vue dans un chapitre ultérieur.

En tout cas le mot thread a un sens très précis en système et un sens très précis en archi, et ce ne sont pas tout-à-fait les mêmes. Vous voici prévenus.

2. Il était une fois une boîte qui s'appelait Tera et développait un calculateur appelé Tera MTA (*multi-threaded architecture*). Elle a levé tellement d'argent avec cette idée qu'elle a racheté le vénérable constructeur de supercalculateurs Cray, juste pour pouvoir vendre le Tera MTA sous le nom Cray MTA.

3. Intel appelle cela HyperThreading, ce qui veut dire "multithreading à 2 threads seulement"...

4. D'ailleurs `cat /proc/cpuinfo` prétend que j'ai 8 processeurs alors que je sais que j'ai 4 coeurs multithreadés.

### 11.6.1 Taxonomie 1 : la classification de Flynn

Flynn donne une classification du parallélisme explicite, ou si vous préférez du parallélisme dans le modèle de programmation :

- SISD : single instruction, single data. Une instruction travaille sur une donnée. C'est ce qu'on a vu au chapitre précédent.
- SIMD : single instruction, multiple data. Une instruction travaille sur plusieurs données. Historiquement on a eu des calculateurs exploitant un SIMD massif (256 données en parallèle). Googlez Connection Machine, MasPar si cela vous amuse. De nos jours on a dans chaque processeur des instructions SIMD à plus petite échelle : une instruction applique la même opération à un vecteur de 4 à 32 nombres. Dans les deux cas on a un seul automate qui contrôle tous les opérateurs, c'est économique. Par contre ce n'est pas très flexible.
- MIMD : multiple instructions, multiple data : on a plusieurs programmes qui vivent leur vie.
- MISD n'a pas de sens.

Attention de ne pas avoir des œillères de Flynn.

- On peut exploiter le parallélisme même en SISD par la technique du pipeline et de l'exécution superscalaire. C'est même une très bonne chose puisque pour le programmeur c'est caché/sans prise de tête/automatique/implicite.
- Un processeur moderne multicoeurs c'est du MIMD où chaque I peut faire du SIMD... cette classification date d'avant les processeurs modernes.

### 11.6.2 Taxonomie 2 : la classification de Raina

C'est une classification des architectures MIMD suivant la manière dont elles accèdent à la mémoire.

Les deux premières lettres décrivent le *modèle de programmation* : mémoire partagée, ou bien mémoire distribuée. La mémoire partagée c'est plus facile à programmer, mais c'est impossible à construire d'une manière qui passe à l'échelle. La mémoire distribuée implique qu'il faut alourdir le modèle de programmation avec des communications explicites entre les processeurs.

Les deux lettres suivantes (et plus si nécessaire) décrivent la réalité du matériel. Cela donne :

- DADM : disjoint address spaces, distributed memory : en terme de mémoire, les processeurs sont isolés les uns des autres. NORMA veut dire "no remote memory access", comme ça c'est clair. Le plus simple à construire mais il faut expliciter toutes les communications.
- SASM : Shared address space, shared memory : tous les processeurs se battent pour la même mémoire. Cela ne passe pas à l'échelle.
- SADM : Shared address space, distributed memory. On a le beurre (la mémoire partagée entre les processeurs) et l'argent du beurre (la mémoire physique distribuée, ce qui permet des accès concurrents donc le passage à l'échelle), mais il va falloir le payer par du matériel. Les options possibles pour ce matériel :
  - SADM-NUMA (non uniforme memory architecture). On place les données physiquement dans les mémoires distribuées, et après elles ne bougent plus. Si on veut accéder à une donnée sur une mémoire lointaine cela va marcher mais cela va prendre le temps de la communication.
  - SADM-OSMA : c'est l'OS (donc le logiciel) qui ajoute l'abstraction de la mémoire partagée par dessus une architecture DADM.
  - SADM-CCMA (cache-coherent memory architecture) : il y a du matériel qui assure la cohérence des caches des différents processeurs avec la mémoire principale. Un exemple de protocole de cohérence de cache est le protocole MESI survolé en section 13.1.8.
  - SADM-COMA (cache-only memory architectures) : la mémoire est uniquement composée des caches de chaque processeur, pas besoin de s'embêter avec une mémoire centrale derrière.

Dans les protocoles de cohérence de cache distribués, il y a toujours un moment où il faut qu'un processeur acquière l'exclusivité sur une adresse (car il ne peut en écrire qu'un, comme dit Yoda dans Highlander). Cela veut dire qu'il doit envoyer un message à tout le monde et surtout attendre la réponse de tout le monde. Ceci prend toujours un temps qui ne passe pas bien à l'échelle.

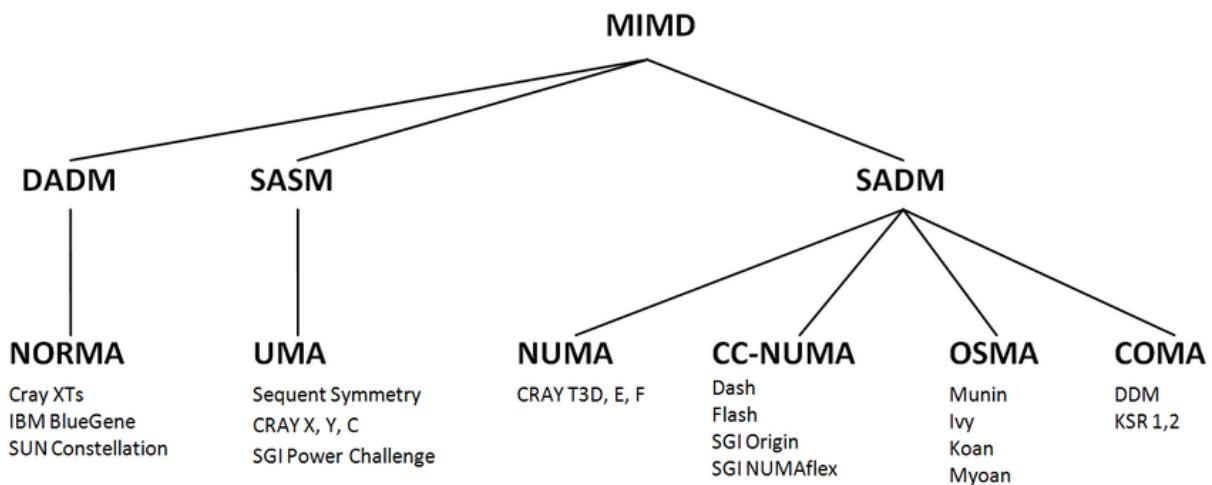


FIGURE 11.4 – Classification MIMD de Raina, avec des exemples de vraies machines

# Chapitre 12

## Du langage au processeur (\*)

*Ce chapitre a eu sa raison d'être à une époque... il ne sera sans doute pas traité en cours parce que des gens plus compétents vous feront des cours de compilation plus sérieux, mais on le laisse quand même.*

### 12.1 Introduction : langages interprétés, langages compilés, processeurs virtuels

... à l'oral.

C, C++, Fortran, qui sont compilés directement en instructions machines.

Python ou Javascript sont interprétés.

Java ou C# sont compilés, mais pour un processeur virtuel qui va tourner dans une machine virtuelle munie d'un processeur virtuel (purement simulé) et d'un espace d'adressage virtuel. Avantages : portabilité, *sandboxing*. Inconvénient : performance. Solution à l'inconvénient : compilation *Just In Time*, qui traduit les parties les plus utilisées du programme assembleur virtuel directement en "vraies" instructions machines. Exemple : Dalvik pour Android.

**Ne pas confondre...** *Au passage, ne pas confondre machine virtuelle comme dans JVM (Java Virtual Machine) et machine virtuelle comme dans VirtualBox.*

*La première émule<sup>a</sup> un processeur différent du processeur matériel, qui peut même être irréaliste en matériel.*

*La seconde émule la machine au sens boîtier avec ses périphériques etc, mais avec le même processeur, en faisant au final tourner le code de la machine émulée aussi directement que possible sur le matériel.*

*Il y a aussi des émulateurs de machines. Par exemple je peux faire tourner dans une machine virtuelle VirtualBox un navigateur qui utilise une machine virtuelle Java pour faire tourner dedans un programme émulateur de Gameboy qui émule entre autre le processeur Z80 de la Gameboy. Merci à la loi de Moore, avec ces 4 couches d'émulation/virtualisation, les jeux tournent quand même plus vite que sur la console d'origine...*

*Il y a aussi la notion de mémoire virtuelle, qu'on verra dans le chapitre suivant, et qui est encore autre chose.*

a. Un jour on m'a demandé ce que cela veut dire "émuler" et j'étais bien embêté. Le Larousse dit que ce verbe ne sert qu'en informatique : c'est quand un ordinateur fait semblant d'être un autre ordinateur. On peut chercher un rapport lointain avec l'émulation en classe.

Dans toute la suite on parle de langages compilés nativement, et en fait on va surtout compiler le C<sup>1</sup>.

### 12.2 L'arrière-cuisine du compilateur

#### 12.2.1 Variables et expressions

La sémantique du C est orientée mémoire : il faut comprendre une variable comme le contenu d'une case mémoire (ou de plusieurs cases consécutives si nécessaires). D'ailleurs on peut demander l'adresse de n'importe quelle variable par l'opérateur &. Par exemple si la variable A est stockée à l'adresse @A, la ligne de C A=A+1 ; se décompose en une séquence d'instructions machines qui vont 1/rapatrier le contenu de la case d'adresse @A dans un registre du processeur ; 2/ ajouter 1 à ce registre, et 3/ écrire en mémoire le résultat à l'adresse @A.

1. C'est une contrepéterie.

Si vous ne demandez pas d'optimisation particulière à gcc, le code assembleur produit ressemble exactement à cela, et c'est pourquoi vous trouverez des séquences de code visiblement stupides, par exemple dans mon programme jouet compilé vers ARM :

```
str r3, [fp, #-8]
ldr r3, [fp, #-8]
```

Oui, il s'agit bien d'une écriture d'un registre à une adresse, immédiatement suivie d'une lecture de la même adresse pour remettre son contenu dans le registre où il est déjà. C'est idiot. Explication : le programme est compilé instruction C par instruction C ; l'écriture termine une instruction C (variable=expression;). La lecture commence l'instruction C suivante.

Au passage, voilà pourquoi on utilise le même mot dans "compilation d'un programme" et "compilation des meilleurs tubes des années 70". Dans les deux cas c'est la juxtaposition de morceaux indépendants.

Si vous branchez les optimisations (gcc -O1, -O2, etc) le compilateur fera le ménage dans ces instructions inutiles.

Une expression arithmétique est un arbre, et en général il faut parcourir l'arbre pour réaliser les calculs intermédiaires dans le bon ordre, éventuellement en les stockant dans des cases mémoires temporaires.

### 12.2.2 Désucrage des opérations de contrôle de flot

**Conditionnelle if-then-else** On peut toujours implémenter le `if-then-else` par un test et un saut conditionnel (et possiblement deux autres sauts non conditionnels en cas de `else`).

Si le jeu d'instruction supporte les instructions conditionnelles (le *conditional move* est assez courant) ou les instructions prédiquées, on peut aussi implémenter un `if-then` ou un `if-then-else` par prédication. La prédication consiste à conditionner au résultat du test l'exécution d'instructions qui ne sont pas des sauts. Par exemple, pour un `if-then-else`, on exécutera les deux branches pour au final ne retenir que celle qui correspond au résultat du test.

Cela peut valoir le coup parce que les branchements sont coûteux dans un processeur pipeliné.

**Boucles** L'implémentation d'une boucle (For, Do, While) se fait selon une variation du schéma suivant (se reporter au document donnant la sémantique du langage) :

1. évaluer l'expression contrôlant la boucle
2. si elle est fausse, brancher après la fin de la boucle, sinon commencer le corps de boucle
3. à la fin du corps de boucle, réévaluer l'expression de contrôle
4. si elle est vraie, brancher au début de la boucle sinon continuer sur l'instruction suivante

Il y a donc un branchemennt par exécution du corps de boucle. En général un branchemennt produit une certaine latence (quelques cycles).

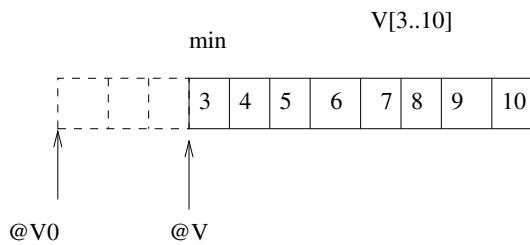
**Case aka switch** La difficulté d'implémentation des instructions `case` (e.g. `switch` en C) est de sélectionner la bonne branche efficacement. Le moyen le plus simple est de faire une passe linéaire sur les conditions comme si c'était une suite imbriquée de `if-then-else`. Dans certains cas, les différents tests de branchemennt peuvent être ordonnés (test d'une valeur par exemple), dans ce cas on peut faire une recherche dichotomique du branchemennt recherché ( $O(\log(n))$  au lieu de  $O(n)$ ).

On peut aussi faire des sauts à une adresse calculée à partir de la variable du switch, ou utiliser une table d'adresses. Dans ce cas, la décision est en temps constant.

### 12.2.3 Tableaux

Le stockage et l'accès aux tableau sont extrêmement fréquents et nécessitent donc une attention particulière. Commençons par la référence à un simple vecteur (tableau à une dimension). Considérons que  $V$  a été déclaré par  $V[min \dots max]$ . Pour accéder à  $V[i]$ , le compilateur devra calculer l'*offset* de cet élément du tableau par rapport à l'adresse de base à partir de laquelle il est stocké. L'offset est  $(i - min) \times w$  où  $w$  est la taille des éléments de  $w$ . Si  $min = 0$  et  $w$  est une puissance de deux, le code généré s'en trouvera simplifié (les multiplications par des puissances de 2 se font par décalage de bits, ce qui est en général plus rapide qu'une multiplication).

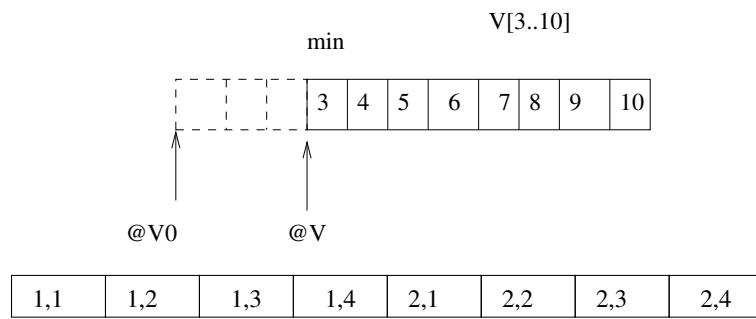
Si la valeur de  $min$  est connue à la compilation, le compilateur peut calculer l'adresse (virtuelle)  $@V_0$  qu'aurait le tableau s'il commençait à 0 (on appelle quelquefois cela le faux zéro), cela évite une soustraction à chaque accès aux tableaux :



Si la valeur de  $min$  n'est pas connue à la compilation, le faux zéro peut être calculé lors de l'initialisation du tableau.

Pour les tableaux multidimensionnels, il faut choisir la façon d'envoyer les indices dans la mémoire. Il y a essentiellement trois méthodes, les schémas par lignes et par colonnes et les tableaux de vecteurs.

Par exemple, le tableau  $A[1..2, 1..4]$  comporte deux lignes et quatre colonnes. S'il est rangé par lignes (*row major order*), on aura la disposition suivante en mémoire :



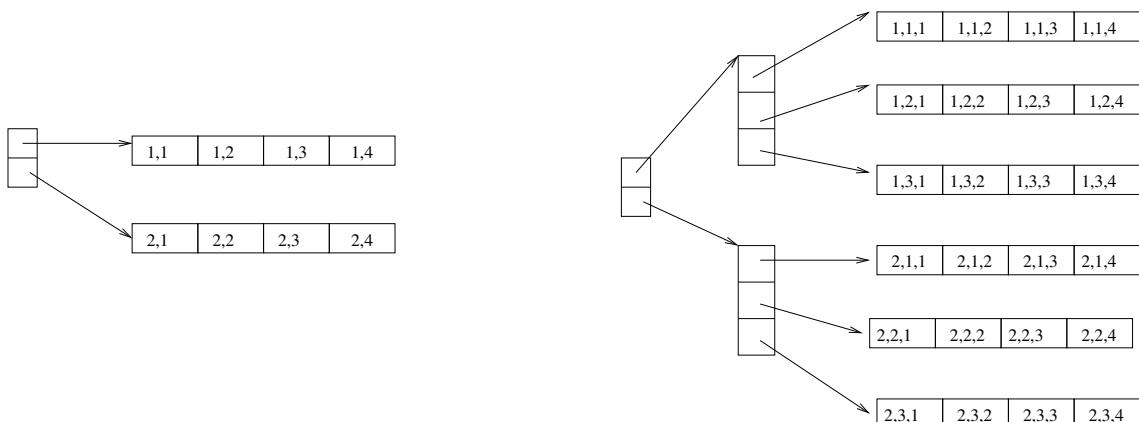
S'il est rangé par colonne on aura :



Lors d'un parcours du tableau, cet ordre aura de l'importance. En particulier si l'ensemble du tableau ne tient pas dans le cache. Par exemple avec la boucle suivante, on parcourt les cases par lignes ; si l'on intervertit les boucles  $i$  et  $j$ , on parcourra par colonnes :

```
for i ← 1 to 2
    for j ← 1 to 4
         $A[i, j] \leftarrow A[i, j] + 1$ 
```

Certains langages (java) utilisent un tableau de pointeurs sur des vecteurs. Voici le schéma pour  $A[1..2, 1..4]$  et  $B[1..2, 1..3, 1..4]$



Ce schéma est plus gourmand en place (la taille des vecteurs de pointeur grossit quadratiquement avec la dimension) et l'accès n'est pas très rapide, de plus le programmeur doit écrire une boucle pour allouer le tableau ligne par ligne. Concernant la place, ce schéma permet toutefois d'implanter efficacement des tableaux non rectangulaires (matrices triangulaires) mais c'est sans doute anecdote.

Considérons un tableau  $A[min_1..max_1, min_2..max_2]$  rangé par ligne pour lequel on voudrait accéder à l'élément  $A[i, j]$ . Le calcul de l'adresse est :  $@A[i, j] = @A + (i - min_1) \times (max_2 - min_2 + 1) \times w + (j - min_2) \times w$ , où  $w$  est la taille des données du tableau. Si on nomme  $long_2 = (max_2 - min_2 + 1)$  et que l'on développe on obtient :  $@A[i, j] = @A + i \times long_2 \times w + j \times w - (min_1 \times long_2 + min_2 \times w)$ . On peut donc aussi précalculer un faux zéro  $@A_0 = @A - (min_1 \times long_2 \times w + min_2 \times w)$  et accéder l'élément  $A[i, j]$  par  $@A_0 + (i \times long_2 + j) \times w$ . Si les bornes et la taille ne sont pas connues à la compilation, on peut aussi effectuer ces calculs à l'initialisation. Les mêmes optimisations sont faites pour les tableaux rangés par colonnes.

