



Nanoélectronique : impact sur les architectures et le logiciel

Hervé Fanet
CEA-LETI



- La loi de Moore mais....
- Limites théoriques et pratiques
- Les nouvelles architectures
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- Tentative de conclusion



- La loi de Moore mais....

Limites théoriques et pratiques

- Les nouvelles architectures
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- Tentative de conclusion

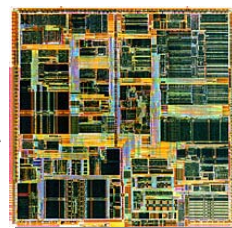
Les deux secrets de la micro-électronique

Miniaturisation

De plus en plus de transistors par centimètre carré de Silicium



En 2015, l'ensemble des livres de la BNF F. Mitterrand sur une seule puce de silicium !



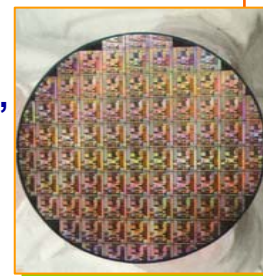
1 lot de fabrication à ST Crolles 1 : = 100 milliards de transistors !!!

Traitement collectif

De plus en plus de centimètres carrés de Silicium traités simultanément

Plaques de silicium de 100, 200, puis 300 mm de diamètre

Lot de fabrication



Réduction des coûts



La micro-électronique et la réduction des coûts

75 000 €
(500 000 F)

Une réduction des coûts unique dans l'histoire de l'industrie



1973

6 000 €
(40 000 F)



1977

450 €
(3 000 F)



1981

120 €
(800 F)



1984

30 €
(200 F)



1987

5 €
(35 F)



1990

0,5 €
(3 F)



1995



0,06 €
(40 centimes)

2000

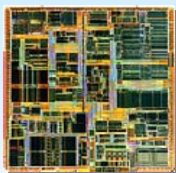


0,005 €
(3 centimes)

2005

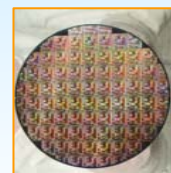
**Coût de 1 million de transistors
(équivalent à un livre de 200 pages)**

Miniaturisation



De plus en plus
de transistors par
mm² de silicium

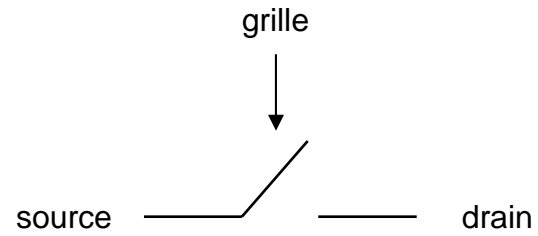
Traitement collectif



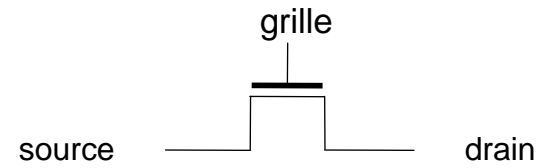
De plus en plus de mm²
de silicium fabriqués
simultanément

Une brique de base : le transistor MOS

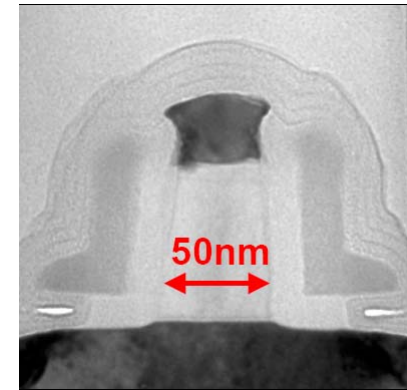
switch



transistor

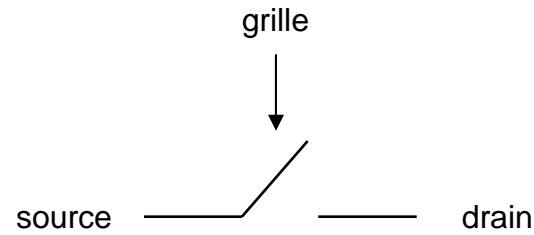


Transistor

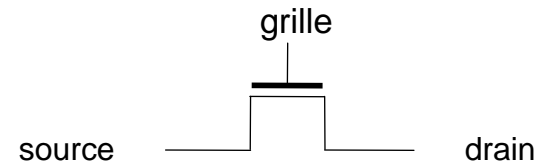


Une brique de base : le transistor MOS

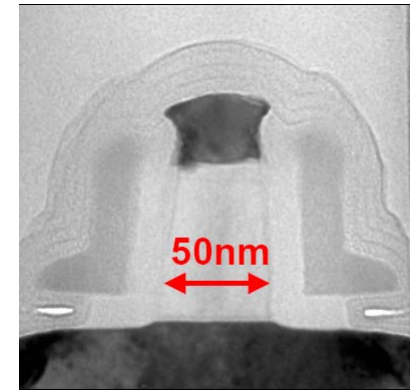
switch



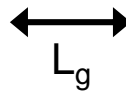
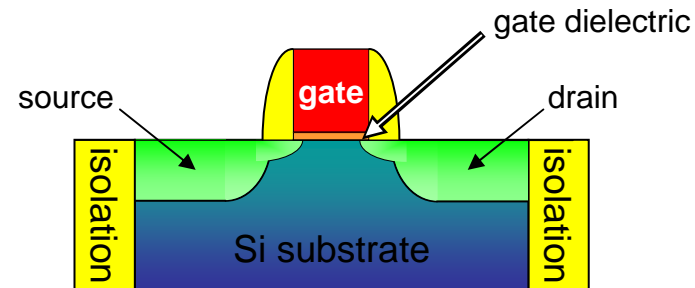
transistor



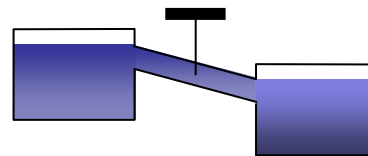
Transistor



physical
cross section

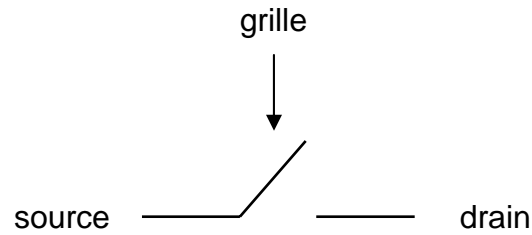


hydrodynamical
analogy

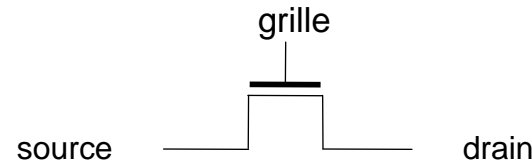


Une brique de base : le transistor MOS

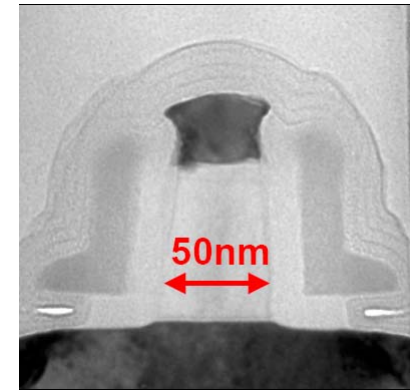
switch



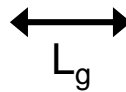
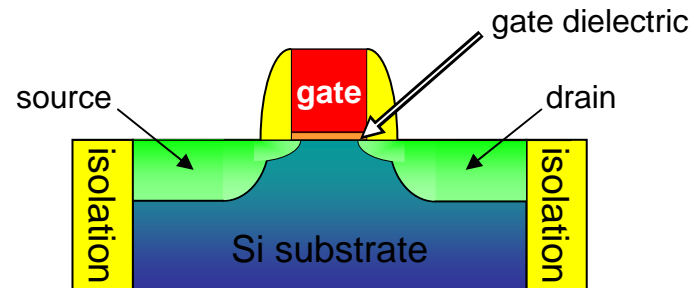
transistor



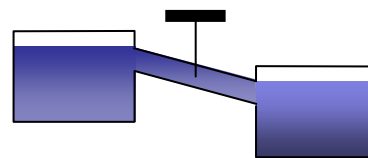
Transistor



physical
cross section



hydrodynamical
analogy

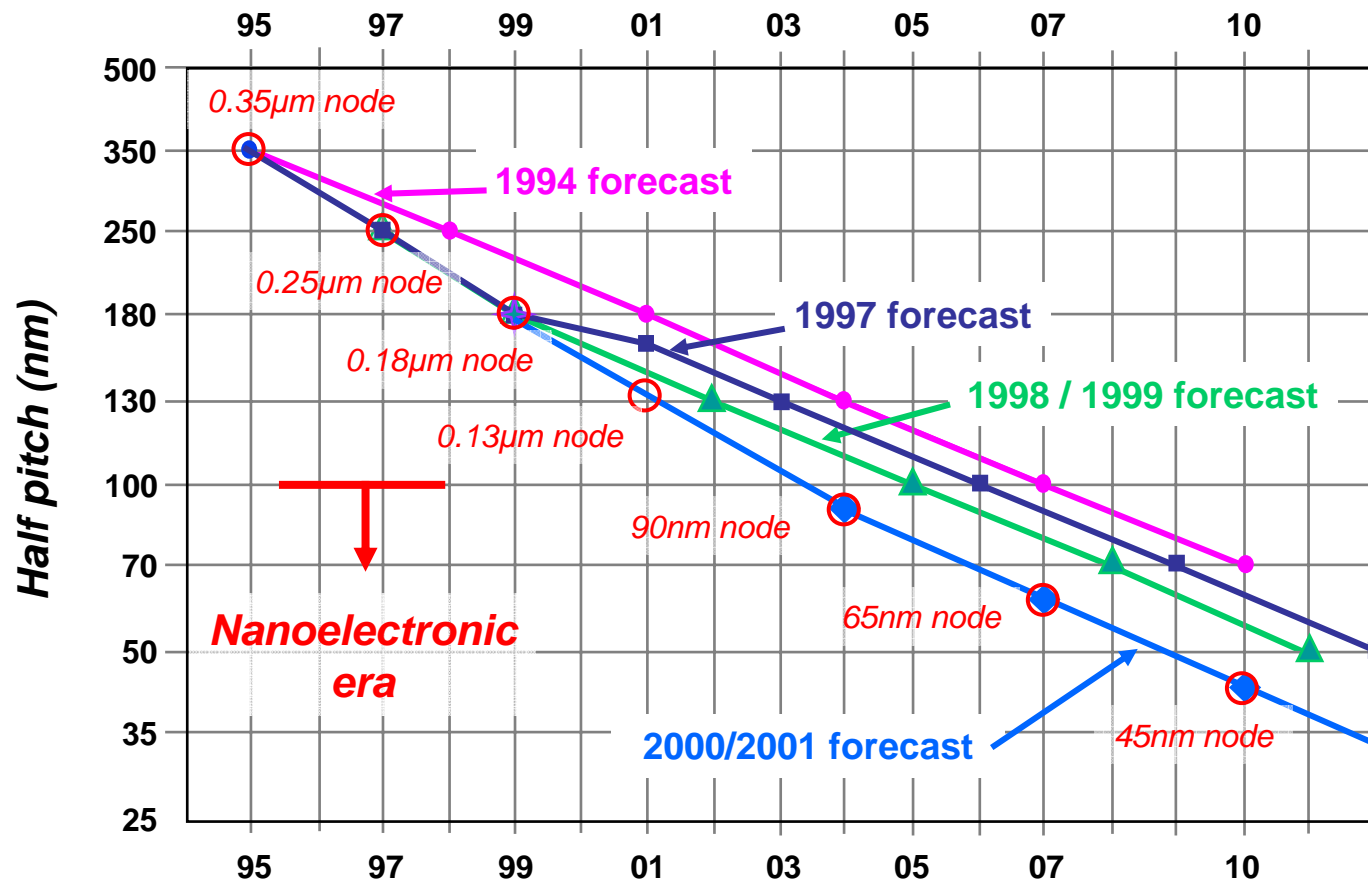


Deux paramètres
décisifs :

I_{off}

I_{on}

La loi de Moore





La loi de Moore mais....

