# Module 1: Introduction


INF8601: Systèmes informatiques parallèles

Michel Dagenais

# Sommaire

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Site Moodle
- Situation du cours
- Objectifs du cours et des laboratoires
- Déroulement du cours
- Evaluation
- Lectures
- Contenu du cours
- Horaire

# Evolution technologique: 5 générations

ÉCOLE POLYTECHNIQUE MONTRÉAL

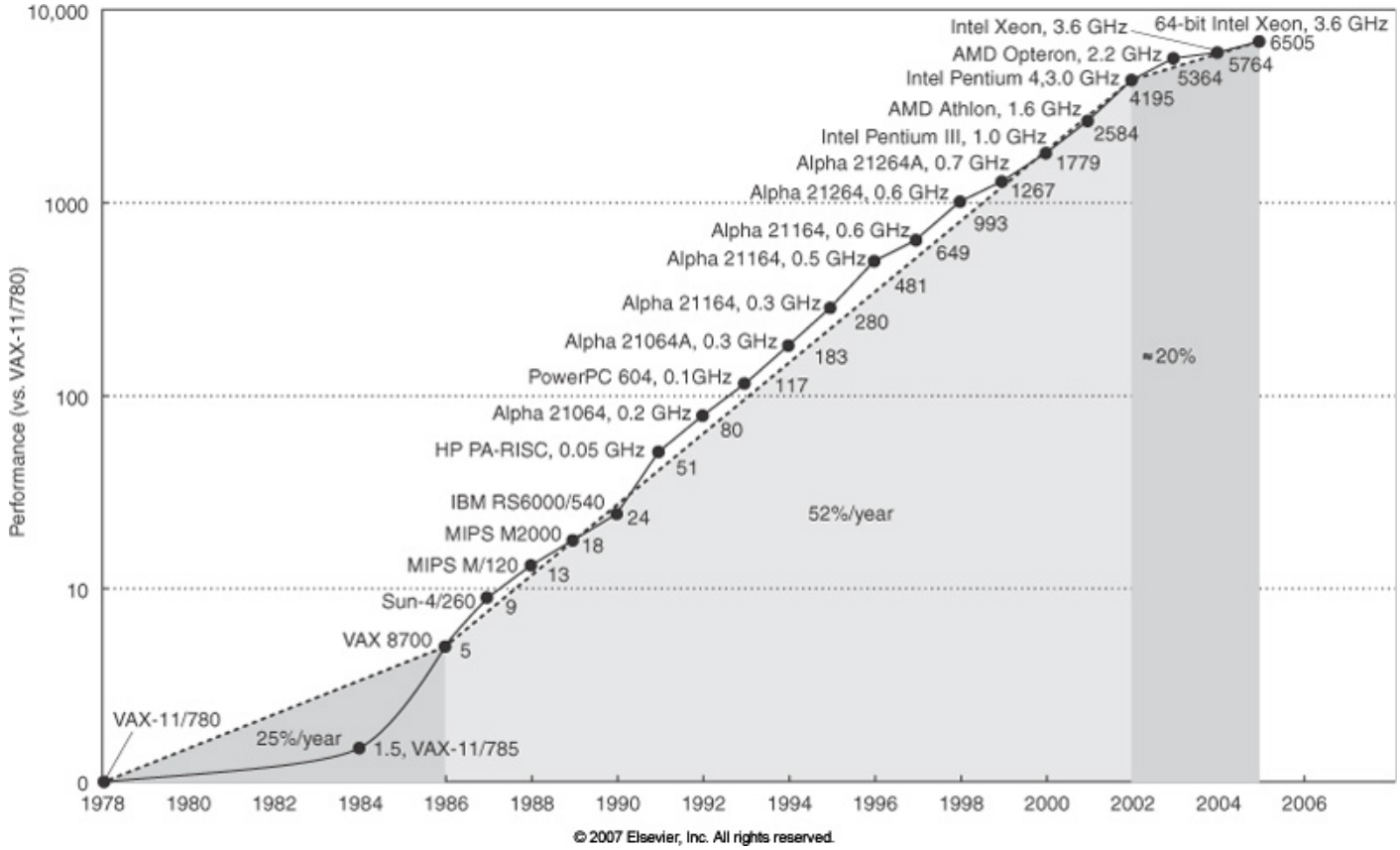
- 1940 - Lampes
- 1960 - Transistors
- 1970 - Circuits intégrés
- 1980 - Circuits LSI/VLSI
- 2000 - Ordinateurs de 5ème génération, parallèles, intelligents...