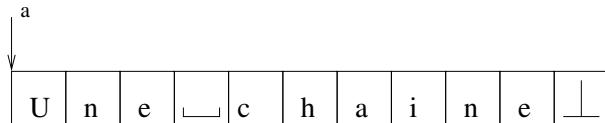
L'accès aux tableaux utilisant des pointeurs de vecteurs nécessite simplement deux instructions par dimension (chargement de la base de la dimension  $i$ , offset jusqu'à l'élément dans la dimension  $i$ ). Pour les machines où l'accès à la mémoire était rapide comparé aux opérations arithmétiques, cela valait le coup (ordinateur avant 1985).

Lorsqu'on passe un tableau en paramètre, on le passe généralement par référence, même si dans le programme source, il est passé par valeurs. Lorsqu'un tableau est passé en paramètre à une procédure, la procédure ne connaît pas forcément ses caractéristiques (elles peuvent changer suivant les appels comme en C). Pour cela le système a besoin d'un mécanisme permettant de récupérer ces caractéristiques (dimensions, taille). Cela est fait grâce au descripteur de tableau (*dope vector*). Un descripteur de tableau contient en général un pointeur sur le début du tableau et les tailles des différentes dimensions. Le descripteur a une taille connue à la compilation (dépendant uniquement de la dimension) et peut donc être alloué dans l'AR de la procédure appelée.

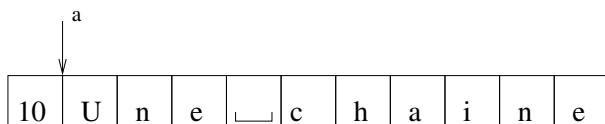
Beaucoup de compilateurs essaient de détecter les accès en dehors des bornes des tableaux. La méthode la plus simple est de faire un test sur les indices avant chaque appel à l'exécution. On peut faire un test moins précis en comparant l'offset calculé à la taille totale du tableau.

#### 12.2.4 Chaînes de caractères

La manipulation des chaînes de caractères est assez différente suivant les langages. Elle utilise en général des instructions opérant au niveau des octets. Le choix de la représentation des chaînes de caractères a un impact important sur les performances de leur manipulation. On peut stocker une chaîne en utilisant un marqueur de fin de chaîne comme en C traditionnel :



ou en stockant la longueur de la chaîne au début (ou la taille effectivement occupée, si ce n'est pas la taille allouée) :



La plupart des assembleurs possèdent des opérations pour manipuler les caractères (sur les ARM, LDRB au lieu de LDR pour charger un octet. Pour les MSP430, exercice : allez chercher dans la doc donnée en TP). Pour cette raison, la mémoire est le plus souvent également adressée par octet, même pour une ISA pure 32bit, comme ARM.

Toutefois, sur ARM on impose que les chargements de mots de 32 bits (par LDR, mais aussi le chargement des instructions) soit uniquement sur des adresses multiples de 4 (octets). Sur les pentium modernes aussi, il est très conseillé d'*aligner* les accès mémoire pour les données 32 et 64 bits sur des adresses multiples respectivement de 4 et 8 octets. Sinon, je crois que cela marche mais je suis certain que cela prend plus du double du temps.

Cela dit il y a eu des architectures 32 bits où la mémoire était adressée par mots de 32 bits (le DEC Alpha arrivait à tourner à des fréquences bien supérieures à ses concurrents, grâce en partie à ce genre de sacrifice). Dans ce cas on doit faire des décalages et des masques pour extraire un octet particulier d'un mot 32bits.

Pour les manipulations de chaînes (comme pour les tableaux d'ailleurs), beaucoup d'assembleurs possèdent des instructions de chargement et de stockage avec autoincrément (pré ou post), c'est à dire que l'instruction incrémentera l'adresse servant à l'accès en même temps qu'elle fait l'accès. Ceci permet de faire directement l'accès suivant au mot (resp. caractère suivant) juste après.

### 12.2.5 Structures

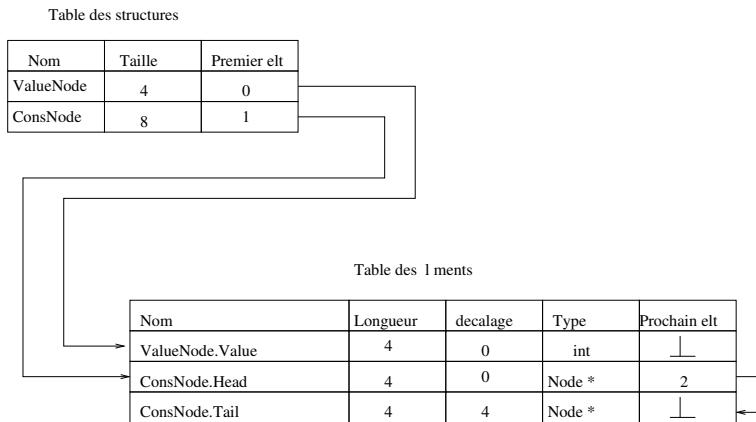
Les structures sont utilisées dans de nombreux langages. Elles sont très souvent manipulées à l'aide de pointeurs et créent donc de nombreuses valeurs ambiguës. Par exemple, pour faire une liste en C, on peut déclarer les types suivants :

```
struct ValueNode {
    int Value
};

struct ConsNode {
    Node *Head;
    Node *Tail;
};
```

```
Union Node{
    struct ValueNode;
    struct ConsNode;
};
```

Afin d'émettre du code pour une référence à une structure, le compilateur doit connaître l'adresse de base de l'objet ainsi que le décalage et la longueur de chaque élément. Le compilateur construit en général une table séparée comprenant ces informations (en général une table pour les structures et une table pour les éléments).



Avec cette table, le compilateur génère du code. Par exemple pour accéder à  $p1 \rightarrow \text{Head}$  on peut générer l'instruction :

$loadA0 \ r_{p1},0 \Rightarrow r_2 \quad // 0 \text{ est le décalage de 'Head'}$

Si un programme déclare un tableau de structure dans un langage dans lequel le programmeur ne peut pas avoir accès à l'adresse à laquelle est stockée la structure, le compilateur peut choisir de stocker cela sous la forme d'un tableau dont les éléments sont des structures ou sous la forme d'une structure dont les éléments sont des tableaux. Les performances peuvent être très différentes, là encore à cause du cache.

## 12.3 Application binary interface

l'ABI (*application binary interface*) est un contrat entre le compilateur, le système d'exploitation et la machine cible. Le but de ce contrat est de fixer les règles qui permettront à ces trois acteurs de coopérer harmonieusement :

- Ce contrat répartit les responsabilités de chacun, et fixe des règles du jeu pour
- l'allocation des ressources (quel registre est réservé pour quoi ? qui est chargé de réserver l'espace mémoire des différentes variables ? quel registre sera utilisé comme le pointeur de pile lorsqu'il y a le choix comme sur ARM ?),
  - le comportement (quels registres une procédure a le droit d'écraser, quels registres une procédure a le devoir de rendre dans l'état où elle l'a trouvé),
  - le placement en mémoire (où est le point d'entrée d'une procédure, où sont ses variables locales et globales ?).

Nous allons décrire une ABI typique pour le C et Fortran, qui doit permettre

- la compilation séparée des procédures et la construction de bibliothèques.
- les procédures récursives

L'ABI est contrainte par les possibilités de l'ISA, mais est indépendante de l'ISA : c'est essentiellement un ensemble de conventions. Par exemple Windows et Linux, sur le même processeur, n'ont pas la même ABI.

### 12.3.1 Procédures et compilation séparée

Pour la plupart des langages impératifs, la procédure est l'unité de base. Le compilateur permet par exemple la *compilation séparée* des procédures. C'était indispensables aux temps héroïques de l'informatique. De nos jours, le moindre PC peut en théorie gérer dans sa mémoire centrale des programmes énormes, mais on continue à faire de la compilation séparée pour ses nombreux avantages :

- gain de temps lors du développement : on ne recompile que ce que est nécessaire,
- dualement, possibilité de passer plus de temps à faire des optimisations poussées sur des morceaux de codes assez petits,
- possibilité de distribuer des bibliothèques pré-compilées, sans distribuer le code source
- Possibilité d'édition de lien dynamique.

Tous ces avantages viennent au prix d'une étape d'*édition de lien*, qui assemble le code compilé des différentes procédures pour en faire un programme complet. L'édition de lien autorise à compiler du code pour une procédure, indépendamment du programme dans lequel elle va s'insérer.

La compilation séparée est encore plus intéressante si l'on peut appeler depuis un programme en C une bibliothèque écrite en Fortran, et réciproquement. Pour que ce soit possible, il faut que l'interfaces des procédures au niveau du code objet soit compatible avec tous les langages.

Cette interface doit donc offrir l'union de tous les langages cibles envisagés, et pourra donc paraître inutilement compliquée lorsqu'on regarde le code compilé pour un langage simple comme C. Deux exemples :

- Pascal a des procédures imbriquées, pas C<sup>2</sup>. Une procédure imbriquée, c'est une déclaration de procédure locale à une autre procédure. Il y a donc plusieurs espaces de nommage imbriqués.
- Les langages objets nécessitent des mécanismes spécifiques pour permettre héritage et surcharge, encore une fois pour savoir à quelle sur-classe appartient une variable d'un objet. Ces mécanismes seront inutilisés en C.

La compilation d'une procédure doit faire le lien entre la vision haut niveau (espace des noms hiérarchique, code modulaire) et l'exécution matérielle qui voit la mémoire comme un simple tableau linéaire et l'exécution comme une simple incrémentation du PC.

### 12.3.2 Récursivité et pile

Considérons l'exécution d'une procédure récursive (pensez à la factorielle ou à Fibonacci). Lorsque j'appelle `factorielle(10)`, l'ordinateur doit

1. réservé de la place pour les variables locales de la procédure `factorielle()`
2. brancher au début du code de `factorielle()`
3. revenir d'où il vient à la fin de ce code.

Les processeurs offrent tous une instruction machine qui permet ce type de "branchement en se souvenant d'où on vient".

- Sur la plupart des processeurs (IA32 et MSP430 inclus), on a deux instructions, `call` et `ret`. L'instruction `call` `adresse` réalise deux opérations :

1. empiler l'adresse de l'instruction suivant le `call` (PC+1, pour une certaine valeur de 1)
2. faire `PC<-adresse`

L'instruction `ret` n'a pas d'argument, mais c'est aussi un branchement : elle dépile une adresse, puis branche à cette adresse.

- Sur les ARM et quelques autres, `call` est remplacé par `BL` `adresse` qui fait également deux choses :

1. une copie de R15+4 (le PC de l'instruction suivant le `BL`) dans R14 (le *link register*)
2. puis un saut vers `adresse`

L'instruction `ret` est remplacée par un `MOV R14 -> R15`. L'intérêt est de ne pas faire d'accès mémoire quand on n'en n'a pas besoin (appel à une procédure terminale, c'est-à-dire qui n'appelle personne d'autre). En cas d'appel récursif, la procédure appelée devra quand même empiler R14 au début, et le dépiler à la fin.

2. A vrai dire, `gcc` accepte les procédures imbriquées, c'est une extension GNU à C. Mais je vous défends formellement de vous en servir

Je rappelle que toutes ces instructions sont atomiques, c'est-à-dire ininterruptibles. On pourrait en principe les émuler chacune par plusieurs instructions (des push, des pop et des sauts) mais ce serait le bordel en cas d'interruption au milieu.

Donc on a une *pile d'appel*. La récursivité nécessite fondamentalement un mécanisme de pile.

### 12.3.3 Variables locales et passage de valeurs sur la pile

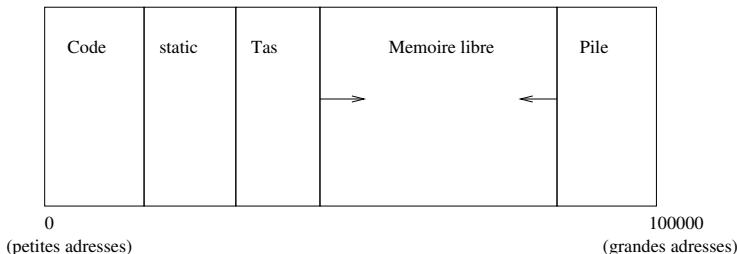
Cette pile est également l'espace de stockage idéal pour les variables locales d'une procédure (voir la figure 12.1) :

- L'allocation et la déallocation se font par des additions et soustractions sur le pointeur de pile, au début et à la fin de la procédure. Pas besoin de `malloc` compliqué.
- Pour ma factorielle, les différentes *instances* de la fonction (`factorielle(10)` a appelé `factorielle(9)` qui a appelé `factorielle(8)`, etc) ont chacune leurs variables locales sur la pile. En haut de la pile, j'ai toujours les variables locales de la procédure en cours d'exécution.
- Du coup, pour le processeur, l'adresse d'une variable locale est  $SP+cste$ . La constante est appelée aussi *offset* de la variable.
- Le passage des paramètres peut également se faire sur la pile : lorsque `toto` appelle `tata(17, 42)`, les valeurs 17 et 42 seront empilées juste avant le `call tata`. La procédure appelée sait qu'elle a deux paramètres et qu'elle les trouvera aux adresses  $SP-1$  et  $SP-2$ . On pourrait aussi les passer dans des registres, mais on a parfois plus de paramètres à passer que de registres<sup>3</sup>
- De même, la procédure peut passer sa ou ses valeurs de retour sur la pile, où l'appelant les retrouvera au retour du `call`<sup>4</sup>.

### 12.3.4 Vision globale de la mémoire

Tous les langages permettent aussi d'allouer de la mémoire dynamiquement (par exemple pour les variables dont la taille n'est pas connue à la compilation). Cette mémoire est allouée sur le *tas*. Typiquement, les variables dont la taille est petite et connue à la compilation (entiers, flottants, pointeurs sur les objets) sont allouées dans la pile. Par contre les tableaux, les objets, etc sont alloués sur le tas. En C, C++ ou Java cette allocation est explicite (il faut des `malloc` ou des `new`).

La mémoire linéaire est généralement organisée de la façon suivante : la pile sert à l'allocation des appels de procédures (elle grandit par exemple vers les petites adresses), le tas sert à allouer la mémoire dynamiquement (il grandit alors vers les grandes adresses). Ceci aussi fait partie de l'ABI, en particulier pour l'ARM pour lequel n'importe quel registre peut servir de pointeur de pile, et les piles peuvent monter ou descendre.



Ce n'est pas encore tout. Une donnée peut, suivant sa portée et son type, être stockée en différents endroits de la mémoire (on parle de classe de stockage, *storage class*). Les principales sont

- stockage local à une procédure (sur la pile, déjà vu),
- stockage dynamique (sur le tas, déjà vu),
- stockage statique d'une procédure (à côté du code),
- stockage global.

On va s'arrêter là dans les détails. Et on pourrait reprendre pour un langage orienté objet.

### 12.3.5 Sauvegarde des registres

Si on ne prend pas de précaution, `call` est une instruction très particulière : elle peut potentiellement changer les valeurs de tous les registres ! En effet, l'appelé va utiliser les registres comme bon lui semble.

3. Les vraies ABI définissent des politiques du genre "si la procédure a moins de 4 paramètres, ils sont passés dans les registres R0 à R4. Sinon, les 4 premiers sont passés dans les registres et tous les suivants sur la pile. Une conséquence est qu'on peut observer une grosse dégradation de performance en ajoutant juste un paramètre à une procédure, cela m'est arrivé..."

4. ... ou dans des registres, voir la note précédente.

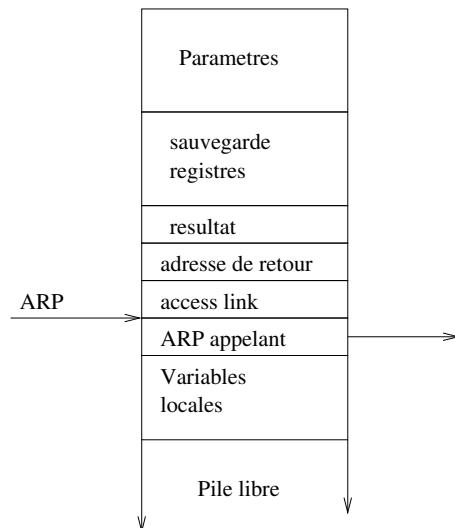


FIGURE 12.1 – Enregistrement d'activation sur la pile

De plus, dans le trip de la compilation séparée,

- l'appelant ne sait pas comment est compilé l'appelé, donc quels registres il va utiliser.
- l'appelé ne sait pas qui l'appelle, donc quel registre l'appelant est en train d'utiliser.

L'ABI doit définir également un code de bonne conduite qui peut être une des variantes suivantes :

1. C'est la responsabilité de l'appelant de sauvegarder (en les empilant) les registres dont il veut conserver la valeur après le `call`. L'appelé n'a aucun souci à se faire, il peut utiliser tous les registres comme bon lui semble.
2. C'est la responsabilité de l'appelé de sauvegarder (en les empilant) les registres qu'il va utiliser. L'appelant n'a aucun souci à se faire, toutes les valeurs des registres seront inchangées après le `call`.
3. Pareil que 2/, mais pas pour tous les registres : les registres Rx à Ry peuvent être écrasés par l'appelé (que l'appelant se le tienne pour dit). Ce sont des *scratch registers*. Par contre, les registres Rz à Rt doivent être préservés par l'appelé (*preserved registers*). L'appelant n'a pas à se soucier de les sauvegarder lui-même.

Pour les détails : lisez l'ABI de votre système.

### 12.3.6 En résumé : l'enregistrement d'activation

L'espace mémoire mis en place en haut de la pile lors de l'appel d'une procédure est appelé *enregistrement d'activation* (en rosbif *activation record* : AR). Cet espace est désalloué lorsque la procédure se termine.

L'enregistrement d'activation comporte la place pour

- les paramètres effectifs,
- les variables locales à la procédure,
- la sauvegarde de l'environnement d'exécution de la procédure appelante (essentiellement les registres)
- l'adresse de retour de la procédure,
- et possiblement d'autres pointeurs, suivant les ABI :
  - un pointeur vers les variables globales;
  - un pointeur sur l'AR de la procédure appelante (utile pour les procédures imbriquées);
  - pour les méthodes de langages objet, un pointeur vers l'objet (`this`), qui est lui-même une structure de donnée compliquée, avec des attributs et des pointeurs vers les surclasses pour accéder aux attributs des surclasses en respectant les règles de visibilité...)
  - ...

Il faut encore insister sur le fait que tout ceci est une convention, un contrat, un choix. Tout est toujours discutable, et les ABI évoluent avec le temps.

Par exemple, voici le code que `gcc` produit typiquement en entrée d'une procédure (`esp` est le pointeur de pile, `ebp` est l'ARP, et la pile descend dans la mémoire comme sur la figure précédente) :

```

.main:
pushl %ebp  (empiler le lien vers l'ARP de l'appelant)
movl %esp,%ebp  (l'ARP de cette procédure sera le pointeur de pile à cet instant)
subl $24,%esp  (réservier sur la pile l'espace pour les variables locales)

```

### 12.3.7 En résumé : code à générer par le compilateur

Le code précédent est le prologue de la procédure. Chaque procédure contient un prologue et un épilogue. De même, chaque appel contient un pré-appel et un post-appel.