- **Limites théoriques et pratiques**

Les nouvelles architectures

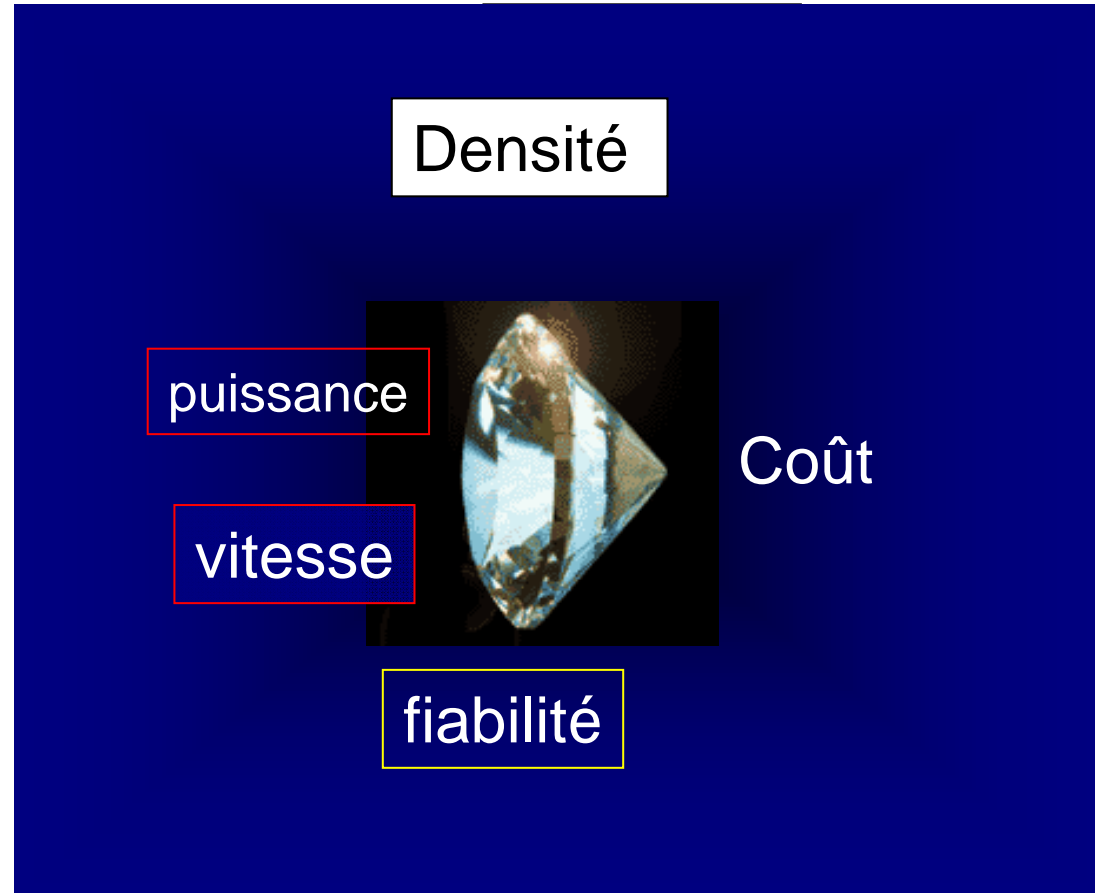
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- Tentative de conclusion

Conséquences de la miniaturisation



R. Dennard © IEEE

dimensions	t_{ox}, L, W	$1/\alpha$
dopage		α
tension		$1/\alpha$
densité		α^2
retard		$1/\alpha$
puissance		$1/\alpha^2$



Smaller = better

Limites théoriques

- Loi de Boltzmann + Shannon : énergie minimale de commutation

$$E \succ k_B T \ln 2$$

- Principe incertitude de Heisenberg : longueur minimale du canal

$$L \succ \frac{\hbar}{\sqrt{2meV}}$$

Limites théoriques

$$x_{\min} = \frac{\hbar}{\Delta p} = \frac{\hbar}{\sqrt{2m_e E_{bit}}} = \frac{\hbar}{\sqrt{2m_e k_B T \ln 2}} = 1.5 \text{ nm}$$

$$n_{\max} = \frac{1}{x_{\min}^2} = 4.7 \times 10^{13} \text{ devices/cm}^2$$

$$t_{\min} = \frac{\hbar}{\Delta E} = \frac{\hbar}{k_B T \ln 2} = 0.04 \text{ ps}$$

$$P = \frac{n_{\max} E_{bit}}{t_{\min}} = 3.7 \times 10^6 \text{ W/cm}^2$$

Sources INTEL

Limites théoriques

$$x_{\min} = \frac{\hbar}{\Delta p} = \frac{\hbar}{\sqrt{2m_e E_{\text{bit}}}} = \frac{\hbar}{\sqrt{2m_e k_B T \ln 2}} = 1.5 \text{ nm}$$

$$n_{\max} = \frac{1}{x_{\min}^2} = 4.7 \times 10^{13} \text{ devices/cm}^2$$

$$t_{\min} = \frac{\hbar}{\Delta E} = \frac{\hbar}{k_B T \ln 2} = 0.04 \text{ ps}$$

$$P = \frac{n_{\max} E_{\text{bit}}}{t_{\min}} = 3.7 \times 10^6 \text{ W/cm}^2$$

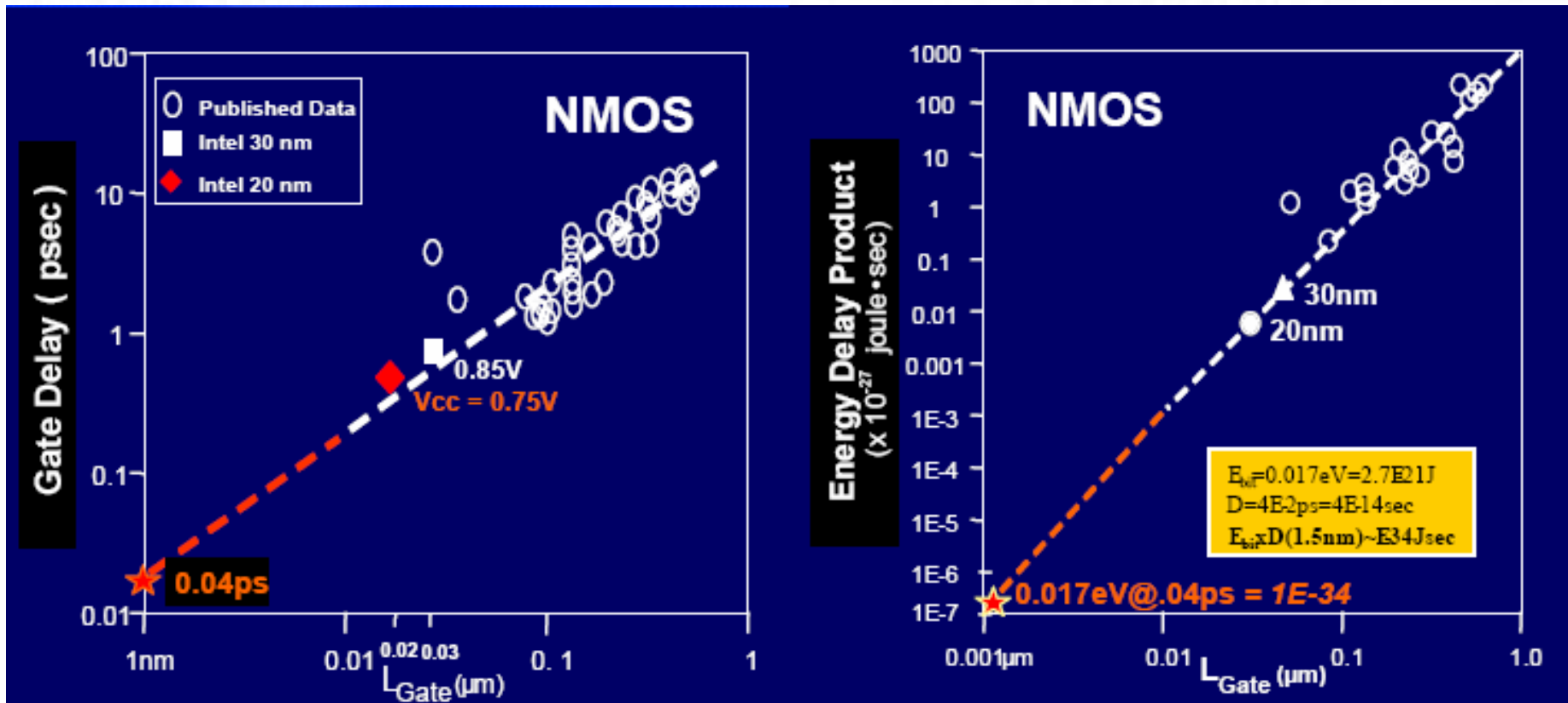
Oui mais....

Logique réversible

Hypothèses physiques

.....

Sources INTEL



Source : INTEL 2006

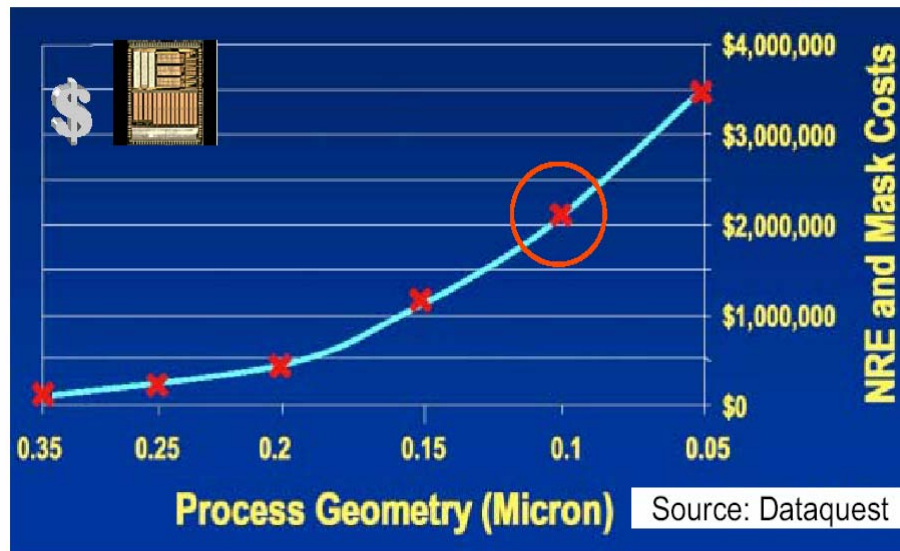


Limites pratiques

- Le coût des masques
- La puissance dissipée
- Le délai des interconnexions
- Augmentation des dispersions