# Evolution technologique

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Circuits logiques: 60% plus de transistors par an; la rapidité augmente presqu'en proportion de la densité;
- Mémoire vive: 60% plus dense à chaque année, vitesse 30% seulement par 10 ans;
- Mémoire disque: 50% plus dense à chaque année, vitesse 30% seulement en 10 ans.

# Vitesse des processeurs



# Densité des processeurs

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Intel 4004, 1971, 10000nm, .74MHz, 2300 transistors
- Intel 8086, 1978, 3000nm, 8MHz, 29000 transistors
- Intel 80386DX, 1985, 1000nm, 33MHz, 275000 transistors
- Pentium Pro, 1995, 600nm, 150MHz, 5.5M transistors
- Pentium 4, 2000, 180nm, 1.7GHz, 42M transistors
- Intel Core 2, 2006, 65nm, 3.0GHz, 291M transistors
- Xeon Westmere, 2012, 32nm, 2600M transistors
- 15 cores Xeon Ivy Bridge EX, 2014, 22nm, 4310M transistors
- 22 cores Xeon Broadwell, 2016, 14nm, 7200M transistors

# Périphériques

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Interface réseau: 1Gbit/s, 12$;
- Disque: 6TB, 750MB/s, $150;
- Mémoire non volatile: 1TB, 750MB/s, $300;
- Rubans: LTO-5, 3TB, $2000;
- Carte graphique: GP100 Pascal, 16nm, 15300M transistors, 3584 coeurs;

# Répertoires d'instruction

ÉCOLE POLYTECHNIQUE MONTRÉAL

- IBM 360
- DEC VAX 11
- Motorola 68000
- Intel 80386
- Sun SPARC
- IBM PowerPC
- ARM
- Intel Itanium
- Intel 64 (AMD64)

# Classes d'ordinateurs

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Super-ordinateur, entre 5 et 20M$, grande pièce dédiée;
- Serveur d'entreprise (mainframe), entre 1 et 5 M$, coin de pièce;
- Serveur départemental, occupe un coin de pièce, entre 50K$ et 1 M$;
- Poste de travail (workstation), dessus de bureau, 4K$ à 50K$;
- Micro-ordinateur, dessus de bureau ou creux de la main, 100 à 4K$;

# Nouvelles classes d'ordinateur

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Le nuage, grappe d'ordinateurs flexible et efficace avec virtualisation, ressources à la demande et migration.
- Clients, de table, portatif, tablette, cellulaire.
- Bon rapport performance prix, importance de la consommation d'énergie, tout est un système parallèle, hétérogène en réseau!

# Tendances

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Plafonnement de la vitesse par processeur (taille des transistors, voltage, dissipation de chaleur);
- Augmenter le nombre de processeurs, meilleure performance / prix, consommation flexible;
- La mémoire centrale et les disques n'augmentent pas de vitesse autant que le processeur;
- Client capable d'exécuter des jeux 3D et du vidéo HD;
- Complexité des systèmes: outils de monitoring, vulnérabilités nombreuses.

# Historique des systèmes parallèles

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Burroughs D825, 1962, 4 processeurs;
- Honeywell Multics System, 1969, 8 processors;
- ILLIAC IV, 1965-1976, Vector machine, 256 units;
- Cray 1, 1976, Vector machine;
- Cray 2, 1985, 8 processors;
- Multi-processeurs (SGI, SUN, IBM);
- Grappes de calcul (Intel, IBM, HP...);
- Grappes de multi-processeurs hétérogènes (GPGPU);

# Programmation des systèmes parallèles

ÉCOLE POLYTECHNIQUE MONTRÉAL

- FORTRAN, compilateurs parallélisants, HPF;
- C, Parallel C, UPC, OpenMP (C et FORTRAN);
- PVM, Linda, MPI;
- CUDA, OpenCL