- **Pre-appel** Le pré-appel met en place l'AR de la procédure appelée (allocation de l'espace nécessaire et remplissage des emplacements mémoire). Cela inclut l'évaluation des paramètres effectifs, leur stockage sur la pile, l'empilement de l'adresse de retour, l'ARP de la procédure appelante, et éventuellement la réservation de l'espace pour une valeur de retour. La procédure appelante ne peut pas connaître la place nécessaire pour les variables locales de la procédure appelée.
- **Post-appel** Il y a libération des espaces alloués lors du pré-appel.
- **Prologue** Le prologue crée l'espace pour les variables locales (avec initialisation éventuelle). Si la procédure contient des références à des variables statiques, le prologue doit préparer l'adressabilité en chargeant le label approprié dans un registre.
- **Épilogue** L'épilogue libère l'espace alloué pour les variables locales, il restore l'ARP de l'appelant et saute à l'adresse de retour. Si la procédure retourne un résultat, la valeur est transférée là où elle doit être utilisée.

De plus, les registres doivent être sauvegardés, soit lors du pré-appel de l'appelant, soit lors du prologue de l'appelé.

En général, mettre des fonctionnalités dans le prologue ou l'épilogue produit du code plus compact (le pré-appel et post-appel sont écrits pour chaque appel de procédure).

## 12.4 Compilation des langages objets

Fondamentalement, l'orientation objet est une réorganisation de l'espace des noms du programme d'un schéma centré sur les procédures vers un schéma centré sur les données. Donc il faudrait tout reprendre... J'espère vous avoir donné les clés pour que vous compreniez la doc si vous avez un jour à la lire.



# Chapitre 13

## Hiérarchie mémoire

Dans ce chapitre on va voir différents raffinements du système mémoire de notre PC, pour plus de performance ou pour plus de fonctionnalité. Le matériel chargé de cette gestion s'appelle la MMU (memory management unit) et est de nos jours directement intégré sur la puce du processeur.

### 13.1 Mémoire cache

Le principe du cache est de reprendre la table 4.1 page 51, et d'y intercaler des mémoires de taille et de performance intermédiaire à deux endroits :

- entre les registres et la mémoire centrale,
- entre la mémoire centrale et le disque.

On obtient la table 13.1.

#### 13.1.1 Principes de localité

Pourquoi un cache est utile ? Parce que

1. Une adresse mémoire actuellement accédée a toutes les chances d'être accédée à nouveau dans un futur proche (localité temporelle)
2. Une adresse proche d'une adresse mémoire actuellement accédée a toutes les chances d'être accédée aussi dans un futur proche (localité spatiale)

Exemples : les boucles (il s'agit alors d'adresses du code). Le tri d'un tableau (adresses de données). Etc. Ce principe de localité est très général. Contre-exemple : le produit de matrices si la matrice est très grande. Solution : organiser le produit de matrice par blocs. Il existe une bibliothèque, ATLAS pour Automatically Tuned Linear Algebra, qui définit la taille des blocs en fonction de la taille du cache.

#### 13.1.2 Scratchpad versus cache

On peut laisser au programmeur/compilateur la charge de rapatrier manuellement dans le cache les données actuellement/localement utilisées. Dans ce cas on parle plutôt de *scratchpad memory*, ou mémoire brouillon. Le surcoût matériel se limite alors à la mémoire cache elle-même. Par contre il y a un surcoût de difficulté de programmation/compilation, ainsi qu'un surcoût en terme de taille de code, puisqu'il faut des instructions qui servent à rapatrier des données dans le cache.

Pour éviter cela, le plus souvent, lorsqu'on parle de cache, on parle d'un mécanisme automatique qui intercepte les requêtes mémoires, et les aiguille soit vers le cache soit vers la mémoire, en gérant au

Type	temps d'accès	capacité typique
Registre	0.1 ns	1 à 128 mots (de 1 à 8 octets)
Mémoire cache	1-4ns	4Koctets – 4Mo
Mémoire vive	10 - 100 ns	4 Goctets
Cache disque	100ns	512Koctets
Disque dur	10ms	100Gocets
Archivage	1mn	(illimité)

TABLE 13.1 – Hiérarchie mémoire d'un ordinateur en 2010

passage le remplissage du cache avec les données actuellement utilisées. C'est à ce mécanisme qu'on s'intéresse dans la suite.

Ce mécanisme sera implanté en matériel pour les accès mémoire, et plutôt en logiciel pour les accès disques qui sont de toute manière lents. Mais les principes sont les mêmes. Nous allons nous concentrer sur le cache mémoire.

Construire ce mécanisme de gestion du cache a un coût matériel et énergétique qui s'ajoute au coût de la mémoire cache elle-même. Mais cela simplifie la vie du programmeur.

Considérant tout cela, on trouve des scratchpad sur les systèmes embarqués : on veut éviter la consommation électrique de la MMU, et par ailleurs leur code est statique et connu à l'avance. Par ailleurs le scratchpad est explicite dans les environnements de programmation OpenCL et OpenVX. On trouve des vrais caches sur tous les systèmes un peu plus ouverts. Scratchpad versus cache, c'est un peu la même discussion que VLIW versus superscalaire. Le processeur Kalray offre au choix scratchpad (pour le temps réel/la prédictibilité) et cache (pour la simplicité de programmation).

### 13.1.3 Cache hit et cache miss

Et donc on va garder dans le cache les adresses (et les données correspondantes) accédées récemment (cf figures 13.1 ou 13.2 ci-dessous). Un accès mémoire commencera par comparer l'adresse demandée avec celles qui sont présentes dans le cache. En cas de succès (souvent) la donnée sera lue depuis le cache (*cache hit*). En cas d'échec (défaut de cache ou *cache miss*, rarement espère-t-on), la donnée sera d'abord chargée de la mémoire vers le cache. Elle y sera présente pour cet accès et aussi pour les accès suivants (localité temporelle). Et au passage, le cache chargera aussi les quelques données suivantes pour exploiter la localité spatiale.

Concrètement, le cache charge depuis la mémoire centrale une *ligne de cache* à la fois. La ligne de cache est l'unité d'échange entre le cache et la mémoire, et fait quelques centaines d'octets. Une ligne de cache de taille  $2^l$  est définie par "toutes les adresses ayant tous leurs bits identiques sauf les  $l$  de poids faible".

Une fois définie la ligne de cache, et si on a les moyens, on peut organiser notre mémoire physique pour qu'elle soit adressée par ligne uniquement. La mémoire physique sera alors construite pour réagir à des adresses sans les  $l$  derniers bits et envoyer  $2^l$  octets en parallèle sur un bus très large entre le processeur et la mémoire. On verra cela en 13.1.7.

### 13.1.4 Hiérarchie mémoire

Une mémoire cache se comporte, vu de l'extérieur, comme une RAM normale, mais réagit plus rapidement. Par exemple, le cache disque est à peine plus lent que la mémoire vive – un peu tout de même car les bus qui y accèdent sont moins larges que ceux qui accèdent à la mémoire vive.

Il peut en fait y avoir plusieurs niveaux de mémoire cache :

- le premier sur la puce même du processeur, au plus près de chaque cœur;
- le second toujours sur la puce, mais partagé entre plusieurs cœurs, donc un peu plus loin de chacun;
- le troisième sur une puce séparée, mais avec un bus très large (128-256 bits) avec le processeur.

De même, il peut y avoir un cache disque dans le disque dur, et un autre dans la mémoire centrale. Dans la suite, on ne va considérer qu'un niveau car les mécanismes sont les mêmes lorsqu'il y en a plusieurs.

### 13.1.5 Construction d'un cache

Comme le montre la figure 13.1, à chaque ligne de cache est associée dans le cache son adresse virtuelle, ainsi qu'un bit de validité (par exemple, toutes les lignes du cache sont invalides lorsqu'on allume l'ordinateur). On voit que d'organiser le cache en lignes (et pas en mots) ne sert pas qu'à exploiter la localité spatiale : cela sert aussi à minimiser le surcoût de ces données supplémentaires. Pour un cache organisé par mots, il faudrait autant de bits pour les données que pour leur adresse...

Lorsque le processeur envoie une demande de lecture à une adresse donnée, il faut savoir si cette adresse est présente dans le cache. Pour cela, le cache devrait comparer l'adresse recue (sans ses  $l$  derniers bits) avec toutes les adresses présentes dans le cache. C'est coûteux, donc on fait des compromis :

- Cache *Direct-mapped* : une ligne ne peut aller qu'à un seul endroit dans le cache, défini par les bits de poids faible de l'adresse de la ligne. Simple à réaliser, mais conflits nombreux.
- Cache *Fully associative* : une ligne peut aller n'importe où dans le cache. Le cache est alors une mémoire associative, comme un dictionnaire dans lequel l'adresse est le mot cherché, et la donnée est la définition. Cela implique que le cache doit comparer une adresse d'entrée avec toutes ses adresses, et en parallèle pour que ce soit rapide : c'est coûteux en matériel.

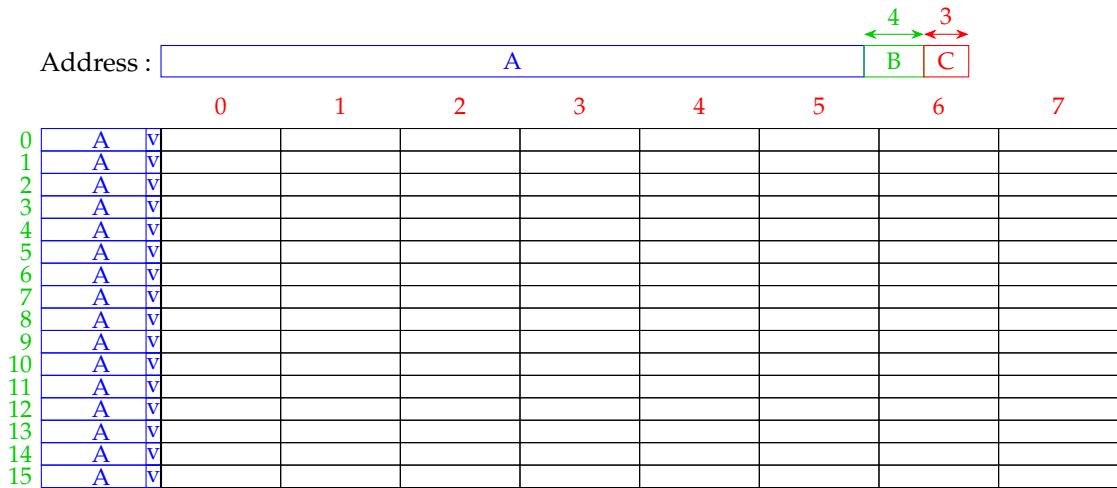


FIGURE 13.1 – Structure d'un cache direct-mapped avec 16 lignes de 8 mots. C est la position dans la ligne, B est directement le numéro de la ligne dans le cache, et A est l'adresse de la ligne en mémoire, stockée avec la ligne et assortie d'un bit de validité.

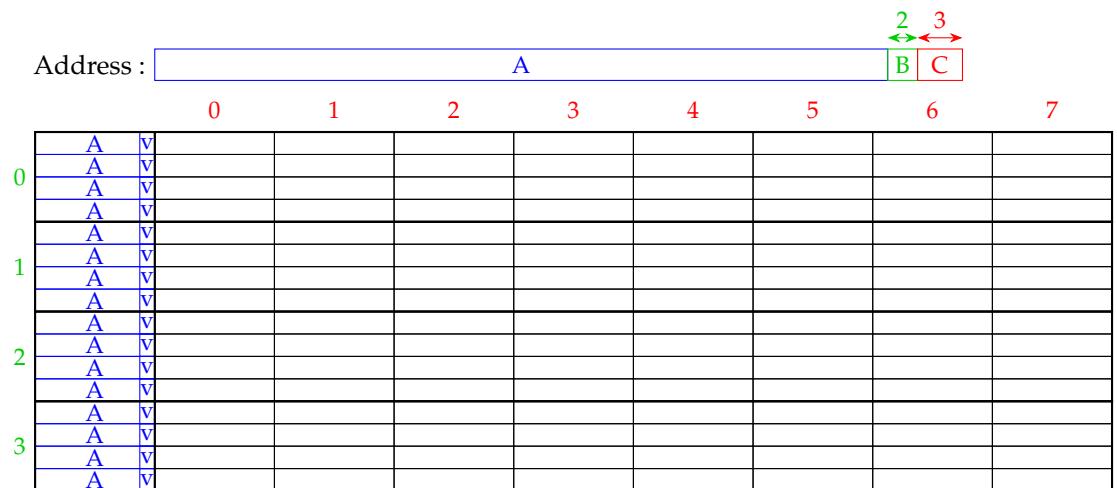


FIGURE 13.2 – Structure d'un cache 4-way set associative. C est la position dans la ligne, B est le numéro de la voie dans le cache, et A est l'adresse de la ligne en mémoire, stockée avec la ligne et assortie d'un bit de validité.

- Cache *n-way set associative* : c'est un compromis intermédiaire. Une ligne peut aller dans le cache dans un ensemble de taille  $n = 2, 4$  ou  $8$  lignes. Il faut comparer l'adresse de la ligne avec les adresses juste dans cet ensemble.

On a fait des statistiques, du genre *2 :1 Cache Rule* : *The miss rate of a direct-mapped cache of size  $N$  is about the same as that of a two-way set-associative cache of size  $N/2$ .* (Hennessy-Patterson, CAQA 1ère édition, page 0)

### Éviction du cache

Un problème se pose lorsque le cache est plein, ce qui arrive assez vite puisque sa capacité est réduite. Pour ajouter une nouvelle ligne, il faut en virer une. Laquelle ?

En correspondance directe on n'a pas le choix.

Dans les autres cas, en principe on aimerait virer la moins utile dans le futur, mais... on ne connaît pas le futur. Alors on le prédit en fonction du passé : en première approximation, la ligne à virer est la moins utilisée dans le passé (*least recently used* ou LRU).

Mais pour cela, il faut accrocher une date à chaque ligne. Encore des frais ! Mais pas tant que cela : tout ce qu'on veut savoir, c'est qui est le plus ancien. Par exemple, pour un cache à 2 voies il suffit d'un bit par ligne : à chaque accès on met à 1 le bit de la ligne accédé et à 0 le bit de l'autre. Pour les caches à plus de 2 voies il faut plus de bits et il faudra les comparer tous, donc c'est compliqué. On re-approxime, par exemple on fait deux paquets de 2 voies et on a un bit qui dit quel paquet a été LRU, et dans chaque paquet on a un bit LRU comme pour le cache à deux voies.

Une technique qui marche bien aussi est de remplacer une ligne au hasard...

### Écriture dans un cache

On n'a parlé que de lectures, pas encore des écritures mémoire. Convainquons-nous d'abord qu'elles sont bien moins fréquentes que les lectures. Déjà, le code est surtout lu. Et pour les données, considérez le produit de matrice de service : on fait  $n^2$  écritures pour  $n^3$  lectures.

Deux stratégies en cas d'écriture :

- écriture directe dans la mémoire centrale (*write through*) : lent à chaque écriture, puisqu'on doit attendre d'avoir fini l'écriture dans la mémoire centrale (des fois que l'instruction suivante soit une lecture à cette adresse). Par contre le cache est toujours *cohérent* avec la mémoire centrale, donc on peut virer une ligne sans plus de travail
- écriture différée (*write back*) : on n'écrit que dans le cache, ce qui est rapide. Mais le cache devient incohérent avec la mémoire centrale. Lorsqu'on vire une ligne (éviction), il faut d'abord la recopier en mémoire centrale.

Ces questions de *cohérence* deviennent cruciales lorsqu'on a un ordinateur multiprocesseur (ou multicœur, c'est plus la mode) à mémoire partagée. Dans ce cas il faut que tous les caches soient cohérents entre eux. Ce qui veut dire que toute écriture à l'adresse  $a$  doit invalider toutes les lignes de caches contenant  $a$  dans tous les autres processeurs (c'est moins cher d'invalider que de propager des copies dont on n'est pas certain qu'elles seront utiles). Les protocoles qui assurent cette cohérence deviennent vite subtils, et il y a souvent des bugs de ce type dans les révisions initiales des processeurs. Heureusement, contrairement au bug de la division du pentium, ils peuvent être cachés à l'utilisateur par le système d'exploitation... au prix d'une performance réduite.

### Cache unifié ou cache séparé

Il peut être utile d'avoir un cache *séparé programmes/données*, ce qui a l'avantage supplémentaire d'éviter le conflit sur le bus mémoire entre lecture d'instruction et lecture de donnée. Toutefois, il risque de ne pas être rempli aussi optimalmement qu'un cache *unifié*, par exemple dans le cas d'un petit nid de boucle traitant beaucoup de données – encore moins produit de matrices. On voit souvent des caches séparés pour le premier niveau de cache. Et des caches unifiés pour les niveaux suivants.

Attention toutefois, les deux caches séparés doivent tout de même être cohérents.

### En résumé

En résumé, il y a un paquet de paramètres à considérer pour un cache :

- unifié ou séparé
- taille de la ligne de cache
- taille du cache
- associativité

- politique de remplacement
- write-through ou write-back

et ceci, pour chaque niveau de la hiérarchie mémoire (combien en faut-il ? Encore un paramètre).

On gère ces paramètres en simulant des processeurs exécutant des benchmarks.

Au final, les hiérarchies actuelles assurent un *miss rate* inférieur à 1% en moyenne.

**Et ya quoi dans mon PC ?** Essayez-voir :

`lscpu --caches`

### 13.1.6 Statistiques et optimisation des caches (\*)

Il faut faire des statistiques sur les accès mémoire pour définir les paramètres du cache.

Ces statistiques dépendent des programmes, et diffèrent selon qu'on parle de code ou de données.

Ensuite on combine ces statistiques à coups de loi d'Amdahl (calculs p. 564 du Patterson/Hennessy).

On constate en particulier que la performance d'un cache peut se dégrader rapidement si on augmente juste la fréquence du processeur sans toucher à la hiérarchie mémoire.

### 13.1.7 Entre le cache et la mémoire physique (\*)

Si on résume, le cache doit, en cas de *miss*, rapatrier toute une ligne de la mémoire physique. Les options sont (de gauche à droite sur la figure 13.3)

- d'avoir un bus de la largeur de la ligne de cache (ici  $8 \times 32$  bits) : on rapatrie une ligne en une latence mémoire, mais c'est coûteux en filasse.
- d'avoir un bus de la largeur d'un mot, et un automate qui génère les adresses consécutives. On rapatrie une ligne en 8 latences mémoires dans cet exemple.
- de pipeliner les lectures mémoires sur ce bus : on envoie toutes les requêtes en lecture, et (après une latence mémoire) on reçoit en rafale toutes les données, une par cycle. Ainsi on rapatrie une ligne en (*nombre de mots mémoires*) + (latence mémoire) cycles. Cette solution présente un bon rapport performance/coût. Les mémoires modernes (DDR et compagnie) sont faites pour cela : elles ont une grande latence initiale, mais ensuite peuvent envoyer des données consécutives en rafale, une donnée par cycle (mode *burst*).