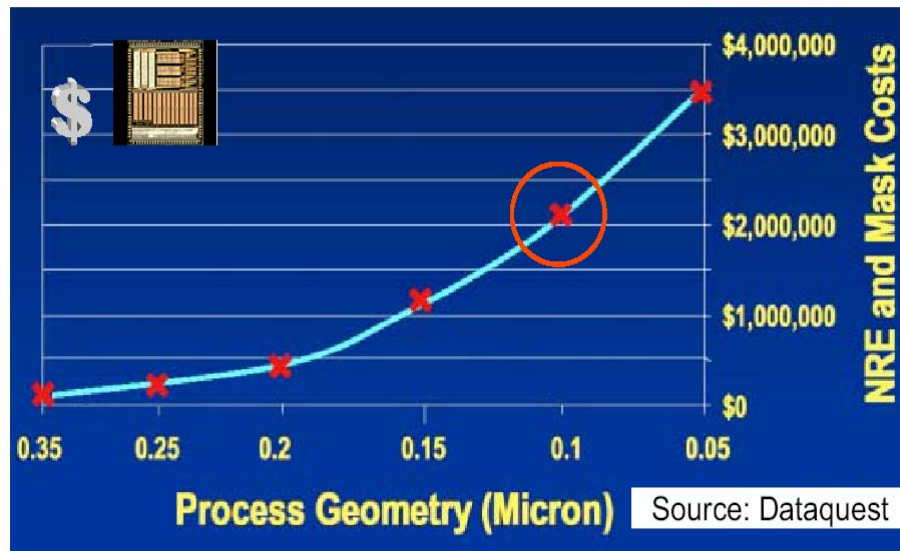
Limites pratiques

- Le coût des masques



Limites pratiques

- Le coût des masques

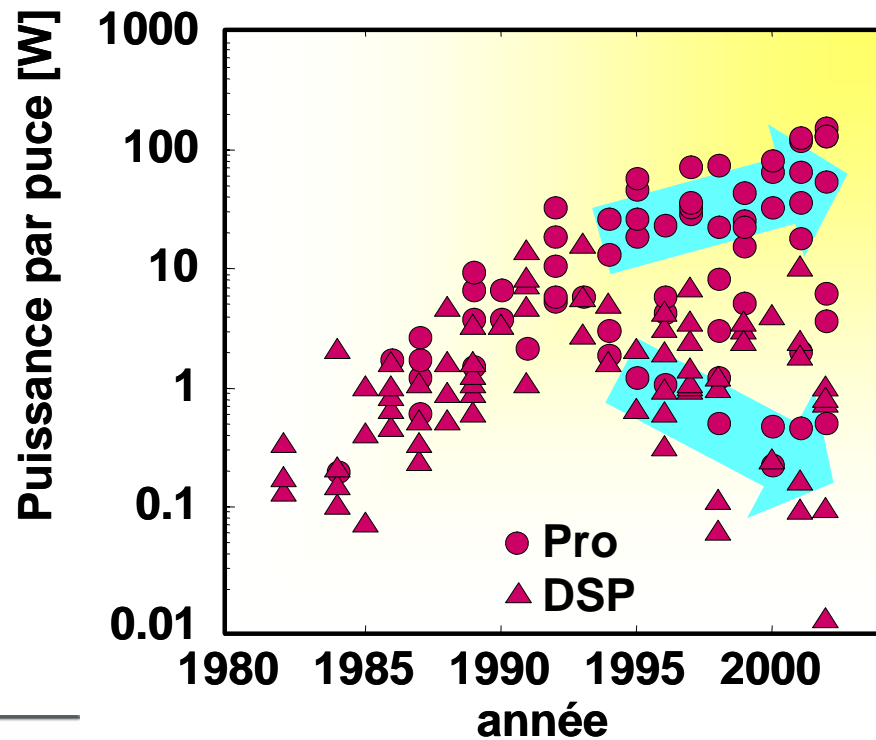


Solutions

- Se passer de la lithographie (nanotechnologies)
- Faire des circuits multi-usages en grande série (reconfigurables)

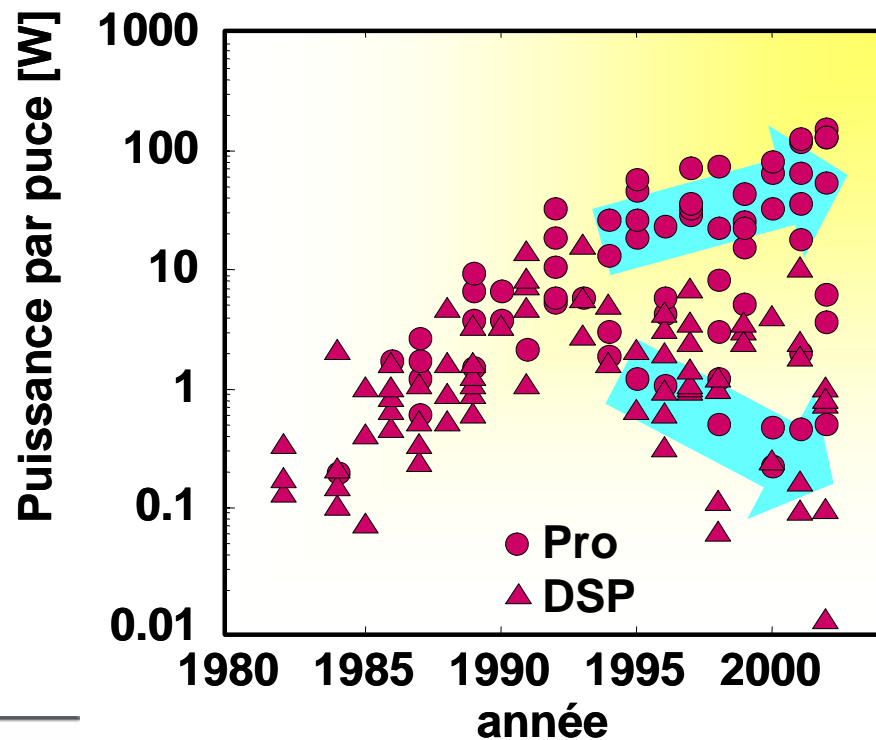
Limites pratiques

- La puissance dissipée



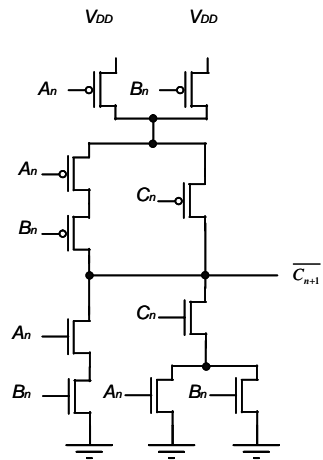
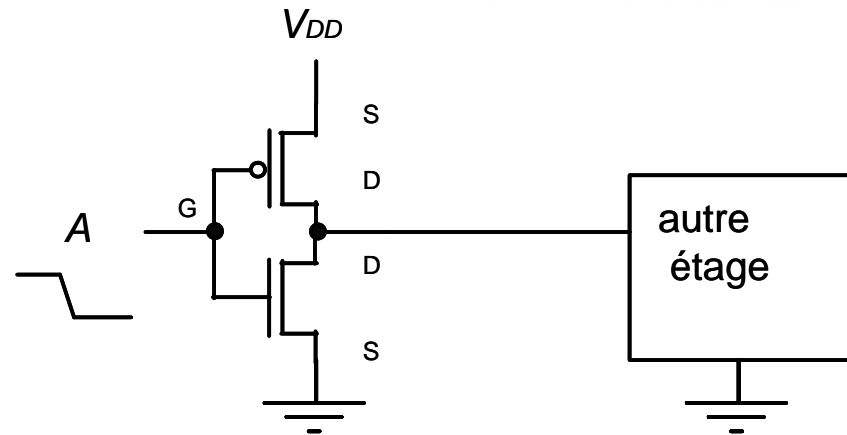
Limites pratiques

- La puissance dissipée



Solutions

- Nouveaux composants (SET, molécules...)
- Nouvelles architectures (parallélisme)



$$P = a f C V_{DD}^2 + I_0 V_{DD} \exp^{-\frac{V_T}{S}} + V_{DD} I_t$$



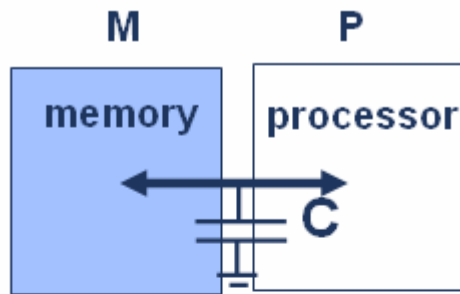
$$P = a f C V_{DD}^2 + I_0 V_{DD} \exp^{-\frac{V_T}{S}} + V_{DD} I_t$$

- Arrêt à la course à la fréquence
- Réduire au maximum la tension d'alimentation