# Hiérarchie de mémoire

ÉCOLE POLYTECHNIQUE MONTRÉAL

- De haut (vite) en bas (dense): registres, cache L1/L2/L3, mémoire vive (NUMA), disque SSD, disque magnétique;
- L'adresse se décompose en numéro de bloc et en adresse dans le bloc;
- Taux de succès: proportion du nombre d'accès complétés au niveau le plus haut sans interaction avec le niveau inférieur;
- Taux d'échec: (1-succès);
- Temps d'accès succès: temps d'accès lorsque c'est complété au niveau le plus haut;
- Pénalité d'échec: temps pour compléter la communication entre les deux niveaux et transmettre la donnée au niveau supérieur (CPU), (accès, transfert complété), outre le temps d'accès succès;
- Temps d'accès moyen = temps d'accès succès + taux d'échec * pénalité d'échec

# Mémoire cache

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Blocs de 4 à 128 octets;
- Temps d'accès succès 1-4 cycles;
- Pénalité d'échec 8-32 cycles (accès 6-10), (transfert 2-22);
- Taux d'échec 1% à 20%;
- Dimension 1K à 512K (Intel Core I5, 32K I et 32K D);
- Correspondance directe, associative, par ensemble;
- Remplacement aléatoire ou LRU;
- Ecriture simultanée ou réécriture;

# Mémoire centrale

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Pages de 4K, 2M ou 4M octets;
- Temps d'accès succès 10-100 cycles;
- Coût de l'échec 500000-6000000 cycles (quelques dizaines de milisecondes);
- Taux d'échec .00001%-.001%;
- Table de pages à 1, 2 ou 3 niveaux, (associatif, remplacement LRU);
- Conversion adresse logique à adresse physique par matériel avec cache (TLB);
- Chargement des pages par le système d'exploitation;
- Une portion du système d'exploitation est toujours en mémoire centrale.

# Le système d'exploitation

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Interface avec les périphériques (pilotes);
- Interface avec les processus (appels systèmes);
- Gestion de la mémoire (mémoire des processus, mémoire virtuelle, tampons E/S);
- Systèmes de fichiers;
- Contrôle des accès;
- Ordonnancement;
- Gestion des interruptions.

# Top 500.org

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Ordinateurs les plus puissants au monde;
- Modélisation de phénomènes physiques (météo, turbines, fuselages, réactions chimiques/nucléaires, déchiffrement);
- Ordinateurs multi-coeurs, parallèles;
- Architecture? OS? Pays?

# Taxonomie de Flynn

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Flux d'instructions et de données.
- SISD: ordinateur conventionnel.
- SIMD: une opération peut opérer sur plusieurs données parfois de manière conditionnelle (Connection Machine 65536 processeurs de 1 bit, Ordinateur vectoriel, GPGPU).
- MIMD: SMP, multi-core, many-core, grappe, grille, nuage...

# Parallélisme

ÉCOLE POLYTECHNIQUE MONTRÉAL

- La même opération sur des données différentes.
- Le même programme sur des données différentes.
- Différents programmes sur les mêmes données ou des données différentes.

# Partage de données

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Mémoire partagée uniforme.
- Mémoire partagée répartie, accès non-uniforme (NUMA).
- Mémoire partagée répartie, par logiciel.
- Envoi de messages.

# Fiabilité

ÉCOLE POLYTECHNIQUE MONTRÉAL

- MTBF: Mean Time Between Failure. Temps moyen entre les pannes.
- MTTF: Mean Time to Failure. Temps moyen jusqu'à la panne.
- MTTR: Mean Time to Repair. Temps moyen pour réparer la panne.
- Fiabilité: F = MTBF= MTTF + MTTR
- Disponibilité: D = MTTF / (MTTF + MTTR)

# Fiabilité

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Déterminer un taux de panne (“failures in time”), FIT = 1 / MTTF, exprimé en # de pannes par milliard d’heures (environ 114000 ans).
- Les taux de pannes s’additionnent si la probabilité de panne ne dépend pas de l'âge (accident aléatoire plutôt qu'usure).
- Exemple: MTTF total pour 10 disques (MTTF=1 Mh), 1 contrôleur (MTTF= 0,5 Mh), 1 alimentation (MTTF= 0,2 Mh), 1/(10/1Mh + 1/,5MH + 1/.2MH) = 58823 heures