Rappel : cela correspond à l'architecture interne d'une puce mémoire : allez relire la section 4.2.3.

Si on veut faire une puce de mémoire d'1Mbit ( $2^{20}$  bits) de forme carrée, on devra l'organiser en  $2^{10}$  lignes de  $2^{10}$  bits. Mais la RAM n'a pas  $2^{10} = 1024$  broches pour sortir une ligne. Elle va donc d'abord, en utilisant les  $2^{10}$  bits de poids fort, sortir une ligne en parallèle dans un grand registre interne de 1024 bits. Ceci est lent (parce qu'une toute petite cellule mémoire doit charger un long fil) mais ce n'est pas très grave. Ensuite, l'accès au registre interne sera rapide : il utilise des points mémoire statiques, et peut être construit avec des transistors plus gros puisque sa capacité n'est "que" de 1024 bits. Sortir un mot de 32 bits par cycle en utilisant les poids faibles de l'adresse est rapide. Au fait, ce registre interne est aussi utilisé pour re-écrire en mémoire la ligne qu'on vient de lire (car la lecture a été destructrice).

Jadis, le truc était d'utiliser des mémoires *entrelacées*, répondant chacun à son tour, un mot par cycle.

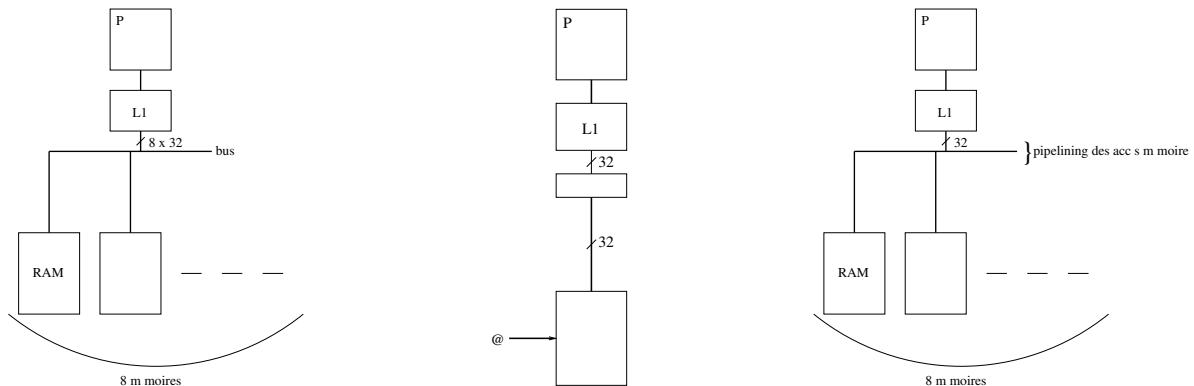


FIGURE 13.3 – Entre cache et mémoire physique

### 13.1.8 Les caches dans un multicœur à mémoire partagée (\*)

On avait déjà un goulot d'étranglement au niveau de l'interface mémoire, cela ne va pas s'arranger si on a plusieurs cœurs qui se battent pour la mémoire. La solution est bien sûr de donner un cache à chaque cœur. La grosse question est alors la *cohérence* des différents caches. Si plusieurs coeurs travaillent sur la même adresse et que l'un l'écrit, il faut propager cette écriture aux autres avant leur prochaine lecture.

Voici un aperçu d'une technique qu'on peut mettre en œuvre dans ce cas : le protocole MESI. On associe, à chaque ligne de cache dans chaque cache, un état qui peut être l'un des suivants :

- I pour Invalide : un accès à une adresse de cette ligne provoquera un défaut de cache (cache miss) pour faire remonter la donnée de la mémoire centrale.
- S pour Partagé (*Shared*). Le cœur a accès en lecture mais pas en écriture, car d'autres cœurs ont aussi cette ligne dans leur cache, également en lecture seulement pour l'instant.
- E pour pour Exclusif : Le cœur a accès en lecture et écriture, parce qu'aucun autre cœur n'a cette ligne dans son cache. Mais il n'a pas encore écrit dedans : la ligne est cohérente avec la mémoire, on peut la virer du cache sans autre forme de procès.
- M pour Modifié : pareil qu'Exclusif, mais il y a eu des écritures dans la ligne, et elle n'est plus cohérente avec la mémoire centrale. Avant de la virer du cache il faudra l'écrire en mémoire (*write back*).

Il n'y a plus qu'à construire l'automate associé à ces 4 états. Les transitions sont provoquées par les instructions de lecture/écriture mémoire des différents cœurs. Vous le trouverez dans tous les bons bouquins. Le matériel responsable de la cohérence des caches sera chargé d'implémenter cet automate.

Il y a plein de communications cachées dans ce protocole. Par exemple, lorsqu'un cœur veut écrire une donnée à une ligne qui est dans l'état S, il doit d'abord lancer une demande d'invalidation de cette ligne à tous ses copains. Il n'est pas question de réaliser l'écriture avant d'être certain que toutes ces invalidations ont été effectuées, et donc que l'écriture ne va pas créer une incohérence.

Une conséquence est que l'automate qui implémente en pratique les 4 états MESI a un zillon d'états intermédiaires. En 2017 j'en ai vu un dans une thèse, il prenait plusieurs pages et les patates étaient toutes petites.

## 13.2 Mémoire virtuelle

Dans cette section, on va voir une des vraies différences entre un microcontrôleur (comme le MSP430) et un ordinateur. Le second est conçu pour pouvoir gérer plusieurs processus, et offre un certains nombre de mécanismes matériels pour cela. La mémoire virtuelle est l'un de ces mécanismes.

Le principe est simple : on veut partager la mémoire entre plusieurs processus, mais on veut que chacun ait l'impression qu'il est tout seul et a toute la mémoire pour lui. Donc en particulier, en termes de jeu d'instruction, tout l'espace d'adressage. Seul l'OS sait qu'il y a plusieurs processus et arbitre la "vraie" mémoire entre eux. La "vraie" mémoire peut du reste être partagée entre plusieurs processus (code des bibliothèques par exemple) – tant qu'on ne fait que la lire c'est facile.

Et donc les adresses manipulées par les instructions du jeu d'instruction sont des adresses *virtuelles*, qui ne correspondent pas du tout aux adresses *physiques*. La traduction de l'une en l'autre est réalisée automatiquement (en matériel), et est fonction du processus.

### 13.2.1 Vue générale

Le processeur possède un registre "numéro du processus en cours" ou PID qui contient le numéro du processus courant, mettons sur 16 bits. Ce registre est mis à jour à chaque changement de processus (on dit aussi *changement de contexte*).

L'adresse virtuelle (sur 32 bits) générée par une instruction est d'abord étendue par le PID. On obtient une adresse virtuelle unique de 48 bits. Cette adresse virtuelle unique est ensuite traduite en une adresse physique (de 28 bits si vous êtes pauvre, de 36 si vous êtes très riche) qui est envoyée sur le bus physique de la mémoire.

On voit que par ce mécanisme, on peut avoir deux processus qui exécutent un même code, produisant les mêmes adresses absolues, sans se marcher sur les pieds. Par exemple, il est typique d'organiser la mémoire ainsi : le code en bas, suivi par le tas qui croît en montant, et la pile descendant de l'adresse maximale  $2^{32} - 1$ . Tous les processus auront leur pile partant de l'adresse (virtuelle) maximale, mais ce ne sera pas la même pile.

A ce point un petit dessin à deux processus s'impose. A gauche, les espaces d'adresses virtuelles. A droite, l'espace d'adressage physique.

Remarque : si vous avez un processeur 64 bits (genre AMD64), cela signifie que ses adresses virtuelles sont de 64 bits, et donc que chaque processus peut adresser jusqu'à  $2^{64}$  octets. Toutefois, par le même mécanisme il pourra vivre dans une mémoire physique plus petite, et tout de même avoir la pile qui descend de l'adresse  $2^{64} - 1$ .

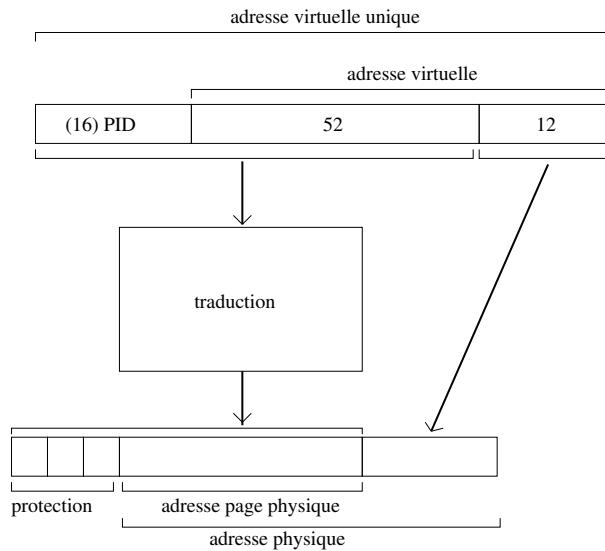


FIGURE 13.4 – Vue d'ensemble de l'adressage à travers la mémoire virtuelle pour des adresses virtuelles de 64 bits

### 13.2.2 Avantages de la mémoire virtuelle

Puisque nous allons étendre les adresses virtuelles avec le PID, on peut en profiter pour établir un système d'allocation et de protection de la mémoire. Et puisque les adresses manipulées par le programme ne sont plus des adresses physiques, on peut envisager de modifier l'emplacement des adresses physiques pour une adresse virtuelle donnée suivant les besoins.

Voici un résumé de ce que la mémoire virtuelle va permettre :

1. Donner à chaque processus l'impression qu'il est tout seul dans l'ordinateur.
  - impossibilité mécanique pour un processus d'écrire dans la mémoire des autres processus (isolation mémoire)
  - possibilité pour chaque processus d'allouer toute la mémoire physique (pas tous en même temps bien sûr).
2. Proposer des mécanismes de protection de la mémoire plus fins que l'isolation, par exemple :
  - un bit associé à chaque adresse dit si elle est accessible
  - un bit dit si elle est accessible en lecture seule, ou accessible en lecture/écriture.
  - un bit dit si l'adresse peut contenir du code exécutable, ou pas. Lorsque le processeur tente d'exécuter une instruction venant d'une adresse non exécutable, une exception est levée et le système tue le processus sans autre forme de procès. Ainsi, il est impossible d'exécuter directement du code malveillant caché dans une image par exemple.
  - etc.
3. Donner l'illusion qu'il y a plus de mémoire que la mémoire physique réellement disponible
  - les adresses (virtuelles) qui ne sont pas actuellement activement utilisées par l'ordinateur ne seront pas traduite en mémoire physique, mais en mémoire disque. La partie du disque correspondant s'appelle alors *mémoire d'échange* ou *swap* en swahili.
  - L'OS remontera ces adresses en mémoire physique automatiquement quand le besoin s'en fera sentir.
  - rappel de la LCDE : le Mo sur le disque est bien moins cher, mais bien plus lent que le Mo en mémoire vive.
4. Faciliter l'allocation mémoire : une plage d'adresses virtuelles consécutives peut être traduite en plusieurs pages de mémoires physiques placées n'importe où en mémoire. Deux exemples :
  - Sur un processeur avec des adresses virtuelles de 64 bits, malloc devient trivial :  $2^{64}$  est tellement énorme qu'on n'a pas à se soucier de la fragmentation mémoire, on trouve toujours une plage virtuelle assez grande. Ensuite on associe à chaque page virtuelle une page physique au hasard, et on est bon.

- A l'initialisation de chaque processus, on alloue une page pour la pile, pour un SP initialisé à  $2^{64} - 1$ . En général cela suffit. Mais parfois, un processus fait assez de récursion pour que tout à coup un empilement provoque un défaut de traduction. Cette exception est reçue par le système d'exploitation, qui teste si par hasard l'adresse fautive ne correspond pas à SP. Si c'est le cas, au lieu de faire un segfault, le système alloue une page de plus pour la pile, puis relance le processus interrompu à partir de l'instruction fautive : elle va désormais marcher.

L'idée d'un registre de PID n'est qu'une idée parmi d'autres. Par exemple, le powerPC n'a pas de registre de PID, mais a à la place 16 "registres de segments" qui font l'expansion des 4 bits de poids forts de l'adresse virtuelle en une adresse de page virtuelle sur 24 bits. Ce sont ces 16 registres qui sont changés par l'OS à chaque changement de processus. Lorsque certains de ces registres contiennent des valeurs identiques pour deux processus, la mémoire correspondante est partagée par les deux processus.

### 13.2.3 Aspects architecturaux

La traduction d'une adresse virtuelle en adresse physique peut se faire en principe par une lecture de table. Deux problèmes pratiques, que nous exposons dans le cas d'adresses étendues par le PID :

1. La table de traduction a  $2^{48}$  entrées, donc est plus grosse que la mémoire physique.
2. Chaque accès mémoire se traduit désormais par deux accès mémoire...

La solution au premier problème va être de faire cette traduction par *pages mémoire*, une page étant une unité de mémoire intermédiaire, typiquement 4 Ko. Ainsi, le coût de la traduction (tant en mémoire qu'en temps) sera amorti sur toutes les adresses d'une page.

La solution au second problème est d'utiliser un mécanisme de cache : les traductions récentes sont conservées dans une petite mémoire associative, appelée TLB pour *translation look-aside buffer*. En pratique, cette TLB réalise la traduction d'adresse très rapidement la plupart du temps. Lorsqu'elle échoue, le système ou le matériel doit réaliser la traduction pour de bon, ce qui peut impliquer plusieurs accès mémoire, mais c'est très rare : moins d'une adresse sur 4000 pour une page de 4Ko, espère-t-on. Rappelons que le principe de localité s'applique toujours : les adresses d'une page ont toutes les chances d'être réutilisées de nombreuses fois, ce qui amortit encore plus le coût.

### 13.2.4 Dans le détail : table des pages (\*)

Tables des pages directe, organisée hiérarchiquement (Fig.2 de l'article *Virtual memory, issues of implementation* du magazine *Computer*).

Accès top-down ou bottom-up.

Pb : adresses virtuelles de 64 bits ?

Solution : table des pages inversées (Fig. 5).

### 13.2.5 Cache d'adresses virtuelles ou cache d'adresses physiques ? (\*)

Au fait, de quelles adresses on parle ?

Le principe de localité fonctionne aussi bien en adresses virtuelles qu'en adresses physiques, dès lors que la traduction se fait par pages plus grosses que la ligne de cache. Si l'on choisit un cache en adresses virtuelles, il faut naturellement qu'il utilise les adresses virtuelles étendues par le PID, sans quoi l'isolation des processus n'est plus garantie. Quoiqu'une alternative est que l'OS marque tout le cache comme invalide à chaque changement de processus.

Il est plus naturel d'avoir un cache en adresses physiques. L'inconvénient est qu'il faut alors réaliser la traduction virtuelle-physique, puis l'accès au cache en séquence.

Avec un cache en adresses virtuelles, on économise cette traduction en cas de hit, ce qui permet de retourner la donnée plus vite. L'inconvénient est que le code de `printf`, et en général toute donnée partagée, va se retrouver en plusieurs exemplaires dans le cache. On a du gâchis de cette mémoire coûteuse.

Un bon compromis est sans doute de réaliser le cache de niveau 1 (le plus près du processeur, et celui qui doit être le plus rapide) en adresses virtuelles, et les niveaux suivants en adresses physiques.

On va construire complètement la version d'une hiérarchie mémoire qui a besoin du moins de matériel. Puis on verra ce qui existe dans les "vrais" processeurs grâce à un article paru dans *IEEE Computer*.

## 13.3 Une mémoire virtuelle + cache minimale

- Un registre de PID
- Un cache de niveau 1 séparé, indexé virtuellement, et tant pis pour les doublons.
- En cas de *hit*, l'accès mémoire ne pose pas de pb.
- Un défaut de cache génère une interruption. Le système consulte ses tables, fait la traduction de la ligne virtuelle en ligne physique, vérifie les droits, choisit une ligne de cache à virer, et copie la ligne physique dans le cache. Le tout idéalement en matériel.
- Les niveaux suivants de la hiérarchie sont indexés physiquement.

### 13.3.1 Instructions spécifiques à la gestion mémoire

Et voici quelques instructions à ajouter à notre processeur. Sauf la première, ce sont des instructions accessibles uniquement en mode superviseur du processeur.

- prefetch (pour la performance) : cette instruction demande à ce qu'une adresse soit présente dans le cache.
- lecture et écriture directement en mémoire physique, pour y mettre par exemple la page des tables
- invalidation d'une ligne de cache ou d'une page
- éventuellement écriture dans les tables matérielles si elles existent
- etc...



# Chapitre 14

## Conclusion : vers le système d'exploitation (\*)

L'OS a deux rôles fondamentaux (et largement interdépendants) :

- *virtualiser* le matériel : donner à l'utilisateur l'illusion qu'il est tout seul sur un ordinateur ; donner au programme l'illusion qu'il s'exécute sur un matériel "mieux" (plus simple, moins limité) que le vrai.
- gérer les ressources (les différents niveaux de mémoire, le temps de calcul, les processeurs, les entrées/sorties,...) pour les partager au mieux entre plusieurs utilisateurs/applications.

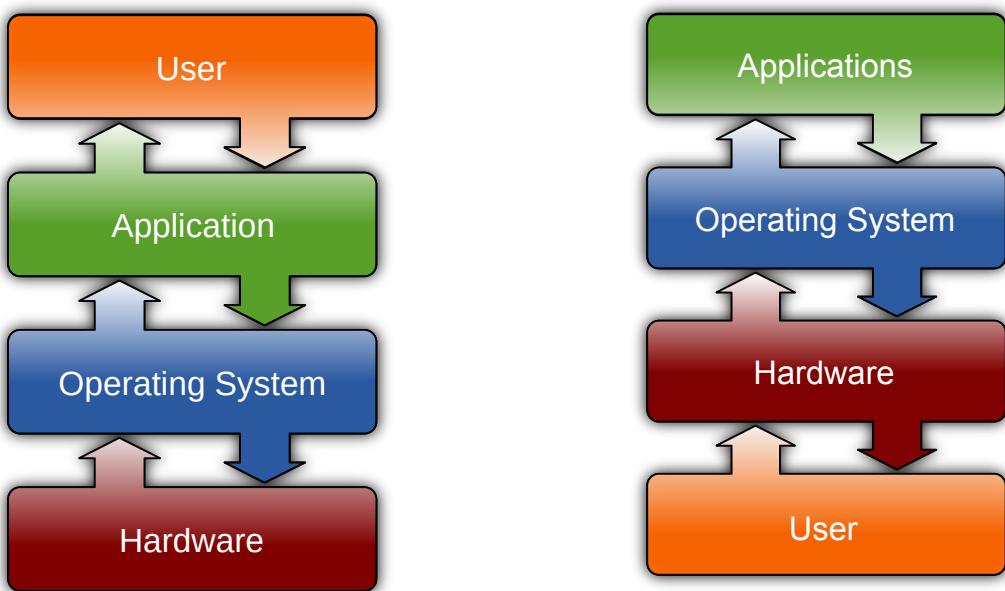


FIGURE 14.1 – La place de l'OS selon Wikipedia (à gauche) et G. Salagnac (à droite). Du point de vue de l'architecte, c'est celle de droite qui correspond le mieux à la réalité.