Minimiser les distances d'échange

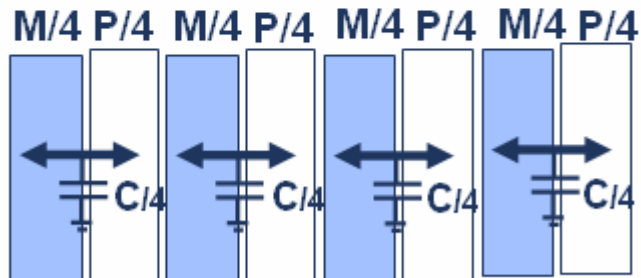
Contrôler les tensions de seuil

Limiter la miniaturisation



$$P_{mono} = CV^2f \text{ (Watt)}$$

Intérêt du parallélisme pour la consommation dynamique



$$4 (C/4)V^2(f/4) = P_{mono}/4$$

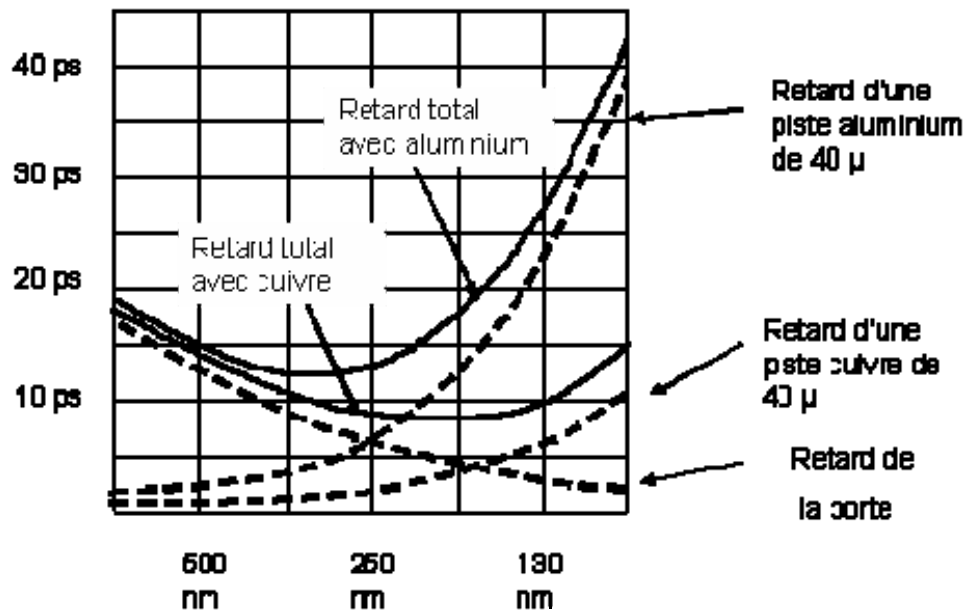
But if $f \sim V$

$$P// = P_{mono}/4^3$$

Attention pour la consommation statique

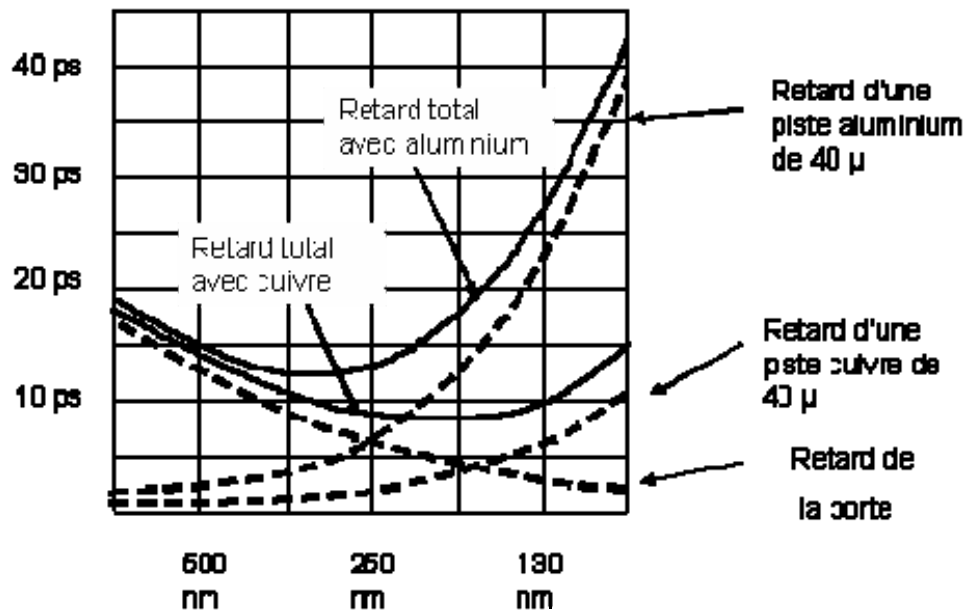
Limites pratiques

- Le délai des interconnexions



Limites pratiques

- Le délai des interconnexions

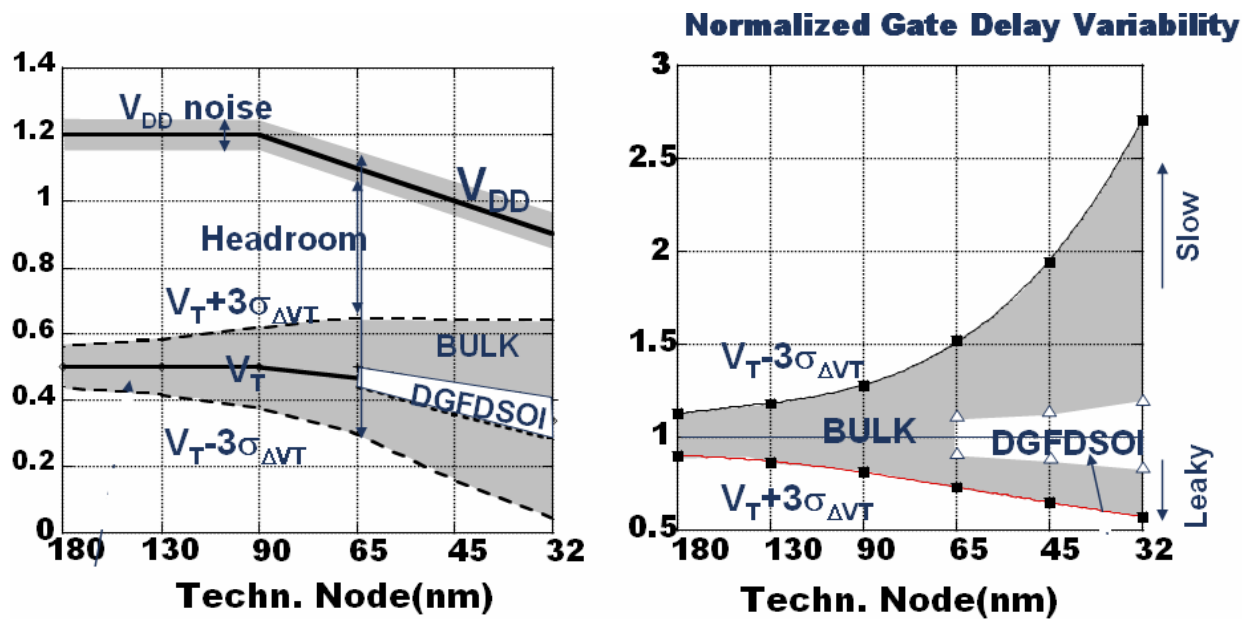


Solutions

- Low k et cuivre
- Liens optiques
- Architectures GAL et parallélisme

Limites pratiques

- Augmentation des dispersions

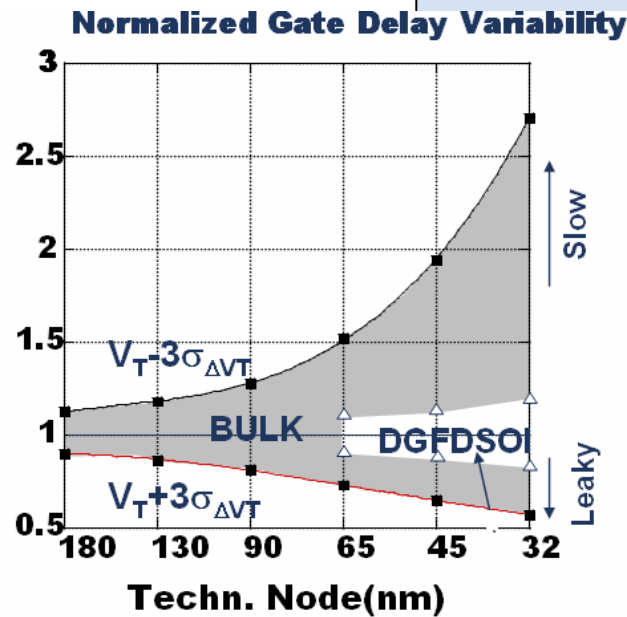
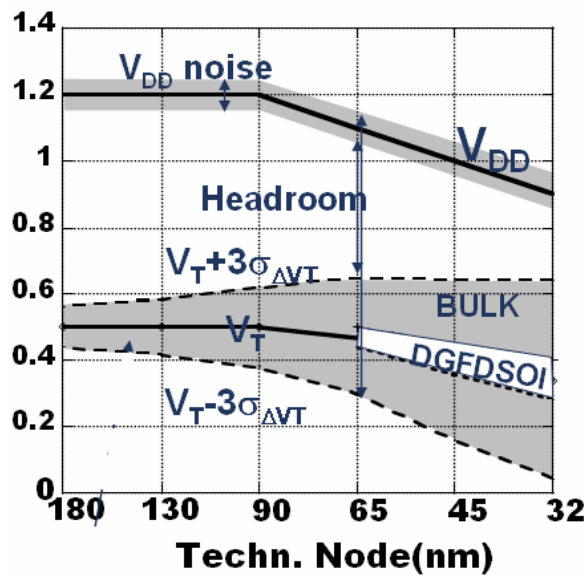


Limites pratiques

- Augmentation des dispersions

Solutions

- Architectures tolérantes
- Nouveaux composants





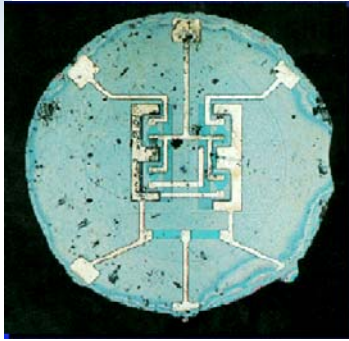
La loi de Moore mais....