# Probabilités

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Probabilité que deux événements indépendants de probabilité p1 et p2 se produisent au même moment: p1 x p2. Par exemple, on ne peut travailler si tous les ordinateurs d'un laboratoire sont en panne en même temps.
- Les probabilités de cas disjoints peuvent s'additionner. La probabilité de pouvoir travailler, si un ordinateur ou plus est fonctionnel sur n, est la somme des probabilités que 1 ou (+) 2 ou (+)... n ordinateurs soient fonctionnels.
- La probabilité que m ordinateurs sur n soient fonctionnels, si la probabilité pour un ordinateur est p, est:

$$\frac{n!}{m! \times (n-m)!} \times p^m \times (1-p)^{n-m}$$

- Il est parfois plus simple de prendre le problème à l'inverse: probabilité d'avoir au moins un fonctionnel = somme des probabililtés que 1, 2... n sur n soient fonctionnels = 1 – probabilité de tous défectueux = 1 - $(1-p)^n$.

# Probabilités: attention

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Probabilité que deux événements indépendants de probabilité p1 et p2 se produisent au même moment: p1 x p2. Par exemple, un ordinateur est fonctionnel si la carte mère (p1) et le disque (p2) sont fonctionnels. Il y a une panne sauf si les deux sont fonctionnels.

  1 - p1 x p2

- Peut-on plutôt prendre la probabilité de panne de carte + la probabilité de panne de disque?

  (1 - p1) + (1 - p2) ≠ 1 - p1 x p2

- Quelle est la différence? La panne de disque se produit indépendamment de la carte qui peut être fonctionnelle ou non. On compte donc deux fois le cas où le disque et la carte sont défectueux. Les deux éléments sommés se recoupent alors qu'il faut bien tout séparer. Si p1 et p2 sont grands, la partie sommée en double est petite et le résultat est incorrect mais très proche. Si p1 et p2 sont petits, le résultat dépasse largement 1!

  (1 - p1) x p2 + (1 - p2) x p1 + (1 - p1) x (1 - p2) = (1 - p1) + (1 - p2) - (1 - p1) x (1 - p2) = p2 - p1 x p2 + p1 - p1 x p2 + 1 - p2 - p1 + p1 x p2 = 1 - p1 x p2

# Probabilités: encore

- La probabilité que m disques sur n soient fonctionnels.

$$\frac{n!}{m! \times (n-m)!} \times p^m \times (1-p)^{n-m}$$

- Pour 4 disques sur 5, peut-on dire $p^4$? Ce serait la probabilité que 4 disques spécifiques soient fonctionnels, e.g. les 4 premiers. Il y a 5! / (4! x 1!) = 5 combinaisons possibles de disques fonctionnels.
- Peut-on dire 5 x $p^4$? Si p = 1, on pourrait alors avoir une probabilité de 5, ce qui est impossible. Les éléments se recoupent, les 4 premiers disques fonctionnels n'empêchent pas le 5ème d'être fonctionnel et cela recoupe donc la probabilité d'avoir 4 autres fonctionnels.

# Performance

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Temps écoulé.
- Temps d'utilisation (CPU, mémoire, disque...).
- Rendement, résultat/temps par rapport aux ressources utilisées.
- Accélération de performance A = performance avec amélioration / performance sans amélioration (ou temps sans amélioration / temps avec amélioration)
- Accélération partielle, (fraction f du traitement), et totale résultante, loi d'Amdahl. At = 1 / ( (1-f) + f / Ap).

# Performance parallèle

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Facteur d'accélération, S(p) = Temps séquentiel (optimal) / Temps avec p processeurs.
- Efficacité, E(p) = S(p) / p
- Temps de réponse optimal versus coût optimal...
- Si fraction parallélisable f accélérée de p, S(p) = 1 / ((1-f) + f/p), S(p) tend vers 1/(1-f) si p très grand.

# Bancs d'essais

ÉCOLE POLYTECHNIQUE MONTRÉAL

- Linpack (Top500.org).
- SPEC2006: 12 entiers (9 C, 3 C++), 17 point flottant (6 FORTRAN, 4 C++, 3 C, 4 C et FORTRAN).
- TPC: Transactions.