### 14.1 Le rôle de l'OS (du point de vue d'un prof d'archi)

Du plus fondamental au plus gadgetesque :

- Définir (ou tout au moins publier en fournissant une interface) des règles du jeu permettant à un programme de fonctionner sur une machine donnée. Par exemple, comment demander (allouer) de la mémoire, comment les procédures se passent des paramètres, où se trouve la pile, comment afficher un caractère à l'écran, comment créer un fichier disque, comment lancer un programme...
- Assurer les tâches de maintenance de routine : compacter la mémoire lorsqu'elle est libérée, envoyer les bonnes couleurs sur l'écran et les bons échantillons sonores dans le tampon de la carte son au bon moment,...
- Démarrer les sous-systèmes dans l'ordre.

- gérer les *processus*, et on va commencer par voir cela.
- Fournir différentes abstractions du matériel et du logiciel, certaines dégénérant franchement en allégories. Exemples : de l'arborescence du système de fichiers jusqu'au bureau avec des fenêtres et du *drag and drop* entre elles.

## 14.2 Multiutilisateur ou multitâche, en tout cas multiprocessus

On dit processus, ou tâche.

Les processus doivent se partager

- la mémoire de l'ordinateur (l'espace) comme vu au chapitre 13 ;
- le temps ;
- d'autres ressources, notamment des entrées/sorties, et depuis peu les ressources d'exécution : les cœurs de vos processeurs multicœurs de gamerz.

Le rôle de l'OS est d'assurer ce partage sans disputes et dans la bonne humeur.

### 14.2.1 Partage du temps

Pour partager le temps, on le découpe en tranches de quelques milliers de cycles, et chaque processus exécute une tranche de calcul à son tour.

Entre deux processus, le processeur doit *changer de contexte*. Un contexte c'est au moins l'état de tous les registres du processeur. Il sera stocké par le système en mémoire en attendant le prochain tour.

La tranche de temps doit être

- assez courte pour que l'utilisateur, qui fonctionne à 100Hz, ait l'impression que les processus s'exécutent en parallèle,
- mais assez longue pour que le temps du changement de contexte reste négligeable devant le temps passé à faire le boulot.

Un processeur de 1GHz qui veut changer de processus à 100Hz peut réaliser 10 millions de cycles de chaque processus entre deux changements de contextes... c'est beaucoup.

Le partage du temps peut être coopératif (chaque processus est responsable de rendre la main au système après le temps qu'il juge raisonnable) ou préemptif (le système alloue les tranches de temps et interrompt les autres processus). Les premiers OS multitâches grand public étaient souvent coopératifs :

- c'est plus simple pour l'auteur du système (c'est essentiellement aux auteurs des applications de gérer le multitâche)
- à une époque où les ordinateurs étaient moins rapide, cela permettait relativement facilement de rendre le système réactif aux interactions avec l'utilisateur : par exemple, lorsque celui-ci déplace une fenêtre, le processus qui s'occupe de cela va décider de garder la main jusqu'à ce que l'utilisateur lache la souris. Ainsi, 100% du temps CPU est consacré à assurer la fluidité du déplacement de la fenêtre.
- enfin cela n'a besoin d'aucun support matériel particulier.

Le gros inconvénient du multitâche coopératif est qu'un plantage d'un processus peut planter tout le système. De nos jours, on n'a plus que du multitâche préemptif.

Celui-ci a besoin d'un support matériel minimal, sous forme au moins d'une interruption, typiquement déclenchée par une horloge extérieure au processeur tous les 100ème de seconde, qui interrompt le processus courant et saute à la routine système qui va endormir ce processus (sauvegarder son contexte) et réveiller le processus dont c'est le tour.

Dans les détails, l'OS maintient une liste chaînée circulaire de contextes, un par processus (un contexte étant une structure de donnée contenant l'état complet du processeur : la valeur de tous ses registres, et une adresse de retour). Lorsqu'un processus est interrompu (au bout de sa tranche de 10ms), la routine de traitement d'interruption sauvegarde ce contexte, puis le remplace par le contexte du processus suivant dans la liste, y compris l'adresse de retour. Tant et si bien que le RETI (return from interruptions) qui termine cette routine de traitement d'interruption ne reprend pas l'exécution du processus qui vient d'être interrompu, mais celle d'un processus qui avait été interrompu il y a longtemps.

Avec de la mémoire virtuelle, le contexte inclut aussi la table de traduction d'adresses du processus (ou un pointeur dessus). Au changement de processus, pour assurer l'isolation, il faut vider la TLB (le cache des traductions d'adresse) – le processeur doit fournir une instruction (en mode privilégié) pour cela. Cela veut dire que les premiers accès mémoire du processus réveillé vont produire des page fault. Si les caches d'instruction et données sont en adresses virtuelles il faut les purger aussi. Bref, le gros du coût du changement de processus va être lié à la mémoire virtuelle.

### 14.2.2 Partage des entrées/sorties

La plupart des processus ne font rien la plupart du temps : ils attendent que l'utilisateur bouge sa souris, son clavier, ou qu'un paquet arrive sur le réseau...

Pour implémenter une telle attente, on peut écrire une boucle infinie qui teste l'arrivée d'un événement. C'est ce que vous ferez sur votre processeur. Toutefois, cette approche, dite *attente active*, est un gros gaspillage de ressources de calcul : si par exemple le système a un gros calcul à réaliser en tâche de fond (par exemple une impression, qui consiste à convertir un document en une image à la résolution, énorme, de l'imprimante), et que personne n'appuie sur une touche du traitement de texte, on préfère donner plus de temps de calcul à l'impression, et moins à l'attente de la touche.

La bonne approche est d'endormir les processus qui attendent un événement (c'est à dire de ne plus leur donner de tranche de temps du tout – en pratique les contextes correspondant sont sortis de la liste chaînée circulaire évoquée ci-dessous, et rangés dans un placard des processus endormis). À l'arrivée de l'événement attendu, le système réveille le ou les processus qui attendent cet événement en les réinsérant dans la liste circulaire.

Autrement dit, le résultat principal de l'appel à `scanf` ou `getc` est d'endormir votre processus.

Ainsi, l'attente active est remplacée par l'attente passive, et c'est ainsi que votre *CPU load* est à 0 quand votre PC ne fait rien, ou que l'impression d'un document est capable d'accaparer 99% du temps CPU, le 1% restant étant ce dont votre PC a besoin pour gérer un étudiant tapant du texte le plus vite possible de ses petites mains à 3Hz.

En fait, même les processus systèmes responsables des entrées/sorties (ceux qui vont réveiller vos processus à vous) sont eux-même en attente passive. Toutefois, eux sont réveillés par des interruptions matérielles : appui sur une touche, la carte son réclame des échantillons à jouer, *timer* à une fréquence de l'ordre de la miliseconde, etc.

Certains processeurs peuvent même arrêter complètement leur horloge jusqu'à la prochaine interruption matérielle. Je ne sais pas ce qui se passe dans votre PC.

### 14.2.3 Partage des ressources d'exécution

On a à résoudre un problème d'équilibrage des charges, ou *load balancing* en australien. Si on vous demande de le faire, vous implémenterez un algorithme glouton qui fera bien l'affaire : quand arrive un nouveau processus, on le refile au processeur qui a l'air le moins chargé.

Si par la suite tous les processus du cœur 0 se terminent ou s'endorment alors que le processeur 1 est toujours à 100%, on peut envisager de redéployer certains des processus du second sur le premier. Un tel rééquilibrage de charge a un coût très supérieur à un changement de contexte, donc il faut le faire avec discernement. La difficulté est de prévoir l'avenir : quelle est la probabilité que tel processus endormi se réveille, et pour combien de temps ? L'OS peut maintenir des statistiques par processus.



## **Troisième partie**

### **Annexes**



## Annexe A

# Rudiments de complexité

Il est utile, en architecture comme en programmation, de raisonner à la louche sur le coût (en temps, en mémoire, en surface, ...) d'une implémentation. Mais ce coût dépend en général de la taille du problème. Nous donnons ici des éléments qui permettent, toujours à la louche, d'évaluer comment un coût croît avec la taille du problème. Ce n'est pas toujours juste proportionnel !

Dans toute la suite,  $n$  décrira la taille du problème.

Les fonctions utiles à ce cours sont les fonctions suivantes. Quand  $n$  est grand,

$$1 \ll \log_2 n \ll n \ll n^2 \ll n^3 \ll \dots \ll 2^n \ll n^n < n! .$$

Là dedans, vous connaissez et visualisez bien  $n$  et  $n^2$ . Il est important de bien se familiariser avec  $2^n$  et  $\log_2 n$ . Non pas comme des objets abstraits, mais comme des objets qui arrivent naturellement en architecture.

On va voir que  $2^n$ , c'est déjà très cher. Par conséquent, les fonctions plus grande comme  $n^n$  et  $n!$  seront de peu d'intérêt pratique.

### A.1 Les fonctions $2^n$ et $\log_2 n$

La fonction  $2^n$  apparaît dès le chapitre du codage. Dans ce cas, il suffit de comprendre que le 2 de  $2^n$  est le 2 de binaire.

- Sur  $n$  bits on peut coder  $2^n$  informations différentes.
- Autrement dit, un mot de  $n$  bits peut prendre  $2^n$  valeurs différentes.
- La table de vérité d'une fonction combinatoire à  $n$  entrées binaires a  $2^n$  lignes.

Sa propriété fondamentale est la suivante :

$$2^{a+b} = 2^a \times 2^b.$$

Mais en fait on utilise surtout la variante suivante (pour  $b = 1$ ) :

$$2^{a+1} = 2 \times 2^a.$$

La fonction  $2^n$  apparaît donc naturellement chaque fois qu'on a une construction algorithmique du type : *pour obtenir truc( $n + 1$ ), j'assemble deux truc( $n$ )*. En effet, le coût de  $\text{truc}(n + 1)$  sera le double du coût de  $\text{truc}(n)$ . Cela suggère que ce coût sera en  $2^n$ .

Une autre propriété utile, car elle sous-tend toute l'arithmétique binaire :

$$2^n - 1 = \sum_{i=0}^{n-1} 2^i.$$

Vérifiez pour les petits  $n$  :

$$1 + 2 = 3 = 4 - 1$$

$$1 + 2 + 4 = 7 = 8 - 1$$

$$1 + 2 + 4 + 8 = 15 = 16 - 1.$$

En fait c'est aussi une propriété que vous manipulez, dans sa version décimale, depuis que vous savez compter : c'est  $999 + 1 = 1000$ . Vous avez donc déjà rencontré cette dernière propriété sous la forme suivante : en binaire,  $10000000 = 0111111 + 1$ . Réfléchissez une minute pour vous convaincre que c'est la même chose. Sinon, la preuve est par récurrence en utilisant  $2^{a+1} = 2 \times 2^a$ .

Enfin, la fonction  $2^n$  croît très vite. Cela se voit dans les deux propriétés ci-dessus.

$2^0 = 1$ ,  $2^1 = 2$ ,  $2^{16} = 65536$ ,  $2^{32} \approx 4.10^9$  (4Giga),  $2^{64} \approx 16.10^{18}$ . Peut-être plus parlant :  $2^{64} \approx 16\ 000\ 000\ 000\ 000\ 000$ .

Quant à la fonction  $\log_2 n$ , c'est l'inverse de  $2^n$  :

$$x = 2^y \iff y = \log_2 x .$$

On peut donc reformuler à l'envers toutes les phrases dans lesquelles on avait un  $2^n$ , par exemple : *pour coder n informations différentes, il faut  $\log_2 n$  bits*.

Attention, si  $n$  n'est pas une puissance de 2,  $\log_2 n$  ne sera pas entier. Il faut l'arrondir à l'entier supérieur. La phrase correcte serait *pour coder n informations différentes, il faut  $\lceil \log_2 n \rceil$  bits*. Mais ce sera souvent implicite, surtout dans les raisonnements à la louche.

## A.2 Raisonner à la louche

Quand on évalue les coûts, on ajoute souvent les coûts de plusieurs composants. Mais quand on raisonne à la louche, on ne s'intéresse qu'au plus important. Les autres coûts sont rassemblés dans une formulation de type "et des chouillas".

Tout cela peut s'appuyer sur des mathématiques propres. Si vous voulez creuser, allez wikipédier *comparaison asymptotique* et *complexité asymptotique*.

Par exemple, quand pour obtenir  $truc(n+1)$ , j'assemble deux  $truc(n)$ , j'ai souvent en fait besoin d'un chouilla en plus. Pour obtenir un multiplexeur pour des adresses  $n$  bits, on assemble deux multiplexeurs pour des adresses de  $n-1$  bits, et un multiplexeur de taille 1. Ce dernier est un chouilla. On obtient quand même un coût total en  $2^n$ .

Quand on a un doute on peut toujours faire une vraie récurrence où l'on voit apparaître une des formules précédentes. On peut partir d'un schéma pour un petit  $n$ . Par exemple, la figure 3.10, p. 35, fait apparaître la récurrence  $c_{n+1} = 2c_n + 1$ , avec  $c_1 = 1$  (ici  $a$  est le nombre de bits d'adresse). Calculez les premiers termes (1, 3, 7, 15, 31) et intuituez-en que  $c_n = 2^n - 1$ , ce que vous démontrerez ensuite en utilisant la propriété déjà vue  $2^{a+1} = 2 \times 2^a$ .

En général, si on a une somme de deux fonctions de la liste ci-dessus, à la louche on peut négliger la plus petite. Je dirai donc que le coût de la gare de triage de la p. 35 est en  $2^n$ . Quel est le coût en fonction du nombre de cases adressées  $k$ ? On voit qu'il faut  $1 + 2 + 4 + 8 + \dots + k/2$  aiguillages, donc  $k-1$ .

Pour des raisons similaires, on ne s'intéresse pas aux constantes qui peuvent multiplier ces fonctions. Si je compare une solution qui coûte  $2^n$  et une solution qui coûte  $n^2$ , je ne suis pas à un facteur 3 près sur ces coûts : le  $n^2$  peut même être un  $17n^2$ , dès que  $n$  sera assez grand ce sera moins cher que  $2^n$ .

Si les constantes sont importantes pour vous, c'est que vous ne raisonnez plus à la louche — je ne dis pas que c'est mal.

Pour conclure, les fonctions listées ci-dessus ont également les propriétés à la louche suivantes :

- Si le coût est en  $\log_2 n$ , il faut que le problème double de taille pour que le coût commence à augmenter. C'est donc pas cher. C'est bien.
- Si le coût est en  $n$ , il est proportionnel à la taille du problème.
- Si le coût est en  $2^n$ , l'augmentation d'une unité de la taille du problème fait doubler le coût. On n'ira pas très loin.

## Annexe B

# VHDL, un langage pour la synthèse de circuits

### B.1 Flot de synthèse

La figure B.1 est coupée de la première page du manuel d'un outil qui sert à concevoir des circuits (en l'occurrence l'outil ISE, qui sert à programmer les FPGA Xilinx).

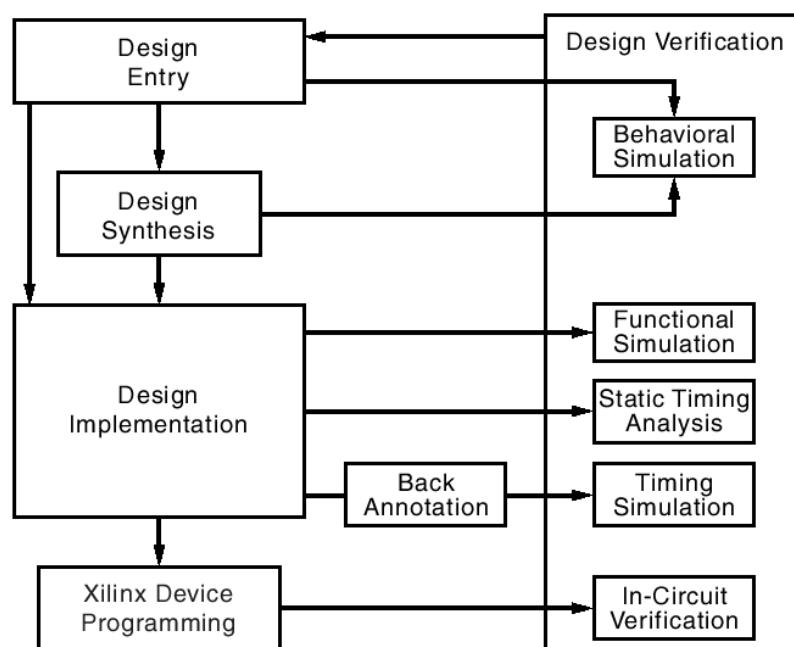


FIGURE B.1 – Le flot de conception pour implémenter un circuit sur un FPGA (vue très simplifiée)

La boîte *Design entry* recouvre différentes techniques de description de circuit. On y trouve des interfaces qui ressemblent à LogiSim, et d'autres interfaces graphiques, par exemple pour dessiner un automate sous forme de patates et de transitions.

Toutefois, la majorité des circuits sont décrits au moyen de deux langages standard et à peu près équivalents : VHDL et Verilog.

Par pure méchanceté, nous allons travailler avec le plus compliqué des deux, car d'une part c'est le plus répandu en Europe, et d'autre part il est plus facile de passer de VHDL à Verilog que le contraire. Il y a aussi des raisons purement techniques, comme un typage plus strict qui protège les débutants que vous êtes. La plupart des concepts que nous allons voir existent dans les deux langages, avec des syntaxes différentes.

## B.2 Décrire un circuit par un langage

### B.2.1 Description de circuit $\neq$ programmation

Nous allons présenter VHDL essentiellement par opposition aux langages de programmation que vous connaissez. En effet, VHDL n'est pas un langage de programmation. C'est un langage de description de circuit intégré (*Very large scale integration Hardware Description Language*).

La différence fondamentale est la suivante. Dans un programme classique, on avance ligne par ligne en exécutant le cycle de von Neumann. Une ligne de votre programme va sans doute correspondre à de nombreux cycles, mais au final la sémantique<sup>1</sup> du langage décrit une exécution séquentielle des différents calculs et accès mémoire.

Par contre, dans un circuit, tout bouge potentiellement en même temps. La sémantique d'un langage de description de circuit est donc naturellement parallèle. Chaque porte vit sa vie en parallèle des autres portes. S'il y a une séquentialité des événements dans un circuit, elle est uniquement causée par la propagation de l'information le long des fils et à travers les portes, pas du tout par l'ordre dans lequel sont décrits les différents composants en VHDL.

Relisez les deux paragraphes précédents avant de passer à la suite.

### B.2.2 Mais au fait, quelle est la sémantique d'un circuit numérique ?

Il est difficile de comprendre VHDL sans se poser cette question. Considérons un circuit dessiné dans LogiSim. La sémantique de ce circuit, c'est ce qu'il fait, son comportement. Comment formaliser ce comportement ?