- Limites théoriques et pratiques
- **Les nouvelles architectures**

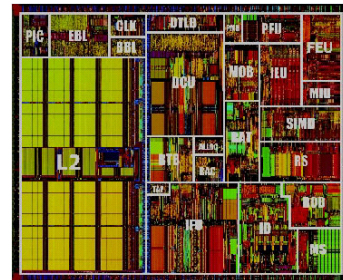
Un grand nombre de nouveaux dispositifs

- Quelles architectures pour les nanocomposants
- Tentative de conclusion

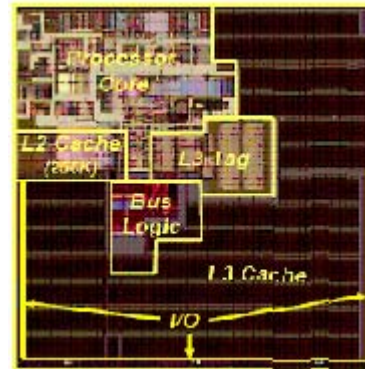
Evolution des architectures



Le premier circuit intégré



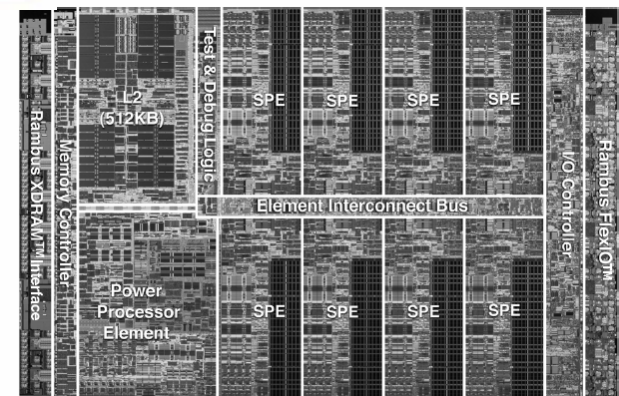
Le pentium 4



Le processeur Itanium

- Plus de transistors
- Plus de mémoire
- Parallélisme

Le processeur Cell



De nouvelles architectures

Ease of programming
Low power
Flexibility
High performance

Few applications
Pseudo static behavior

Dynamic behavior

1 algorithm

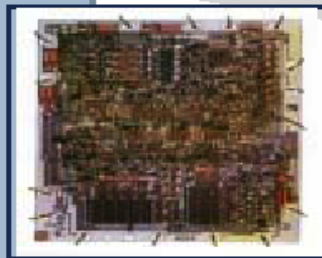
1 application



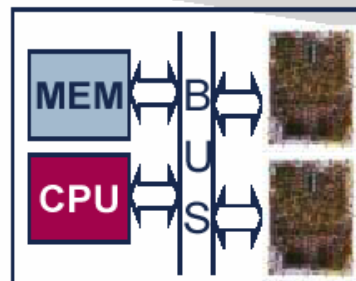
Middleware

Operating system

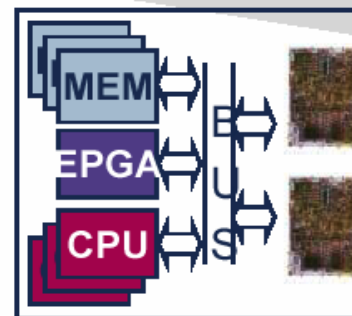
SW & HW multi-tasking



ASIC



SoC



Multi-processor
SoC platform

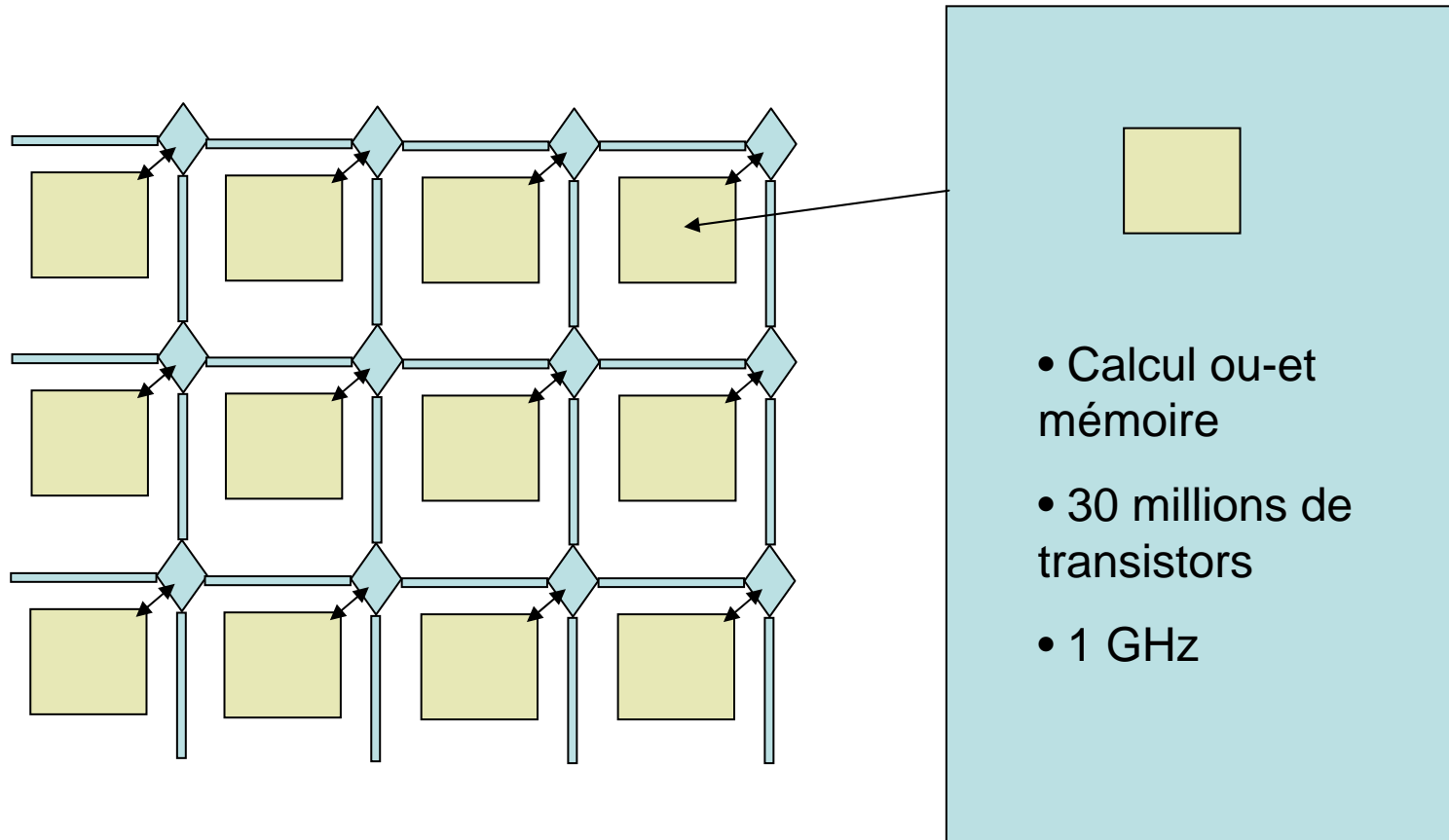
Re-configurable hardware
On-chip communication network

Memory hierarchy



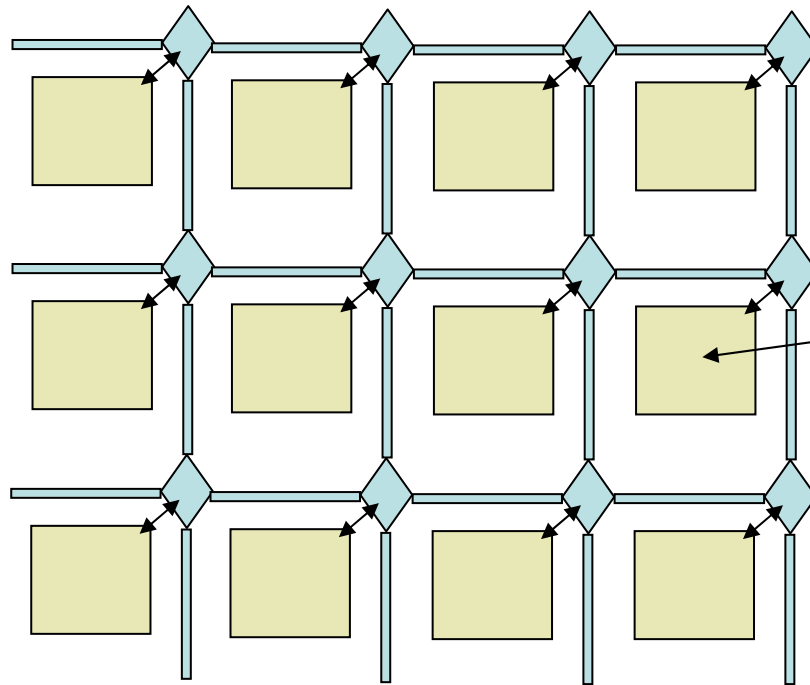
Flexible SoC platform

L'architecture du futur ?



Le Network On Chip de type GAL

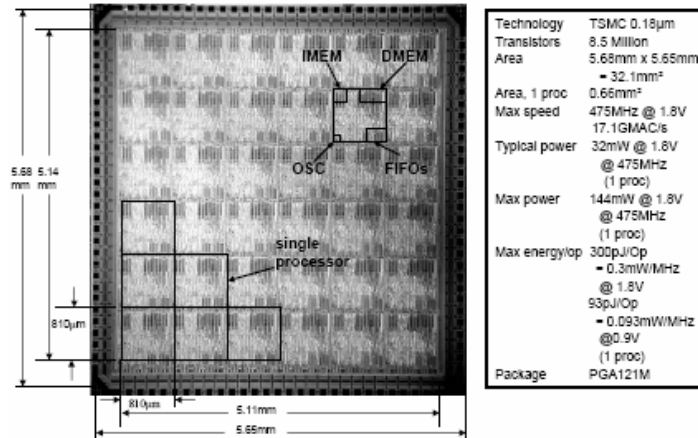
L'architecture du futur ?



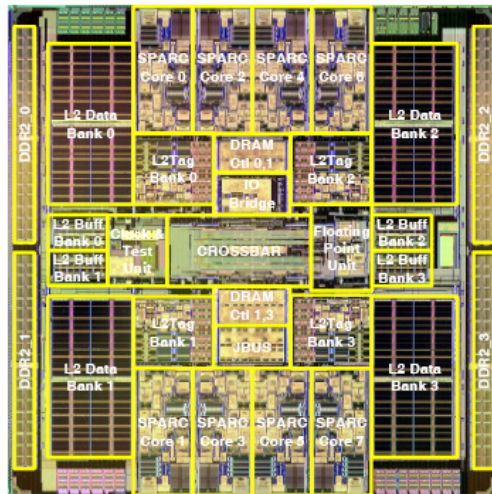
- Mémoire non volatile
- Pilotage dynamique des paramètres physiques (VDD, f, VT)

Le Network On Chip de type GAL

L'architecture du futur : quelques réalisations



Université de Californie :
réseau de 36 processeurs
ISSCC 2006



Features:

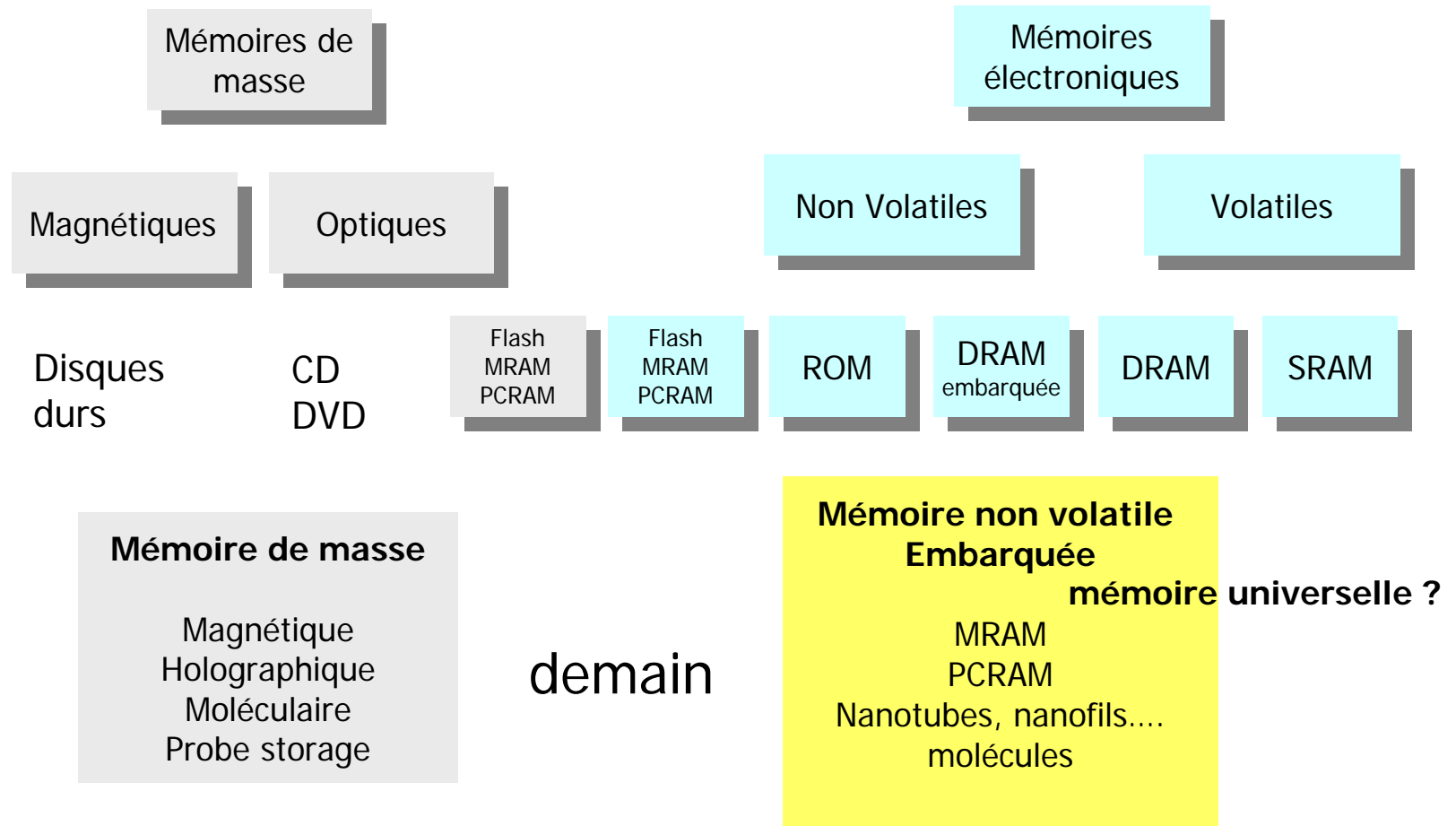
- Eight 64b Multithreaded SPARC Cores
- Shared 3MB L2 Cache
- 16KB ICache per Core
- 8KB DCache per Core
- Four 144b DDR-2 DRAM Interfaces (400 MT/s)
- 3.2GB/s JBUS I/O
- Crypto: Public Key (RSA)
- Extensive RAS

Technology:

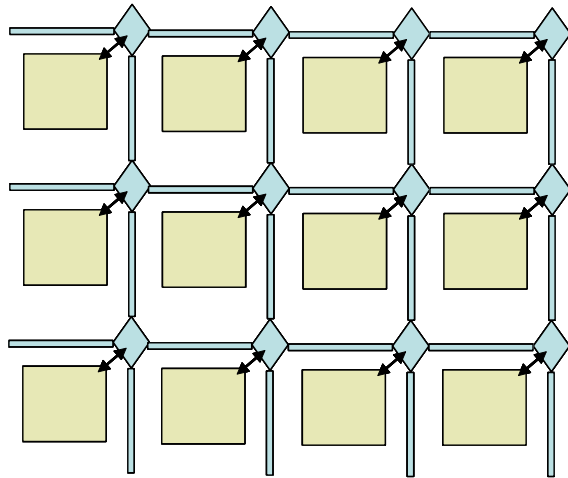
- 90nm CMOS Process
- 9LM Copper Interconnect
- Power: 63 Watts @ 1.2GHz
- Die Size: 378mm²
- 279M Transistors
- Package: Flip-chip ceramic LGA (1933 pins)

SUN : réseau de 8
processeurs
ISSCC2006

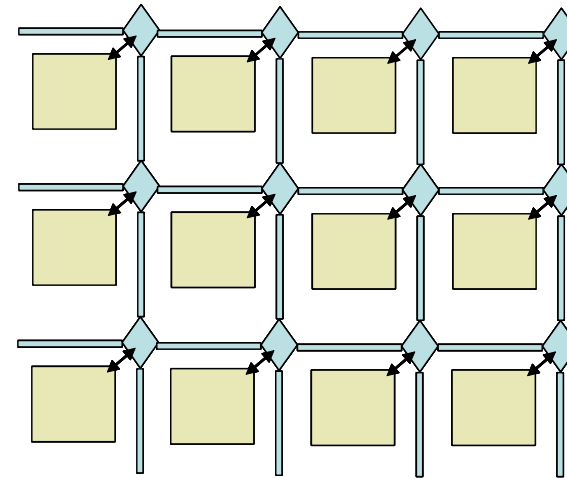
Les Mémoires ne sont pas en reste



Mémoires intelligentes et Processeur



Processeur de
type GAL



Mémoire
intelligente



- Portage des applications sur des architectures parallèles
- Des OS capables de piloter dynamiquement les paramètres physiques de fonctionnement
- Des architectures plus tolérantes

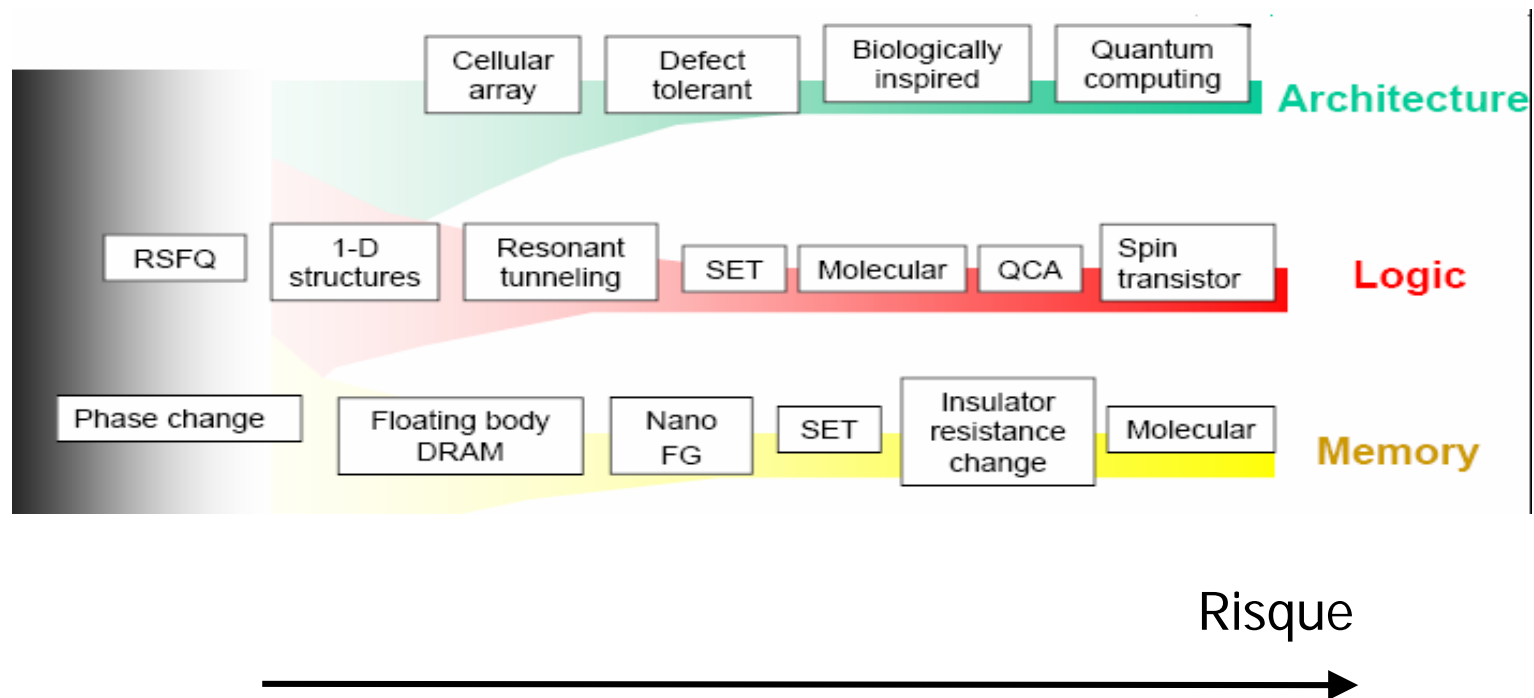


La loi de Moore mais....

- Limites théoriques et pratiques
- Les nouvelles architectures
- **Un grand nombre de nouveaux dispositifs**
- Tentative de conclusion

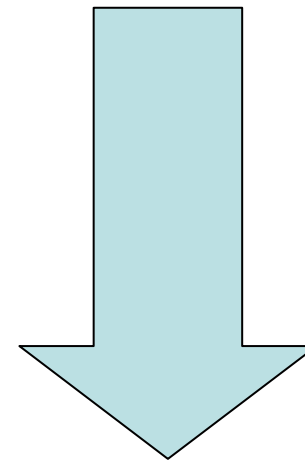
Vision de l'ITRS

Dispositifs émergents





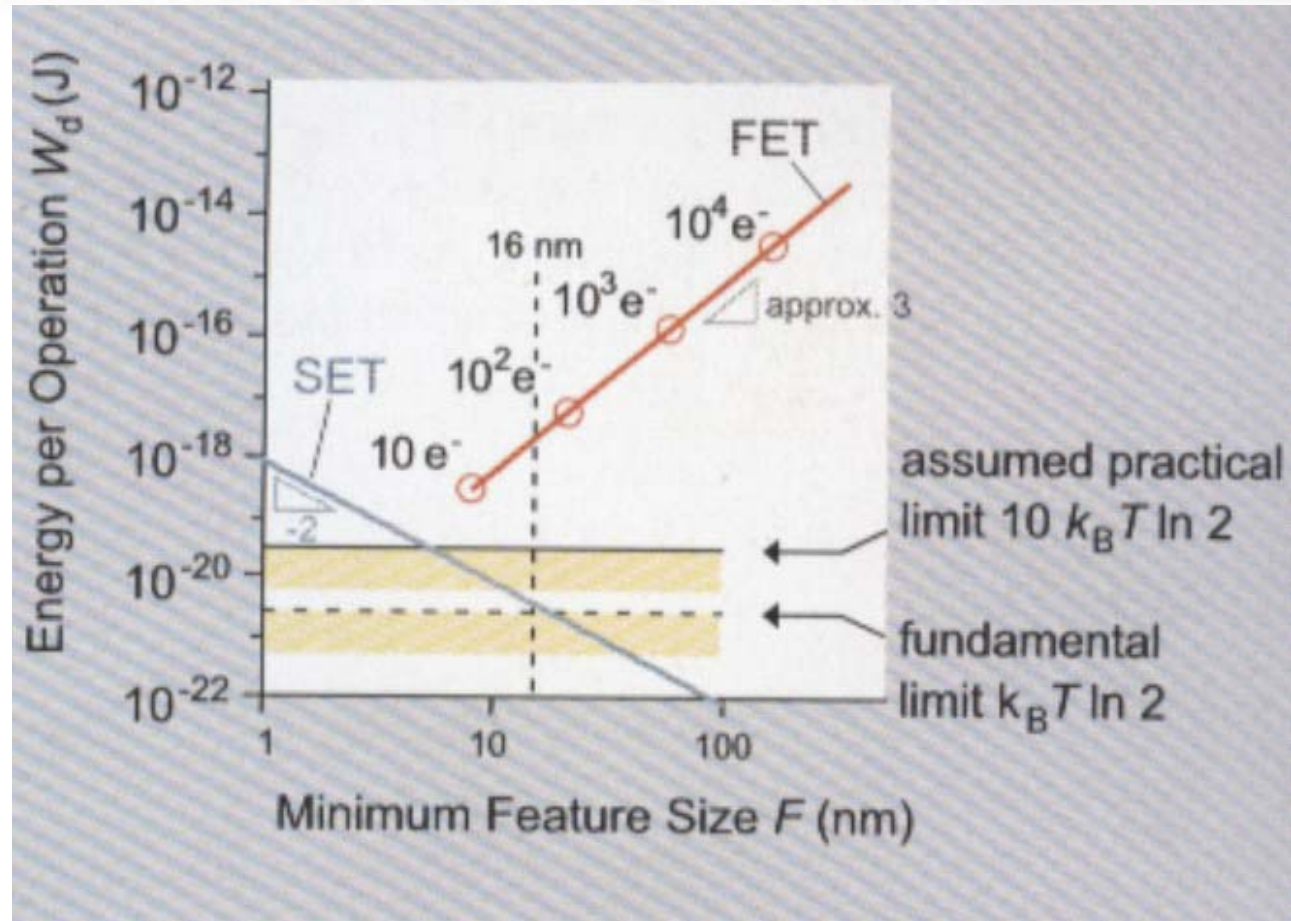
- Transistor à électron unique
- SpinFET
- Nanotubes de carbone et nanofils
- Nanocristaux
- Molécules et atomes



Taille du
dispositif

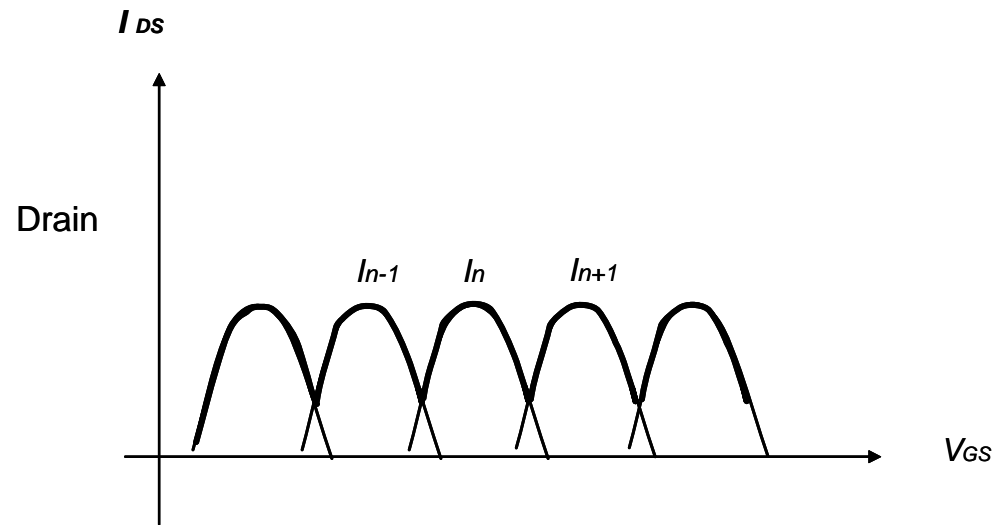
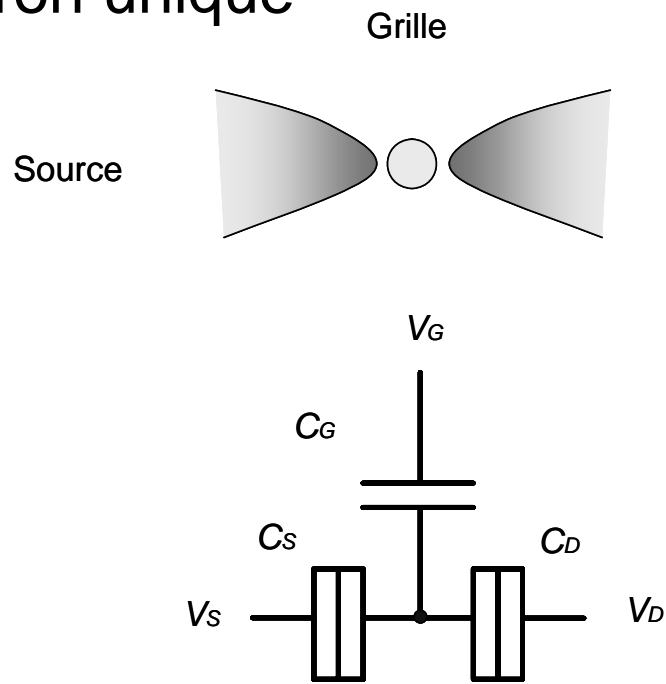