La meilleure réponse à ce jour est ce qu'on appelle une sémantique à événements discrets. Le comportement d'un circuit, c'est la manière dont il réagit à des événements (des transitions de 0 à 1 ou de 1 à 0) sur les entrées.

On peut simuler un circuit avec ce qu'on appelle un simulateur à événements discrets (toutes les boîtes de droite de la figure B.1 avec le mot *simulation* en sont). Un simulateur à événement discret fonctionne de la manière suivante.

- Il utilise un graphe du circuit, composé de portes reliées par des fils. Ce graphe ne bouge pas pendant la simulation (un circuit ne se modifie plus une fois qu'il est assemblé).
- Il maintient liste d'événements triés par date. Un événement est un triplet  $(t, s, v)$  où
  - $t$  est une date (un instant),
  - $s$  est un signal (un fil du graphe du circuit), et
  - $v$  est une valeur de ce fil.

On n'a un événement que lorsque cette valeur change, par exemple de 0 à 1 ou de 1 à 0 (il peut y avoir d'autres valeurs possibles comme *U* pour *Unknown* ou *Z* pour *haute impédance*).

- Le simulateur réalise une seule boucle que voici :

Tant que la liste d'événements n'est pas vide,

- Avancer le temps au prochain événement, c'est-à-dire retirer l'événement  $(t, s, v)$  en tête de liste;
- Regarder dans le graphe à quelles entrées de portes le signal  $s$  est connecté. Propager le changement de valeur à travers ces portes. Cela génère de nouveaux événements (un peu plus tard) : les insérer à leur place dans la liste triée par  $t$ .
- Et recommencer.

Cette sémantique à événements discrets est très sympa puisqu'on la décrit avec une simple boucle séquentielle : elle est facile à programmer, mais aussi facile à appréhender pour nos pauvres cerveaux séquentiels.

Il faut comprendre que certains aspects du langage VHDL sont sous-tendus par cette sémantique.

### B.2.3 Description comportementale ou description structurelle

Dans tout ce cours on a dessiné nos architectures comme des boîtes qu'on assemble par des fils. Il est donc naturel que les outils comme Logisim offrent une interface à base de boîtes et de fils. C'est aussi le cas de VHDL, qui appelle une boîte une *entity* et un fil un *signal*. L'entité est un peu comme la déclaration d'un prototype de fonction C dans un fichier *.h*.

Il faut aussi décrire l'intérieur de la boîte noire. VHDL appelle cela une architecture pour une entité.

On peut donc décrire un gros circuit par l'assemblage de boîtes plus petites. Du LogiSim, quoi. Ceci s'appelle une description structurelle.

1. La sémantique d'un programme, c'est son sens (au sens de "au sens de"), sa signification, ce qu'il fait. De préférence décrit en termes formels et mathématiques.

Mais VHDL permet aussi de décrire un composant par son comportement. L'idée est que c'est plus abstrait, plus près de la sémantique. Et qu'on arrivera bien à compiler automatiquement une telle description en un assemblage de porte. C'est le discours que tenait le prof d'archi au sujet des fonctions booléennes et des automates, n'est-ce pas ? Cela s'appelle une description comportementale.

Les deux ne sont pas antinomiques. En gros, on aime décrire les gros circuits par des boîtes composées de boîtes plus petites, et décrire l'intérieur des plus petites boîtes par une description comportementale. Mais là aussi, c'est l'art de l'ingénieur de trouver le bon équilibre.

Pour décrire un comportement arbitraire, on dispose d'un langage de programmation presque classique, avec variables, conditionnelles, etc. Ce langage est destiné à s'exécuter dans le simulateur. Il faut bien comprendre qu'il ne sera pas transposé directement en un circuit.

## B.2.4 VHDL synthétisable

On arrive à simuler tout programme VHDL. Par contre, on n'arrive pas à transformer en un circuit efficace tout programme VHDL. Tout ce qui est structurel se transpose directement en un circuit. Par contre, les descriptions comportementales sont parfois synthétisables, et parfois non. Et cela dépend des outils utilisés. Souvent, les outils vous fournissent des modèles (*templates*) pour des fonctionnalités qu'ils savent synthétiser. Le plus simple est de partir de cela.

## B.3 Le langage VHDL par l'exemple

Assez de blabla. Observons avec attention la figure B.2.

### B.3.1 Entités, architectures et composants

Une entité c'est une boîte noire. Quand on la déclare on l'appelle `entity`, et il faut juste définir ses `ports` : ses entrées / sorties. Quand on l'utilise, on l'appelle un `component` mais c'est la même chose.

Ensuite on peut définir plusieurs *architectures* pour une entité. L'idée est qu'on commencera par une architecture comportementale (qui définit, dans un langage pas forcément synthétisable, "ce que fait" la boîte noire). Puis quand on sera satisfait du comportement on pourra (plus ou moins progressivement) remplacer cette architecture par une architecture plus *structurelle* : celle-ci sera synthétisable, mais son comportement ne sera plus explicite. On aimerait savoir faire ce raffinement de comportemental vers structurel automatiquement, mais ce n'est pas encore cela.

En attendant, VHDL fournit des constructions permettant d'associer une architecture à chaque entité. Nous, bien sûr, on va définir une seule architecture par entité, cela suffira bien. C'est le cas de la figure B.2.

### B.3.2 Variables, paramètres, signaux et ports

Variables et paramètres sont des notions qui ressemblent à celles des langages de programmation classiques, alors que signaux et ports sont... des signaux. Des fils, quoi.

L'assignation des variables se fait par `:=`, celle des signaux par `<=` qui se lit en gros "se branche sur". Dans `a<=expr`, l'information portée par les signaux se propage de droite à gauche.

Quand-est-ce qu'on utilise une variable ou un paramètre ? Par exemple, le `n` de "additionneur `n` bit" est un paramètre de l'architecture. On aura peut-être envie de déclarer une variable `p` qui vaudra `p:=n+1` ;

Et donc les ports de l'additionneur `n` bits sont des vecteurs de `n` bits. On voit qu'on ne peut pas construire physiquement cet additionneur tant qu'on n'a pas fixé les valeurs de ses paramètres génériques.

Admirez sur la figure B.2 comment on branche un sous-ensemble d'un signal sur un autre au moyen du mot-clé `downto`.

L'autre grosse utilisation des paramètres est pour la spécification du temps, par exemple l'entité suivante définit une infinité de portes avec des délais différents :

```
entity nand is
  generic(delay:time)
  port(a,b:in std_logic; y:out std_logic)
end nand ;
architecture idiote of instr_reg is
begin
  y <= not(a and b) after delay;
end architecture;
```

```

library ieee;
use ieee.std_logic_1164.all;

entity instr_reg is
  port ( clk : in  std_logic;
         ce  : in  std_logic;
         rst : in  std_logic;

         instr : in  std_logic_vector(15 downto 0);
         cond  : out std_logic_vector(3 downto 0);
         op    : out std_logic_vector(3 downto 0);
         updtt : out std_logic;
         imm   : out std_logic;
         val   : out std_logic_vector(5 downto 0) );
end entity;

architecture arch of instr_reg is
  signal instr_r : std_logic_vector(15 downto 0);
begin

  process(clk, rst)
  begin
    if rst = '1' then
      instr_r <= X"0000";
    elsif clk'event and clk = '1' and ce = '1' then
      instr_r <= instr;
    end if;
  end process;

  cond <= instr_r(15 downto 12);
  op   <= instr_r(11 downto 8);
  updtt <= instr_r(7);
  imm   <= instr_r(6);
  val   <= instr_r(5 downto 0);

end architecture;

```

FIGURE B.2 – Un composant du processeur de la figure 9.2

Il faudra l'instancier avec une valeur de `delay`, par exemple `delay:=3ns`;

Comprenez au passage la différence entre cette affectation et `y<=expression;`. Dans `y<=expression`, `y` est un signal et le `<=` signifie "branché sur". Plus précisément, il signifie "propager les événements qui viennent de `expression` vers `y`". Il y a donc potentiellement une infinité de changements de valeur de `y` dans cette ligne.

Par contre, une affectation de variable ou de générique comme `delay=3ns` est effectuée une seule fois, à moins qu'elle ne soit explicitement dans une boucle.

### B.3.3 Instances

Je rappelle qu'un composant c'est une entité. Et donc une instance d'un composant, c'est une copie dont on fixe les paramètres et dont on branche les ports. On peut avoir plein d'instances du même composant, par exemple si tout votre circuit est réalisé à base de porte non-et, vous vous retrouverez au final avec quinze millions d'instances du composant non-et.

Une instantiation définit un nom d'instance (avant le `:`). Le nom de l'entité vient après le `:`. Ensuite viennent les assignations des ports et des paramètres, sous la forme `formel => effectif`. Le `=>` se lit comme "est branché sur" : il n'y a pas de notion de sens d'information. Ce n'est pas le "contraire" de `<!=`. Les événements peuvent se propager dans un sens ou dans l'autre à travers `=>`, suivant que le port est une entrée ou une sortie. À mon avis c'est une syntaxe à la con.

```

porte_1:nand
  generic map (5 ns)
  port map (entree1, entree2, sortie);

-- on peut aussi utiliser les noms des paramètres/ports de l'entité
porte_2:nand
  generic map (15 ns)
  port map (a=>entree1, b=>entree2, c=>sortie);

```

### B.3.4 Types de base

Comme la plupart des langages, VHDL vient tout nu, et n'est utilisable que si on lui ajoute plein de bibliothèques.

Sur la figure B.2 vous avez la syntaxe pour inclure les bibliothèques de logique normalisées IEEE. Elles définissent le type `std_logic`:

Valeur	Nom	Signification
'U'	Non initialisé	Aucune valeur n'a été affectée au signal
'X'	Inconnu	Impossible de déterminer la valeur du signal
'0'	0 logique	
'1'	1 logique	
'Z'	Haute impédance	Le signal est isolé grâce à un transistor (cf. le bus I <sup>2</sup> C)
'-'	Sans importance	La valeur du signal n'est pas importante (don't care)

et encore quelques autres, ainsi que la matrice qui dit ce qui se passe quand on branche ensemble deux signaux avec des valeurs différentes.

Vous rencontrerez assurément 'X' et peut être 'U' en sortie de vos simulations. Souvent ce sera normal.

Ces bibliothèques définissent aussi les opérateurs de base dessus : `and`, `or`, `not`, `xor`, `nand`, `nor`.

Il y a un type `std_logic_vector` qui est comme son nom l'indique, et dont vous avez des exemples de déclaration et d'utilisation sur la figure B.2.

*Remarque* : Pourquoi une notation en `(n downto m)` et non pas `(m to n)` ? Car la notation `downto` nous permet de représenter les bits de poids forts à gauche, ce qui est beaucoup plus lisible dès qu'on veut manipuler des signaux représentant des nombres.

Enfin on peut ajouter

```

use ieee.std_logic_arith.all;
use ieee.std_logic_unsigned.all;
pour pouvoir décrire un additionneur simplement par un + entre deux std_logic_vector. Il faut
qu'ils aient la même taille et que le résultat ait aussi la même taille.

```

## D'autres types

Il existe d'autres types pour les signaux, mais la plupart d'entre eux sont à éviter car ils n'offrent généralement pas de garantie quand à l'implémentation réelle de ces signaux, sauf à utiliser des constructions tellement classiques que le compilateur les reconnaît comme des circuits combinatoires.

Par exemple la manière usuelle de décrire un multiplexeur est la suivante

```
y <= x0 when a = '0' else x1;
```

C'est bien une construction comportementale. Du point de vue du langage, l'expression `a = '0'` est de type booléen, pas signal. Mais les synthétiseurs reconnaissent un multiplexeur.

Attention. Certains de vos collègues ont fort intelligemment utilisé

```
y <= x0 when a = '0' else x1 when a='1';
```

cela se simule de manière identique, mais les cas `a='U'`, `a='X'`, etc ne sont pas prévus, contrairement à la formule officielle. Cela a suffi pour que le synthétiseur décide de l'implémenter comme un registre exotique, et plus rien ne marchait. Si si.

De même il existe un type `integer`, il permet de représenter des nombres entiers signés sur 32 bits (de `-2 147 483 647` à `2 147 483 647`), comme par exemple les indices d'un signal `std_logic_vector`. Mais si on veut manipuler un signal portant un entier 32 bits, on utilisera des `std_logic_vector`.

## Description des automates

Enfin, les types énumérés se déclarent de la manière suivante :

```
type t is (cst1, cst2, ...);
```

Un signal déclaré du type `t` peut donc prendre comme valeur uniquement une des constantes `cst1`, `cst2`, ... Cela est très pratique pour définir par exemple le signal représentant l'état courant d'un automate. Le choix de l'encodage de ce signal est alors laissé au synthétiseur VHDL, qui fera ce boulot généralement mieux que nous de toute façon.

*Remarque* : En fait, les types `std_logic` et `boolean` sont aussi des types énumérés :

```
type std_logic is ('U', 'X', '0', '1', 'Z', 'W', 'L', 'H', '-');
type boolean is (true, false);
```

### B.3.5 Concurrent statement

Entre le `begin` et le `end architecture` d'une architecture, on a une liste de *concurrent statement* : leur sémantique n'est pas une exécution séquentielle dans l'ordre où ils sont écrits, mais bien une exécution concurrente, ou parallèle. Autrement dit, on peut les écrire dans n'importe quel ordre.

Pour nos besoins, les *concurrent statements* sont essentiellement de 4 types :

- l'assignation de signal, déjà vue,
- l'instanciation de composant, déjà vue,
- le processus (`process`) qu'on ne verra pas, à l'exception suivante près : Le pâté entre `process` et `end process` sur la figure B.2 est la manière officielle de décrire un registre avec un reset. On est prié de faire comme cela.
- le `if generate` et `for generate` qui sont un `if` et un `for` concurrents, ou parallèle. Voyez les comme ceci : le `if generate` va produire du matériel si sa condition est vraie. Le `for generate` va répliquer plein de fois du matériel. Attention, ils englobent un bloc de *concurrent statements* mais leur condition ne porte pas sur des signaux mais sur des variables/paramètres.

La suite décrit en détaille le fonctionnement des processus et des `generate`, parce que Jérémie Detrey l'avait tapé. Toutefois, vous êtes invités très expressément à copier et modifier des constructions de code existant et fonctionnel, et à taper le moins possible de VHDL *from scratch*. La suite doit vous permettre de comprendre ce que vous copiez, mais pas (tout de suite) d'inventer du VHDL... en tout cas si vous voulez qu'il soit synthétisable.

### B.3.6 Processes

Les *processes* sont le "côté obscur" de VHDL, c'est-à-dire le VHDL *comportemental*.

### Principe

Un process se présente de la manière suivante :

```
process (signal1, signal2, ... )
begin
  instr1;
  instr2;
  ...
end process;
```

Les signaux indiqués en début de process sont appelés liste de sensibilité (*sensitivity list*) du process. En fait, un process peut être vu comme une sorte de sous-fonction, qui est appelée à chaque fois qu'un événement (changement de valeur) se produit sur un des signaux de sa liste de sensibilité. Relisez la section B.2.2 sur la simulation à événements discrets, et vous comprendrez.

La grande caractéristique du process concerne le mode d'exécution de ses instructions. En effet, même s'il se trouve au milieu d'une architecture (dont les instructions sont donc exécutées toutes en parallèle), le process assure la *sequentialité* de l'exécution de ses instructions, comme pour n'importe quel langage impératif classique.

Les instructions composant le cœur du process peuvent être principalement de trois types (en fait il y en a d'autres, mais ces trois-là sont largement suffisants) :

**Affectations** Même principe que les affectations dans une architecture :

```
signal <= expr;
```

**Constructions de plus haut niveau** Ce sont typiquement des conditionnelles ou des boucles. Par exemple :

```
if cond then
  instr1;
  instr2;
  ...
  elsif cond' then
  ...
  else
  ...
end if;

while cond loop
  instr1;
  instr2;
  ...
end loop;
```

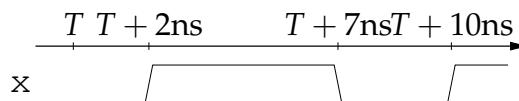
**Temporisations** Ces instructions permettent de donner des indications de temps concernant l'enchaînement séquentiel des instructions. L'instruction de temporisation la plus courante est le *wait* :

```
wait for n ns;
```

Cette instruction permet d'attendre *n* nanosecondes (on peut bien évidemment changer l'unité de temps) avant d'exécuter les instructions suivantes. Par exemple, dans un process, les instructions :

```
x <= '0';
wait for 2 ns;
x <= '1';
wait for 5 ns;
x <= '0';
wait for 3 ns;
x <= '1';
```

génèreront le chronogramme suivant (où le temps *T* correspond au temps au début de l'exécution de ces instructions) :

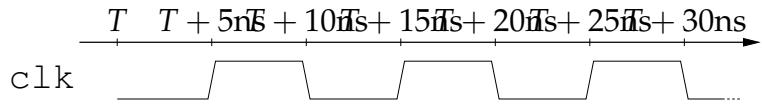


Un process sans liste de sensibilité étant exécuté en boucle, on peut ainsi aisément générer un signal d'horloge, comme le suivant, cadencé à 100MHz (période de 10ns) :

```

process
begin
  clk <= '0';
  wait for 5 ns;
  clk <= '1';
  wait for 5 ns;
end process;

```



### Restrictions d'utilisation

Eh ben voilà, on n'a plus qu'à écrire des bon vieux programmes dans des processus VHDL, pensez-vous ?

Eh bien malheureusement, non : il n'y a pas de compilateur qui sait transformer une telle spécification comportementale en un circuit. Autrement dit, sauf cas rares (que l'on verra dans la suite), les processus ne sont pas synthétisables. Ils sont simulables, et on les utilisera pour décrire des stimuli à nos circuits, mais en ce qui concerne la réalisation effective du circuit (sous forme de fils et de transistors), le synthétiseur se retrouve très vite dans les choux face à des processus.

L'utilisation de ces processus est donc **strictement** limitée aux fichiers de simulation, car ils nous permettront dans ce cadre de générer facilement signaux d'horloge et stimuli.

### Application aux registres

En fait, la seule utilisation des processus autorisée dans une architecture VHDL destinée à la synthèse est pour décrire des registres. Comme vous l'avez remarqué en TD, les éléments de mémoire décrits à partir de portes vous obligent à utiliser des boucles combinatoires, dont la stabilisation n'est jamais très claire.

Les processus offrent une alternative à cela, puisqu'ils sont par définition des éléments synchrones du circuit. Jetons par exemple un oeil au process suivant :

```

process (clk)
begin
  if clk'event and clk = '1' then
    q <= d;
  end if;
end process;

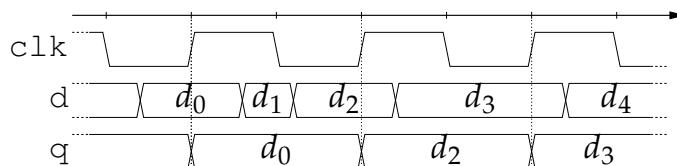
```

Ce process a le signal `clk` dans sa liste de sensibilité. Cela signifie qu'il est exécuté à chaque fois qu'un événement (changement de valeur) se produit sur `clk`. Il sera donc appelé sur chaque front (montant comme descendant) de notre horloge.