Le transistor à électron unique





Le transistor à électron unique





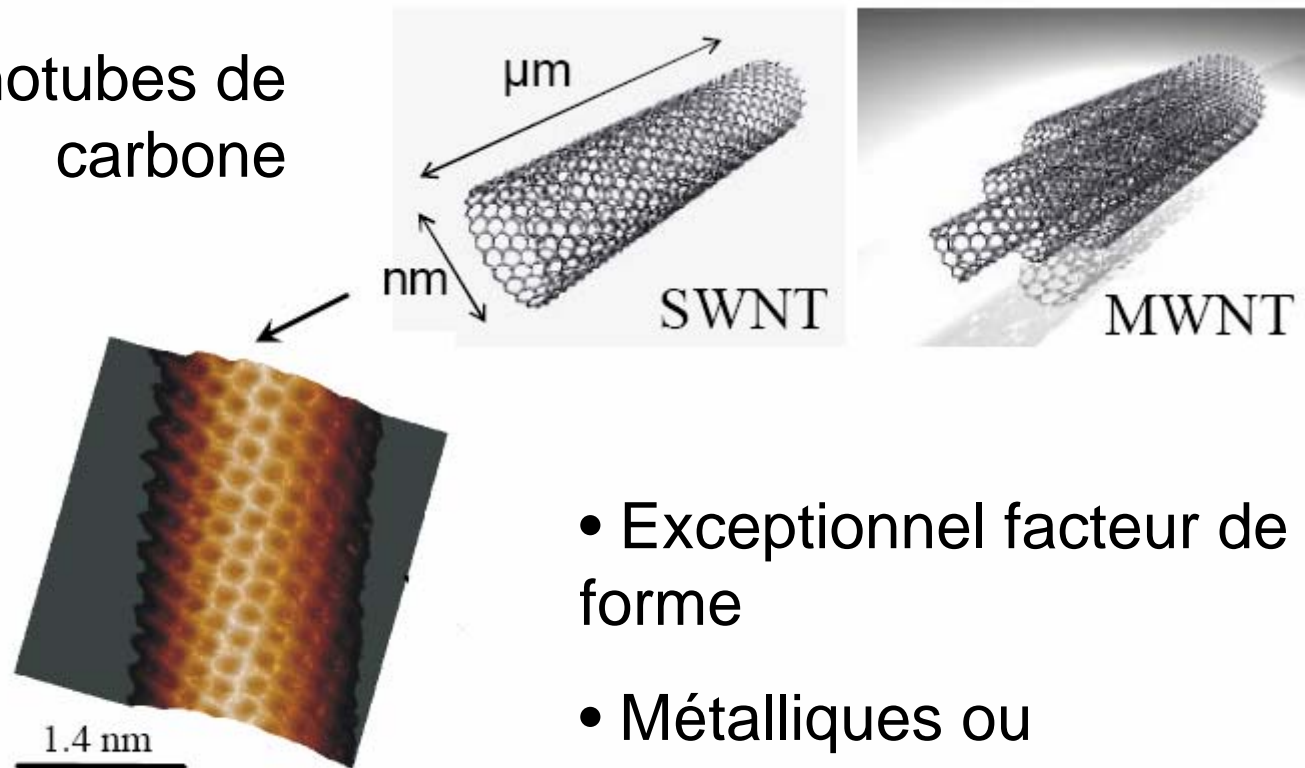
Le transistor à électron unique

- Fonctionnement à température ambiante ?
- Capacités des interconnexions

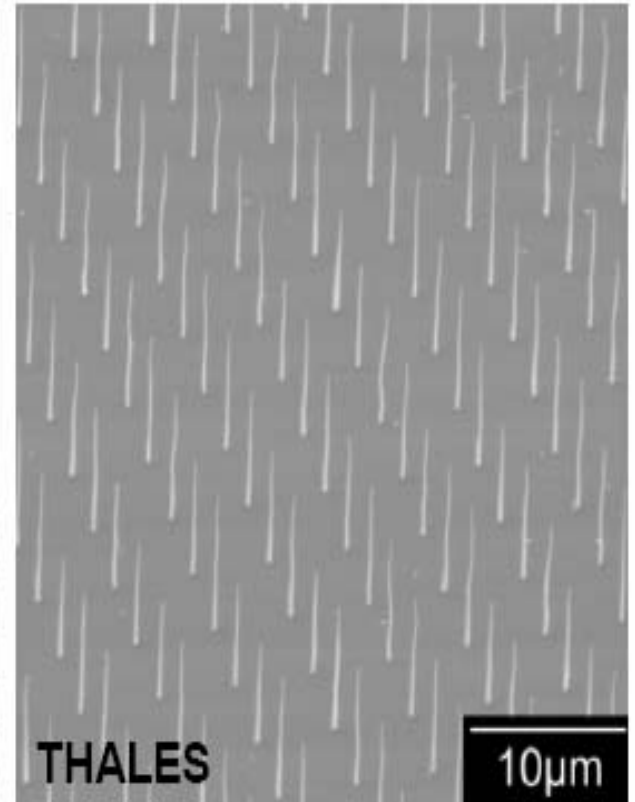
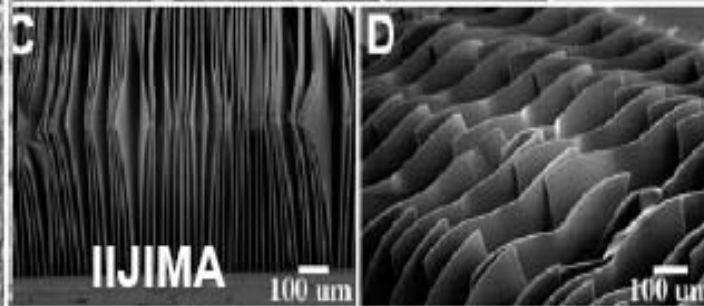
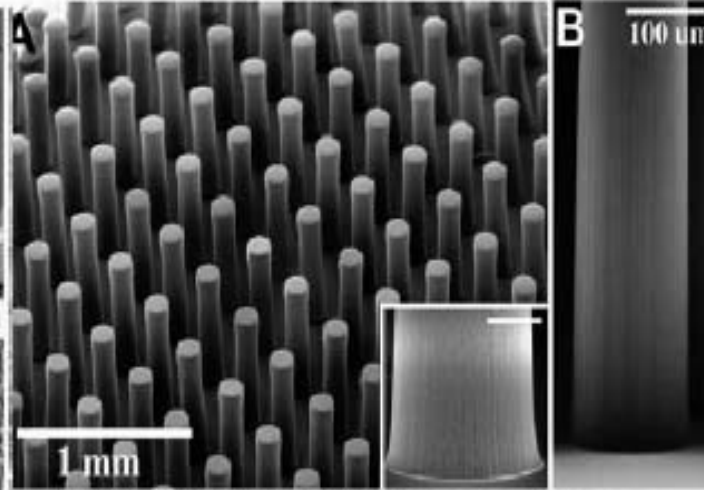
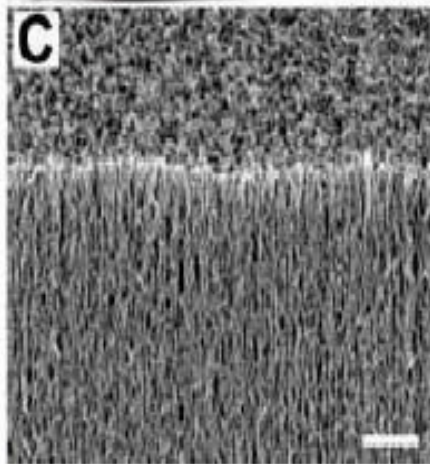
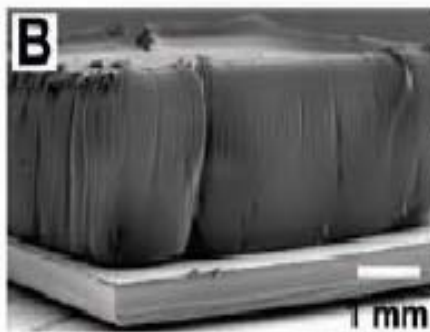
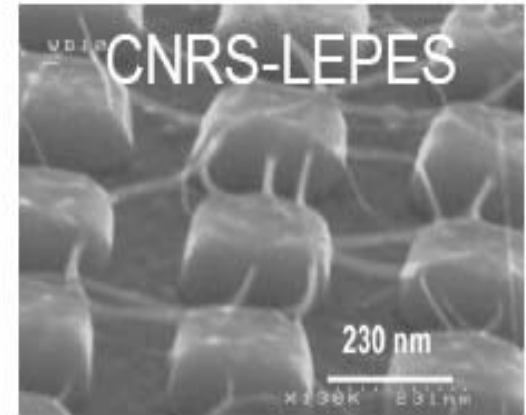
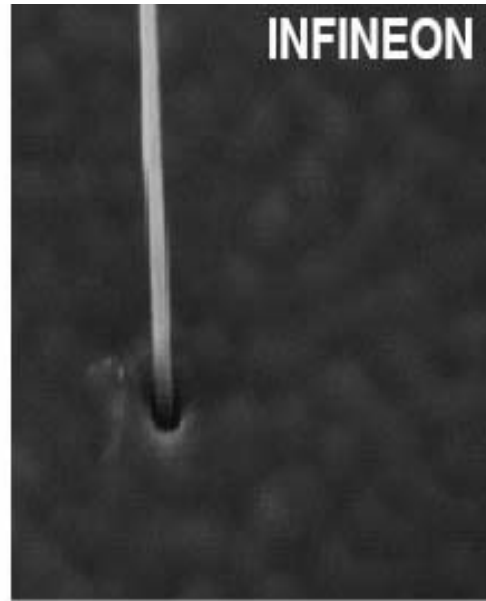
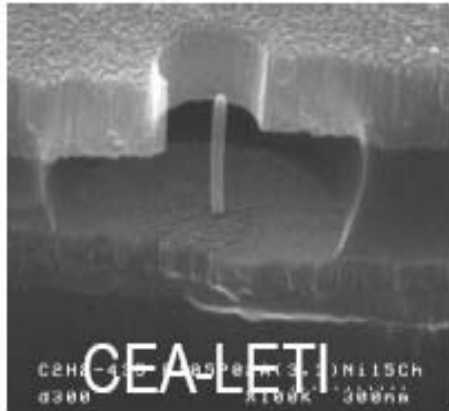
Miniaturisation extrême nécessaire (10 nm)



Les nanotubes de carbone

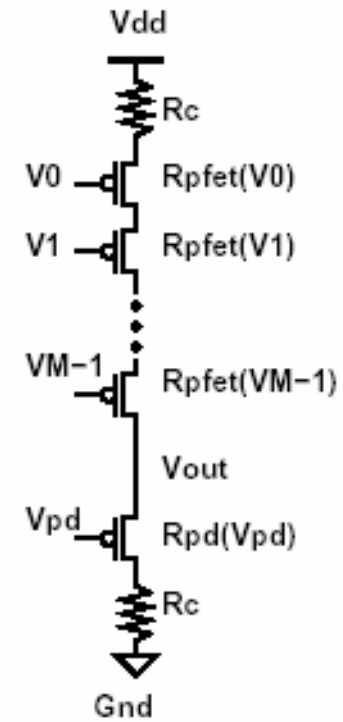
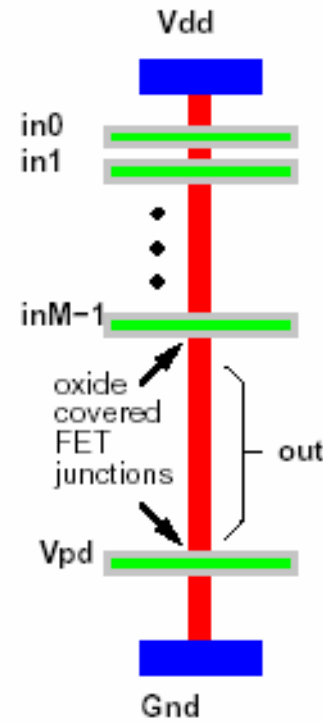
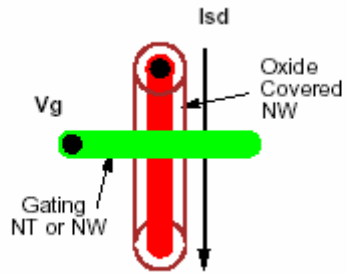