L'instruction `if` nous permet de vérifier les conditions suivantes :

- qu'il s'agisse bien d'un front d'horloge (`clk'event`),
- et que ce front soit montant, c'est-à-dire que l'horloge soit désormais à 1 (`clk = '1'`).

Une fois ces deux conditions vérifiées, le signal `q` reçoit la valeur de `d`. Si l'on trace le chronogramme correspondant, on voit bien que le comportement de ce process est le même que celui d'un registre (*flip-flop*) :



C'est là que le synthétiseur réalise une belle prouesse face à un tel process : en observant le chronogramme du process, il est capable de reconnaître un registre qu'il synthétisera alors proprement en utilisant sa propre bibliothèque de cellules standards.

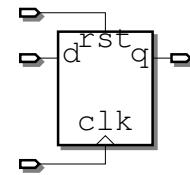
Mais c'est aussi là que se situe sa limite : mis à part ce process et quelques variations, point de salut. Parmi les variations acceptables, on peut trouver :

- le registre (*flip-flop*) avec *reset synchrone* (`rst`) :

```

process (clk)
begin
  if clk'event and clk = '1' then
    if rst = '1' then
      q <= '0';
    else
      q <= d;
    end if;
  end if;
end process;

```

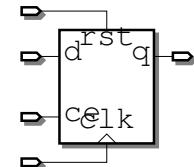


- le registre (flip-flop) avec *clock-enable* synchrone et (pour changer) reset *asynchrone* : on l'a ajouté dans la liste de sensibilité, ainsi un évènement sur reset va déclencher la remise à zéro du registre.

```

process (rst, clk)
begin
  if rst = '1' then
    q <= '0';
  elsif clk'event and clk = '1' and ce = '1' then
    q <= d;
  end if;
end process;

```



### B.3.7 Génération de circuit

#### Boucles de génération

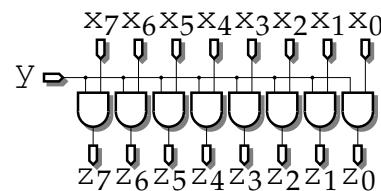
Parfois, lorsqu'un circuit (ou tout du moins un bout de celui-ci) est très régulier, il est plus simple de le générer à l'aide d'une boucle plutôt que d'écrire plusieurs fois la même chose en changeant juste les indices dans les vecteurs de signaux.

Prenons par exemple le bout d'architecture suivant :

```

z(0) <= x(0) and y;
z(1) <= x(1) and y;
z(2) <= x(2) and y;
z(3) <= x(3) and y;
z(4) <= x(4) and y;
z(5) <= x(5) and y;
z(6) <= x(6) and y;
z(7) <= x(7) and y;

```



Cela reste encore lisible pour huit signaux auxquels on applique juste une porte *and*, mais pour un nombre plus important de signaux ou des fonctions plus complexes, ça devient rapidement incompréhensible, source potentielle de plein d'erreurs et bien entendu indébuggable.

VHDL permet d'éviter cela en adoptant une construction à base de boucle :

```

inst : for i in m to n generate
  instr1;
  instr2;
  ...
end generate;

```

où *i* représente l'indice de boucle entier (qui n'a pas besoin d'être déclaré au préalable) qui va parcourir l'intervalle de *m* à *n* (inclus), instanciant à chaque fois le groupe d'instructions indiqué dans le corps de la boucle. L'identifiant *inst* va servir à identifier de manière unique chacune des instances du groupe d'instructions, en les nommant de *inst<sub>m</sub>* à *inst<sub>n</sub>*.

Pour revenir à l'exemple précédent, on peut donc écrire plus simplement les lignes suivantes, qui généreront exactement le même circuit :

```

inst_and : for i in 0 to 7 generate
  z(i) <= x(i) and y;
end generate;

```

### Conditionnelle

On peut aussi trouver la version conditionnelle du `generate` qui fonctionne de la même manière :

```
inst : if cond generate
  instr1;
  instr2;
  ...
end generate;
```

On peut écrire du VHDL récursif, c'est même parfois utile (un arbre est naturellement récursif) mais il faut bien comprendre que le circuit sera complètement déroulé avant la simulation. Dans le processus seront créés des tas de signaux et de composants, qui seront nommés plus ou moins intelligemment par le compilateur (en général, moins intelligemment que si vous les aviez nommés vous même, n'est-ce pas). Du coup, suivre ce qui se passe dans le circuit (déboggage, analyse le chemin critique, etc) devient difficile. Vous voilà prévenus.

## Annexe C

# Anatomie d'un Unix

On a vu du point de vue du matériel pas mal d'aspects systèmes. Voyons comment tout ceci se traduit pour l'utilisateur (en l'occurrence le programmeur).

Ce chapitre est une resucée de l'introduction au bouquin de Tanenbaum.

### C.1 Appels systèmes

La norme POSIX (qui rend les Unix compatibles entre eux) en liste assez peu. Tanenbaum en donne 53, dont quelques uns obsolètes mais conservés par souci de compatibilité.

Remarque : chaque appel système a une interface sous forme d'une fonction C, mais la mécanique sous-jacente variera d'un ordi à l'autre. Un appel système est plus qu'un appel de procédure. En particulier, il doit basculer le processeur en mode privilégié.

### C.2 Processus

Chaque processus est identifié par son PID, un entier. Il dispose de sa mémoire avec le code en bas, suivi du segment de données, et la pile qui part du haut de la mémoire. On peut d'ailleurs avoir des espaces d'adressages différents pour le code et les données.

On crée un processus par l'appel système `pid=fork()` qui crée une copie<sup>1</sup> identique du processus courant (même code, même contenu de la mémoire, même PC, mêmes registres, mêmes fichiers ouverts, tout pareil).

Ensuite, chaque processus continue son petit bonhomme de chemin. Pour que le fils puisse s'émanciper du père, `fork()` renvoie un entier qui vaut

- 0 pour le fils, et
- le PID du fils pour le père.

Un `fork()` dans un programme est donc typiquement suivi d'un test sur l'entier retourné pour que chaque processus sache qui il est.

J'ai parlé de fils et de père... L'ensemble des processus forme une généalogie, et la littérature des Unix est pleine de concepts oedpiens (le père tue son fils etc.).

Autre appels systèmes liés aux processus (sans les détails) :

- `pid=waitpid(pid, & statloc, opts)` : attend qu'un fils meure.
- `s=execve(name, argv, envp)` : remplace le code du processus courant par un autre code – name est le nom d'un fichier exécutable.
- `exit(status)` : fait mourir le processus
- `size=brk(addr)` : redimensionne la taille du segment de donnée du processus – `malloc()` va l'appeler si nécessaire<sup>2</sup>.
- `getpid()` Qui suis-je vraiment? Le fils a reçu 0 de `fork()`, mais un jour il voudra connaître sa véritable identité.
- ...

1. En pratique, on ne recopie pas vraiment la mémoire. On va se contenter de faire pointer les adresses virtuelles du fils vers les mêmes adresses physiques que celles du père, mais en lecture seule. Seul le père conserve le droit d'écrire dans cette mémoire. Tant que le fils ne fait que lire, pas de problème. Si le fils tente d'écrire, cela provoquera une exception qui va, pour le coup, faire une copie de la page physique concernée pour la donner au fils avec accès en écriture. Ainsi on ne fait la copie que si nécessaire. Bref, le prototype de cet appel système est tout simple, mais derrière il y en a du boulot.

2. Tiens, je fais `man brk` sous Linux pour voir pourquoi il s'appelle `brk` (non je ne vous dirai pas, allez-voir vous-même), et ils y disent *Avoid using brk() and sbrk() : the malloc(3) memory allocation package is the portable and comfortable way of allocating memory.*

### C.3 Signaux

Les processus communiquent entre eux de deux manières :

- soit par fichier (les tuyaux, vous savez, `cat *.c | grep "toto"`, sont vus par les processus comme des fichiers.)
- soit par signaux.

Un signal c'est l'abstraction de ce qu'on a appelé une interruption ou une exception du point de vue matériel. C'est envoyé soit par le hard, soit par un autre processus. Il y a un petit nombre de signaux différents.

Donc un processus peut envoyer des signaux. Il peut donc aussi en recevoir. Par défaut, un signal tue un processus, mais le processus peut annoncer qu'il s'attend à recevoir un signal et fournir pour cela l'adresse d'une fonction gestionnaire de signal, qui sera appelée s'il reçoit le signal (si l'adresse est nulle, le signal sera ignoré).

- `sigaction(sig, & action, & oldaction)` définit l'action à exécuter à réception d'un signal
- `sigreturn()` à appeler à la fin du traitement d'un signal pour reprendre le train-train du processus
- `kill(pid, sig)` envoie un signal
- `pause()` suspend ce processus jusqu'au prochain signal
- ...

### C.4 Fichiers

Tous les fichiers sont abstraits dans une seule arborescence. Par conséquent, il faut un appel système `mount()` qui définit la place d'un système de fichier qui arrive avec son arborescence à lui (par exemple un CDROM) dans l'arborescence globale.

Et voilà plein d'autres appels plus ou moins explicites (au choix pour les détails, la commande `man` ou le Tanenbaum) :

- `mknod`
- `open`
- `close`
- `read`
- `write`
- `lseek`
- `pipe`
- ...

Il y a aussi tous les appels qui concernent la gestion de l'arborescence des fichiers :

- `mkdir`
- `rmdir`
- `link`
- `unlink`
- `mount`
- `umount` (unmount était sans doute trop long)
- `chdir`

### C.5 Protection

Enfin, c'est bien beau de protéger les processus les uns des autres, mais il faut aussi protéger les utilisateurs les uns des autres. Chaque utilisateur est identifié par un entier, son UID. Il peut faire partie de différents groupes, chaque groupe identifié par son, devinez quoi, GID. Un processus peut connaître quel utilisateur l'a lancé par `getuid()`.

Comme tout est fichier, on n'a ensuite à protéger que des fichiers. Un fichier a un propriétaire (connu par son UID donc) et un groupe. On peut le changer par `chown()`. La protection se fait par trois fois trois bits (read, write et execute pour propriétaire, groupe et reste du monde) que l'on peut associer à un fichier par `chmod()`<sup>3</sup>. Chaque processus appartient à un utilisateur (à chaque PID est associé l'UID de l'utilisateur qui, consciemment ou non, l'a lancé), et un processus aura sur un fichier les droits correspondant à cet utilisateur.

3. Ces deux appels systèmes ont leur commande shell associée.

Il y a aussi deux bits, SETUID et SETGID. Quand un utilisateur Toto lance le programme `tata`, normalement (si le SETUID de `tata` vaut 0) l'UID du processus créé pour exécuter `tata` est celui de Toto. Si SETUID vaut 1, l'UID du processus `tata` devient celui du *propriétaire* du fichier `tata`. Cela permet de donner à Toto les droits d'un autre utilisateur (souvent le superutilisateur, root), de manière temporaire et dans un contexte réduit à un seul programme.



## Annexe D

# Systèmes de fichiers

Un fichier est une abstraction persistante et nommée des données.

- Persistante : le fichier survit à la mort du processus qui l'a créé.
- Nommée : le fichier est accédé par son nom.

### D.1 L'abstraction du fichier

Vue conceptuelle d'un fichier :

- Aux temps héroïques des cartes perforées, les fichiers étaient des séquences d'enregistrements de 80 caractères. Un enregistrement correspondait à une carte.
- Dans le monde Unix, développé aux temps héroïques des bandes magnétiques, un fichier est une séquence d'octets. Pour l'accéder, on passe par un *file descriptor* qui pointe quelque part dans le fichier. Le fd est typiquement initialisé au début du fichier et se déplace ensuite typiquement en avançant dans le fichier. Autrement dit, la lecture (`getc()`) ne prend pas une adresse.
- Dans le système de fichier des Macs avant qu'ils sombrent dans l'Unix, un fichier était composé de deux flux d'octets appelés *data fork* et *resource fork*, le second contenant les attributs du fichier (nom, permissions, mais aussi position dans le bureau, vignette à afficher, et tout ce qu'une application peut avoir envie d'associer avec un type de fichier particulier – pensez à toutes les infos que votre appareil photo numérique stocke avec chaque photo).
- NTFS généralise : un fichier est un ensemble d'attributs, chaque attribut étant un flux d'octets. Les données sont considérées comme l'un des attributs. C'est vrai ça, la frontière entre donnée et métadonnée étant tout ce qu'il y a de floue, autant tout mettre au même plan.
- Dans un SGBD (système de gestion de base de données) relationnel, un fichier est une base de donnée, composée de tableaux à deux dimensions. Les colonnes sont nommées, et chaque ligne correspond à une donnée. Les données sont accédées dans un tableau par une clé (la donnée d'une valeur d'une colonne).

Ceci montre que, comme tout le reste, la notion de fichier, au départ très proche du matériel, a évolué vers de plus en plus d'abstraction. La tendance est d'abstraire désormais même la localisation d'un fichier dans la hiérarchie des répertoires, au profit d'un accès par le contenu, à la Google.

Cela dit il faut bien distinguer la vue conceptuelle d'un fichier de la manière dont il est implémenté :

- le fichier Unix est implémenté quasi tel quel sur une bande magnétique.
- Stocké sur le disque, il est découpé en secteurs qui sont placés sur des pistes, et l'accès reste séquentiel, mais par bloc. On peut organiser les différents secteurs d'un gros fichier comme une liste chaînée, mais on peut aussi imaginer des mécanismes d'accès plus rapides (on va voir).
- On peut même programmer un *RAMDisk*, un disque virtuel stocké dans la mémoire vive. L'abstraction du fichier Unix ne profite alors pas de l'accès aléatoire de la mémoire.
- Il y a des bouquins entiers sur la manière de stocker une grosse base de donnée sur un disque. La meilleure manière dépend complètement du type d'accès que l'on veut privilégier.
  - Si un tableau est accédé essentiellement par une colonne donnée, on a intérêt à le trier selon cette colonne pour permettre une recherche dichotomique.
  - S'il est accédé souvent par deux colonnes, on ne peut trier que selon une des deux. On peut mettre en place un index (ensemble de couples (clé de la seconde colonne, pointeur vers ligne) trié selon la seconde colonne. Inconvénients : cela consomme de la place, et il faut maintenir l'index à jour.
  - Cela c'était en lecture. Si on doit souvent insérer des lignes, il faut mettre en place une structure de donnée du genre B-tree.

On commence à voir la notion de donnée et de métadonnée (donnée sur les données). Exemples de métadonnées (tous les systèmes de fichiers ne les offrent pas tous) :

- le nom du fichier
- sa taille
- l'identité du propriétaire du fichier
- les dates de création, de modification, etc
- le type du fichier (répertoire, programme exécutable ou fichier de donnée, fichier spécial – voir ci-dessous)
- le type du fichier (une image, du texte, un pdf, un fichier Word)
- eventuellement une vignette à afficher dans les navigateurs (qui écrase la vignette par défaut déduite du type de fichier)
- et des choses moins visibles pour l'utilisateur : le type d'organisation interne du fichier, un pointeur vers la fin si on pense qu'on voudra l'étendre, etc.

Types de fichiers : au moins le fichier standard et le répertoire. Sous Unix, on a en plus les fichiers spéciaux (character special files and block special files) utilisés respectivement comme interfaces vers des écrans/imprimantes et des disques physiques. Je les glisse sous le tapis.

La notion de type est importante. Le système doit savoir si un fichier est un programme exécutable par exemple. Il peut consacrer un bit spécial à cela. Il peut aussi examiner l'extension du fichier, ou les premiers caractères. Unix mélange un peu tout. C'est le résultat d'un équilibre entre le typage fort, qui protège les andouilles mais embête les gourous, et un typage faible, qui rend la vie plus simple mais plus dangereuse.

## D.2 Implémentation d'un système de fichiers

Commençons par les systèmes de fichier pour disque dur. L'unité d'allocation est appelée le bloc, typiquement 1Ko. C'est une abstraction du secteur physique. Les blocs sont attrapés par deux coordonnées, la piste et le secteur. Il y a un ordre séquentiel (une spirale) selon lequel l'accès est rapide.

Vous trouverez dans le Tanenbaum l'organisation de base d'un disque (au boot, le PC lit un bloc particulier, le MBR, et exécute le code qu'il y a dessus. Le MBR se termine par une table des partitions. Le premier bloc de chaque partition contient du code qui sert à lancer le système qui est sur cette partition. Etc. Il n'y a rien là dedans que vous n'eussiez inventé.)

Un répertoire est un fichier comme les autres, qui contient une liste de pointeurs (sur le disque) vers les différents fichiers. Les noms des fichiers sont plutôt stockés dans la structure de donnée du fichier lui-même que dans le répertoire.

Voyons à présent l'organisation des blocs des différents fichiers, du plus simple au plus compliqué.

### D.2.1 Allocation contigue

Bé voila, on met les fichiers à la suite les uns des autres sur le disque, sur la spirale. On consacre quelques blocs au début à contenir des pointeurs vers les débuts et longueur de chaque fichier. La lecture d'un fichier est aussi rapide que possible. L'accès aléatoire au milieu d'un gros fichier qui couvre plusieurs pistes est possible. Tout baigne.

Enfin, tout baigne au début. On remplit le disque petit à petit, en bourrant les fichiers à la fin. Quand on efface un fichier, il laisse un trou. Pour boucher un trou de manière optimale il faudrait connaître à l'avance la taille de chaque fichier au moment du `fopen()`, ce qui est rarement le cas. Donc on préfère remplir par la fin. Et quand le disque est plein, tout à coup le système doit tasser tous les fichiers au début en supprimant les trous, ce qui peut prendre plusieurs minutes. En général, cela arrive juste au moment où vous deviez rendre votre DM.

Donc au début on faisait comme cela, et puis on a fait mieux, (voir ci-dessous). Puis on a inventé le CDROM, et tout-à-coup l'allocation contigue redevient intelligente pour ces nouvelles technos : quand on grave un CD, on connaît la taille de tous les fichiers.

Papy Tanenbaum en tire la morale qu'il faut toujours s'intéresser aux méthodes du passé, car les nouvelles technos peuvent les remettre au goût du jour. Je suis assez d'accord, en tout cas c'est super vrai en arithmétique des ordinateurs.

### D.2.2 Organisation en FAT

FAT c'est *File Allocation Table*. Voici l'idée. Pour gérer les trous qui apparaissent, il est naturel de penser un fichier comme une liste chaînée de blocs de taille fixe (l'optimal est typiquement 1 ou 2 Ko, mais cela dépend du contexte bien sûr. cf ci-dessous). Chaque bloc commence par un pointeur vers le bloc

suivant sur le disque, suivi par les données. Un petit fichier n'utilise qu'un bloc, un gros en utilisera plusieurs, mais ils n'ont pas besoin d'être consécutifs (on essaiera tout de même qu'ils le soient pour accélérer l'accès séquentiel, bien sûr).

Du coup on peut remplir les trous précédents avec des blocs de nouveaux fichiers, quelle que soit leur taille. Mais le problème des listes chaînées, c'est le temps qu'il faut pour faire un accès aléatoire : il faut suivre la chaîne, ce qui représente autant d'accès disques.

L'idée de la FAT est de mettre tous les pointeurs dans un gros tableau qui pourra tenir en mémoire. On numérote les blocs de 1 à  $n$ , et la FAT est un tableau à  $n$  entrées, autant que le disque a de blocs. Un pointeur vers le premier bloc d'un fichier est traduit en un index dans la FAT. Si le fichier consomme plus d'un bloc, l'entrée de la FAT correspondante contient l'index du bloc suivant. Etc.

Avec une FAT il faut toujours suivre la chaîne pour faire un accès au milieu d'un fichier, mais c'est à présent en mémoire que cela se passe, et c'est 1000 fois plus rapide.

Naturellement, la FAT est aussi présente sur le disque lui-même.

C'était une bonne idée quand les disques étaient petits, mais un système de fichier à base de FAT consomme une taille mémoire proportionnelle au nombre de blocs. Pour un disque de 4Go avec des blocs de 1Ko, on a 4Mblocs, chaque entrée de la FAT doit faire 4 octets<sup>1</sup>, cela couté 16Mo. Pour un disque de 512 Go, on arrive à 2Go de RAM consacrées à la FAT. C'est pas terrible.

Bien sur, on peut laisser la FAT sur le disque et l'accéder plus lentement si on est ric rac en mémoire. Et cela peut se faire dans votre dos par un système de mémoire virtuelle.

### D.2.3 I-nodes

La dernière technique, développée dans le monde Unix, est le *index-node* ou i-node. Un i-node est une structure de données qui contient les différents attributs du fichier (nom, protection, etc) et un tableau de pointeurs vers des blocs. Ici aussi, cette structure de donnée est présente sur le disque, et recopiée en mémoire si nécessaire.

Le tableau est de taille fixe (on va faire en sorte qu'un i-node consomme un bloc par exemple), et si le fichier occupe plus de blocs que le tableau ne peut contenir, les trois dernières entrées sont des pointeurs vers un tableau d'indirection simple, un tableau d'indirection double, et un tableau d'indirection triple.

Pour les tout petits fichiers (moins d'un bloc), les inodes consomment plus de place sur le disque qu'une FAT : deux blocs par fichier. Par contre, ils ont deux avantages. Le premier est qu'on ne doit conserver en mémoire qu'autant d'inode qu'on a de fichiers ouverts : la consommation mémoire est indépendante de la taille du disque. Le second est que l'accès aléatoire est plus rapide, puisqu'on peut sauter directement au milieu du tableau.

Un avantage des i-nodes est de permettre facilement d'avoir des *hard links* : on peut avoir plusieurs répertoires pointant vers le même i-node. En fait, chaque i-node a un compteur de références, et il est détruit quand ce compteur atteint zéro – dans Unix, `rm` est en fait identique à `unlink`.

Il existe aussi des *soft links*, qui sont juste de très petits fichiers contenant un chemin (textuel) dans l'arborescence du disque. Contrairement aux *hard links*, ils peuvent pointer vers un fichier qui est sur un autre disque,

### D.2.4 Taille des blocs

Le choix de la taille optimale des blocs est un compromis entre deux questions : la place disque gaspillée par les petits fichiers qui ne rentrent pas dans un bloc (plus le bloc est petit mieux c'est) et la vitesse à laquelle on arrivera à lire un gros fichier (plus il est fragmenté en petits morceaux plus c'est mauvais).

Il y a aussi des contraintes liées à la mécanique : c'est mieux de faire tenir un nombre entier de blocs dans un cylindre, etc.

Bien sûr cela dépend du contexte, et en particulier de la taille médiane des fichiers sur le système, éventuellement pondérée par la fréquence de leur accès. Il vaut mieux considérer la médiane que la moyenne, qui peut être très perturbée par quelques très gros fichiers. Le Tannenbaum mentionne plusieurs mesures sur des vrais serveurs (académiques) ont donné des médianes similaires, à environ 1Ko en 1984 et 2Ko en 2005. Remarque personnelle : à mon avis sur votre PC à la maison, il y a plus de mp3 et de avi que de petits fichiers textes de 2Ko. Et plus sérieusement, même sur mon PC du boulot, j'ai essentiellement des PDF d'articles pas de moi, des sources latex de mes articles, slides et autres polys avec figures et cie, les sources de mes programmes, et la médiane est plus dans les 50Ko. Il faut vivre avec son temps.

1. Devinette : quelle est la différence entre FAT16 et FAT32 ?

### D.2.5 Gestion des blocs libres

Classiquement, soit par une liste chaînée, soit par un bitmap. La liste prend plus de place disque quand le disque est presque vide, mais on s'en fout puisque le disque est presque vide.

## D.3 Cohérence et fiabilité

Un système de fichier est un objet fragile. L'OS a souvent à lire un bloc, le modifier et le réécrire sur le disque. Si le courant est coupé au mauvais moment, au milieu d'une écriture, le disque peut rester dans un état incohérent, par exemple on a déjà écrit des données dans un bloc, qu'on a déjà enlevé de la liste des blocs libres, mais on n'a pas eu le temps de faire pointer aucun inode dessus. Ou le contraire, on a fait pointer un inode vers un bloc qui est encore dans la liste des blocs libres.

Pour éviter cela on a deux techniques. D'abord on a `fsck`, qui construit une liste des blocs occupés, une liste des blocs libres, et vérifie qu'un bloc n'est pas dans les deux listes à la fois, ou deux fois dans la même liste, etc.

Ensuite on a inventé les systèmes journalisés (c'est un australicisme). Le journal est à comprendre comme journal de bord, pas comme le Figaro. C'est juste un fichier spécial.

Avant de réaliser une opération potentiellement dangereuse, on écrit dans le journal qu'on va la faire. Puis on copie dans le journal l'état précédent du système de fichier. Puis on ferme le journal. Puis on fait l'opération dangereuse. Puis on rouvre le journal et on marque la modification comme terminée (*commit*). Ainsi,

- Si le courant est coupé alors qu'on était en train d'écrire l'intention dans le journal, on n'a pas touché au système de fichier. Au redémarrage, on voit un journal pas refermé, on l'efface et tout baigne.
- Si le courant est coupé alors qu'on est en train de faire la modification dangereuse du système de fichier, on a au redémarrage dans le journal tout ce qu'il faut pour restaurer le système de fichier dans son état précédent.

Une coupure de courant aura foutu le bordel, on aura perdu les données sur lesquelles on travaillait, qu'on était en train d'écrire, il n'y a rien à faire. Mais casser le système de fichier, donc même des données sur lesquelles on n'était pas en train de travailler, ce serait beaucoup plus grave.

Dans les détails, naturellement, c'est plus compliqué.

Parenthèse :

Cette idée est héritée du monde des bases de données, dans lesquelles la cohérence est capitale : quand on fait un virement d'une banque à l'autre, un plantage peut faire échouer le virement, ce n'est pas très grave (quelqu'un ralera !). Ce qui serait grave, c'est qu'un plantage laisse les bases de données dans un état incohérent : de l'argent viré d'un côté mais pas débité de l'autre côté, ou le contraire.

L'unité de modification d'une base de donnée est la transaction, qui est définie par les propriétés ACID :

- A - Atomicité : une transaction s'effectue complètement, ou pas du tout.
- C - Cohérence : une transaction prend une base de données dans un état cohérent, et la rend dans un état cohérent.
- I - Isolation : des transactions effectuées en parallèle ne s'influencent pas. Une transaction peut au plus en retarder une autre, car on ne peut accéder en écriture à une même information.
- D - Durabilité : une fois validée, une transaction survivra à toute défaillance dans le système.

## D.4 Performance

Cache disque, prefetch (on dit *read ahead*), on a déjà tout vu quand on faisait les hiérarchies mémoires. Il s'y ajoute le besoin de minimiser les déplacements du bras du disque. Cela passe par

- une allocation des blocs d'un fichier sur des cylindres aussi proches que possibles,
- un ordonnancement des accès, quand il y en a beaucoup à traiter en attente, par cylindres voisins,
- un éparpillement des i-nodes sur tout le disque, au lieu de les mettre tous au début du disque (approche naïve).
- ...

## D.5 Ce que je zappe

C'est tout ce qui a trait à la sécurité, la confidentialité, etc.

# Annexe E

## Le réseau

### E.1 Introduction

#### E.1.1 Besoins

Différents modes d'échange d'information :

- communication point à point orientée message
- communication point à point orientée flux (*stream*)
- diffusion
- réduction
- multidiffusion

Services annexes :

- adressage : adresse MAC, adresse IP, adresse en texte. Traduction de l'une en l'autre.
- encryption, authentification, sécurité
- localisation (réseaux mobiles wifi et téléphone)
- ...
- Partage des ressources physiques entre plusieurs processus/utilisateurs/ordinateurs
- multiplexage temporel (partage du temps, comme les processus se partagent le temps de calcul)
- abstraction possible : canaux logiques, ports
- notion de paquet encapsulant des données
- encapsulation multi-niveaux

Qualité :

- intégrité (pas toujours nécessaire, ex. TV sur friboxe)
- garanties temporelles (pas toujours nécessaires, ex. courriel)
- performance : débit, latence (performance du médium : "bande passante")

#### E.1.2 Architecture globale

C'est une architecture mêlant matériel et logiciel.

De haut en bas :

- Pour l'application : appels systèmes par exemple pour
  - demander l'adresse physique de la machine nommée [www.playboy.com](http://www.playboy.com)
  - envoyer un mail à [Florent.de-Dinechin@insa-lugdu.fr](mailto:Florent.de-Dinechin@insa-lugdu.fr) (opération de haut niveau)
  - envoyer un paquet sur le port 35 de la machine 11.22.33.44 (opération de bas niveau)
  - attendre un paquet
  - ouvrir un canal temps-réel crypté pour recevoir un flux de données de [www.playboy.com](http://www.playboy.com)
- Pour le système d'exploitation :
  - ouvrir/fermer des ports ou des canaux
  - mettre en œuvre les protocoles comme la traduction d'adresse
  - découper/encapsuler les données dans des paquets si nécessaire
  - recopier des tampons, éventuellement sous contrainte de temps réel<sup>1</sup>, jusqu'à avoir rempli le tampon de la carte réseau avec les informations qu'il faut

1. Ici "temps réel" signifie "en assurant un temps de transmission total inférieur à un maximum préalablement spécifié par le protocole, au vu de la lenteur ridicule des canaux d'entrée/sortie de l'utilisateur humain". Par exemple, moins d'un 25ème de seconde pour recevoir puis déchiffrer puis décompresser l'image suivante d'un film. Pour transmettre une information en vrai temps réel il faudrait déjà savoir la transmettre plus vite que la lumière. Heureusement, il y a de la marge entre la durée qu'il faut à la lumière pour faire le tour du monde et 1/25ème de seconde. Oui, je sais, la lumière va droit, elle ne fait pas le tour du monde.

- éventuellement, pour un nœud non terminal du réseau, propager les paquets des autres.
- Pour la carte réseau
  - envoyer et recevoir des paquets sur le réseau.
  - pour cela, implémenter un protocole de communication qui dépend de la technologie qu'elle a derrière. Exemples : ethernet, token-ring, ADSL, ATM
  - certains de ces protocoles (ATM) sont eux-mêmes à base de paquets plus petits.

Certains des services de l'Internet demandent une administration centralisée (annuaire des noms de domaine). Toutefois le gros du travail est décentralisé : l'approche décentralisée est plus robuste en cas de panne, supporte mieux la charge là où elle est, et passe mieux à l'échelle. Le routage des paquets est également totalement décentralisé.

### E.1.3 En résumé : les défuntes couches OSI

Communications physiques dépendant de la technologie, de si on communique avec son voisin de bureau ou avec Hong Kong.

Mais vue applicative unique : clients, serveurs et services.

Abstraction derrière des notions d'adresse, de canal, de port, de paquet, qui se traduit par des couches logicielles et matérielles. Pour vos données, cela se traduit par des couches d'emballage (encapsulation), qui au passage consomment de la bande passante.

Effort de normalisation par l'ISO (*international standard organisation*) : les couches OSI (*open system infrastructure*). Jamais vraiment pris racine, parce que pendant les négociations on continuait de construire un internet qui marchait fort bien – entre autre parce qu'il reprenait des bonnes idées de l'OSI. Même le fait que l'état américain rende obligatoire la compatibilité OSI n'a pas réussi à l'imposer.

Bref, les voilà :

1. couche physique (envoie des bits de point à point)
2. couche lien de donnée (gère l'emballage et le déballage des bits dans des paquets appelés alors *frame*)
3. couche réseau (gère le routage des paquets)
4. couche transport (gère les communications processus à processus)
5. couche session (gère les différents flux associés à une même application, par exemple synchronise les paquets audio et vidéo d'un film)
6. couche présentation (définit la taille en bits d'un entier, le format d'un flux vidéo, ce genre de choses)
7. couche application

Un nœud qui relaye juste un message qui ne lui est pas destiné n'a besoin pour cela que des trois premières couches. La définition des trois couches supérieures est un peu vague, les frontières sont soit mal définies, soit trop bien définies ce qui interdit certaines optimisations.

On va voir que l'internet, lui aussi organisé en couches, est plus flexible.

## E.2 Les couches de l'internet

On n'a que 4 couches :

1. la couche "carte réseau" : elle dépend complètement de la technologie qui est derrière, et les couches supérieures ne veulent pas en entendre parler. Une carte réseau va faire intervenir du logiciel et du matériel, et peut éventuellement travailler sur des paquets, etc.
2. la couche IP (*internet protocol*) : elle définit l'abstraction du réseau logique, qui va permettre à plein de technologies différentes de coopérer dans un seul réseau Internet avec un grand I. Pour cela, cette couche définit la notion d'adresses IP, de paquet IP avec son format incluant l'adresse IP du destinataire, etc (voir ci-dessous).
3. Une couche pour TCP, UDP et quelques autres (en gros la couche transport)
4. La couche application

Ces couches sont organisées de manière plus flexible que chez OSI : il est courant que la couche application parle directement à IP, voire à la carte réseau, sans passer par une couche transport.

### E.2.1 IP

IP *internet protocol*<sup>2</sup> définit :

- un espace d'adressage
- un service de transmission de paquets sans aucune garantie :
  - un paquet peut se perdre
  - un paquet peut arriver deux fois
  - les paquets peuvent arriver dans le désordre.

C'est une philosophie "best effort" : on fait ce qu'on peut. En terme d'implémentation, la traduction est un protocole assez léger qui par conséquent "run over anything". *Anything* veut dire, un réseau fiable ou bien un réseau que quand on envoie des données dessus on n'est jamais sur qu'elles arrivent, et dans quel état. C'est en fait le bon choix, parce qu'il est facile d'implémenter une couche IP au dessus d'une technologie réseau fiable, alors que le contraire serait difficile.

Si vous voulez une garantie de service, ce sera de la responsabilité des couches supérieures (on va voir TCP et UDP ci-dessous).

Un paquet IP est composé d'un *en-tête* (*header* en impérialiste) qui fait typiquement 20 octets mais peut faire plus si nécessaire. Dans cet en-tête, à des emplacements fixés par la norme IP, on trouve

1. la taille totale du paquet, la taille de l'en-tête
2. l'adresse de destination, l'adresse de départ
3. un *checksum* de 16 bits qui permet de vérifier l'intégrité du paquet
4. des informations indiquant si ce paquet est un morceau d'un paquet plus gros, et si oui quelle est sa place dans ce paquet plus gros (au cours de la transmission d'un paquet IP, il pourra être fragmenté).
5. un champ "TTL" pour *time to live* : il est initialisé à 64 typiquement, et décrémenté à chaque passage par un nœud de routage. Arrivé à zéro, le paquet est jeté sans autre forme de procès. Ainsi, en cas de boucle dans le graphe de routage, les paquets ne s'accumulent pas.
6. un champ donnant la version du protocole IP utilisée (de nos jours IPv4 ou IPv6)
7. un champ "protocol" qui décrit le protocole de la couche supérieure qui a envoyé ce paquet
8. des champs optionnels pour mettre d'autres informations sur le paquet

Un gros paquet est fragmenté en paquets plus petits au départ, mais pourra être fragmenté en paquets encore plus petits durant le transport, selon les caprices des nœuds à travers lesquels il passe. Par contre il ne sera réassemblé qu'à destination.

Lorsqu'on doit fragmenter un gros message en plusieurs paquets plus petits, quelle doit être la taille de ces petits paquets IP ? Eh bien la couche IP n'impose rien, c'est la couche supérieure qui choisit. Typiquement, cette dernière peut interroger au préalable la couche "carte réseau", qui pourra lui répondre que ses paquets à lui ne doivent pas dépasser 1500 octets (ethernet) ou 4500 octets (FDDI). La couche supérieure fragmentera alors ses gros paquets en fonction de cela : à tout prendre, c'est le bon choix pour causer avec les voisins. Cela dit, si la couche IP passe des paquets trop gros pour la couche carte réseau, cette dernière les fragmentera et les réassemblera. On y perdra juste un peu en performance.

### E.2.2 Routage (commutation de paquets)

IP définit des adresses hiérarchiquement : il y a une part qui identifie un réseau, et une part qui identifie un nœud dans le sous-réseau. Certains nœuds (appelés routeurs) connectent plusieurs réseaux : ils ont alors plusieurs adresses IP, une dans chaque sous-réseau. Un pied dans chaque réseau, quoi. De plus, les routeurs ont une table qui liste, pour tous les sous-réseaux connus, sur quelle interface ils doivent transmettre un paquet.

Il y a un protocole, DHCP (dynamic host configuration protocol) qui définit la manière dont un ordinateur nouvellement arrivé sur le réseau va recevoir une adresse IP.

Il y a aussi un protocole ami de IP, ICMP (control message) qui en cas d'échec à livrer le message (les raisons ne manquent pas), envoie un message à l'expéditeur, sans garantie qu'il arrive bien sûr. ICMP définit aussi des messages plus positifs, comme "Il existe un chemin plus court de tel réseau à tel réseau" qui permet de mettre à jour les tables de routage.

En fait, la construction de ces tables de routage mérite tout un chapitre à elle seul. Il faut aussi éviter les embouteillages, minimiser la latence, etc.

2. On peut utiliser le mot "internet" pour '*internetwork*, réseau inter-réseau. Il ne prend alors pas de majuscule. Avec une majuscule, il s'agit de l'Internet omniscient que vous connaissez. Certains ne lui mettent alors plus d'article. Et le vouvoiement s'impose quand on s'adresse à Lui.

### E.2.3 UDP

User Datagram Protocol. Sans garantie non plus, mais il vérifie des *checksums*. Couche qui assure la communication de processus à processus. Se contente de démonter et réassembler des paquets pour les passer par IP.

Comment identifier un processus ? Le PID peut varier d'une session à l'autre. On l'abstrait donc par la notion de *port*. La couche UDP distribue les paquets selon des ports, et chaque processus surveille un port donné.

### E.2.4 TCP

Transmission Control Protocol.

Tout comme UDP, mais lui assure des garanties.

Du point de vue applicatif, flux d'octets envoyés et reçus dans l'ordre.

Derrière, système de requête/acquittement + fenêtre glissante.