- Exceptionnel facteur de forme
- Métalliques ou semiconducteurs





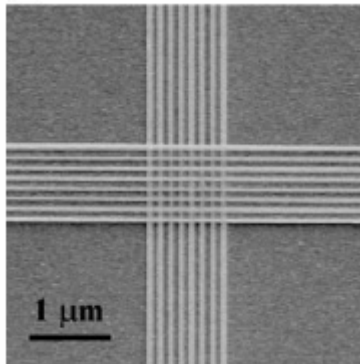
Transistors fet de type p



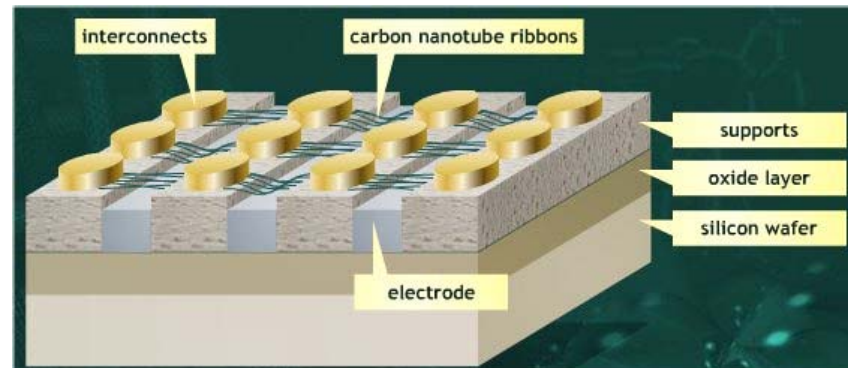
NOR

$$V_{out} = V_{dd} \left(\frac{R_{pd} + R_c}{R_c + \sum_{i=0}^{M-1} (R_{pfet}(V_i)) + R_{pd} + R_c} \right)$$

Quelques rares réalisations



HP 2003



Nantero 2005

Applications possibles des nanotubes de carbone

Transistors améliorés

Connexions à haute densité de courant

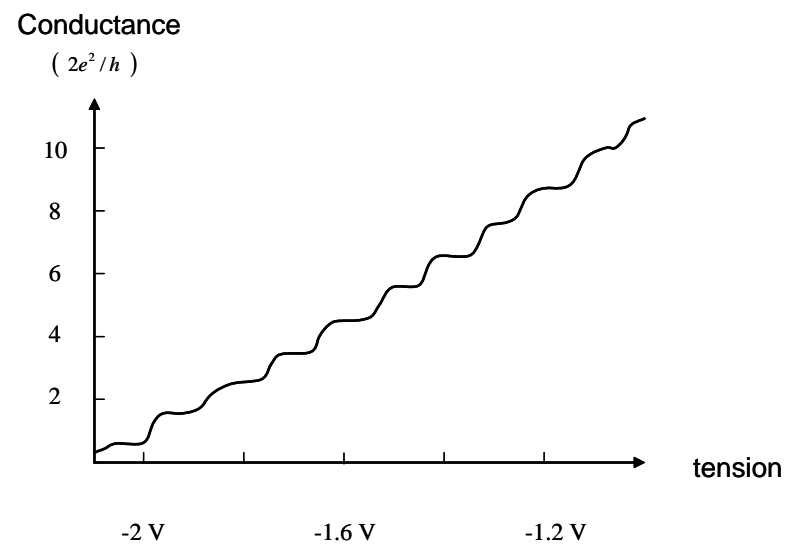
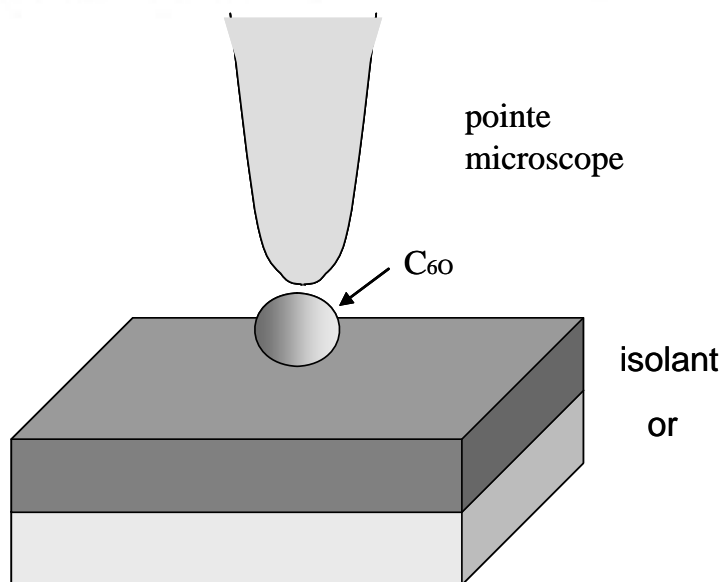
Emetteurs d'électrons

Nano-interrupteurs

.....

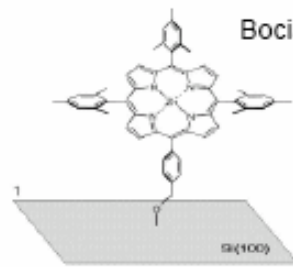
Mais contacts et sélection

Propriétés électriques de la molécule unique

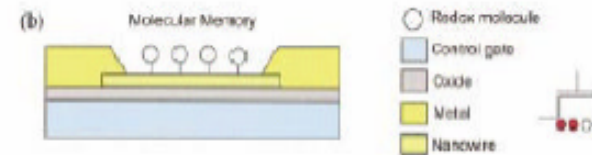


Applications des molécules

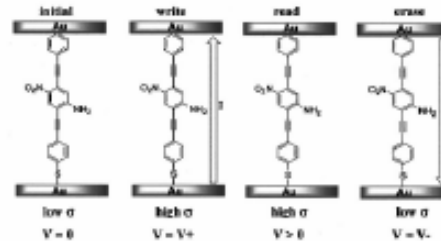
Stockage
multivalué



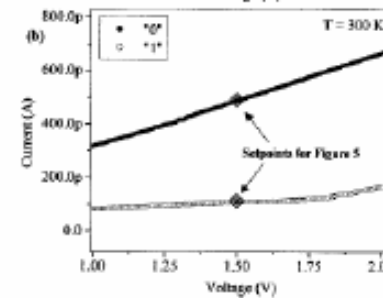
meyyapaan et al APL 2004



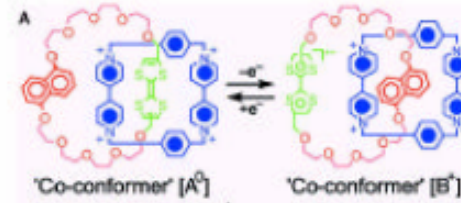
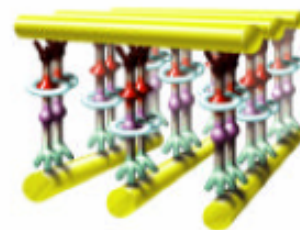
Résistance
négative



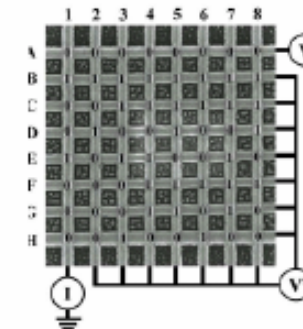
Chen et al APL 2001



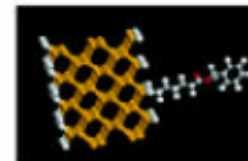
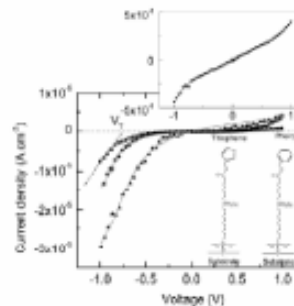
Interrupteurs



Williams, Stoddart, Heath
UCLA/HP



Diodes



Vuillaume et al Nanolett 2003

le 30 09 2005



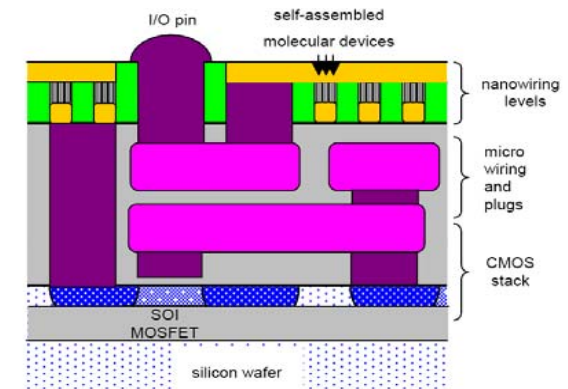
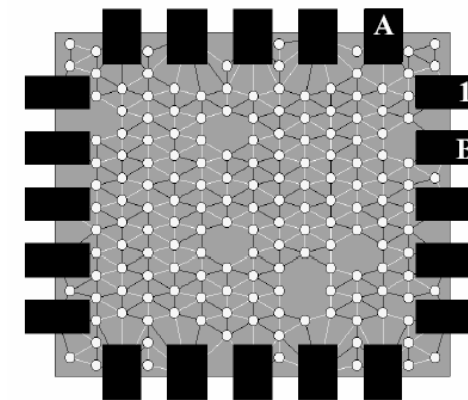
La loi de Moore mais....

- Limites théoriques et pratiques
- Les nouvelles architectures
- Un grand nombre de nouveaux dispositifs
- **Quelles architectures pour les nanocomposants**

Tentative de conclusion

Les trois approches

- On ne change rien
le transistor est remplacé
- Tout est à refaire
calcul quantique
réseau de neurones
réseau automates cellulaires
- Approche « nano inside »



Quelles architectures pour les nouveaux composants

- Se passer le plus possible de la lithographie en tirant parti de l'auto-assemblage
- Maîtriser la consommation
- Tolérer les défauts et les dispersions
- Se comparer aux performances de la CMOS actualisée

On se contente de remplacer le transistor

- SET
- SpinFET
- CNTFET
- Nanocristaux

Comparaison nécessaire avec un MOS avancé (22 nm) et à température ambiante

Aujourd'hui aucun candidat n'a prouvé sa supériorité et les « variantes » du MOS sont les plus probables

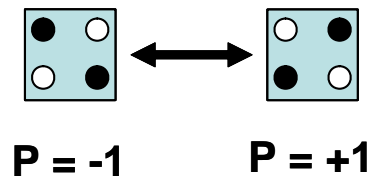
De nombreuses idées associées :

- Calculer au niveau des atomes ou des molécules
- Pousser le parallélisme à l'extrême
- Repousser les limites par la logique réversible

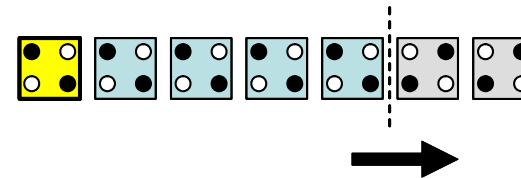
Des difficultés considérables de mise en œuvre pratique et une utilisation dans l'avenir peut-être limitée à quelques algorithmes particuliers.

Les réseaux d'automates cellulaires

Semble résoudre les difficultés des interconnexions



- electrostatic
- magnetic
- etc.



En pratique, il faut ajouter un système de distribution d'horloge

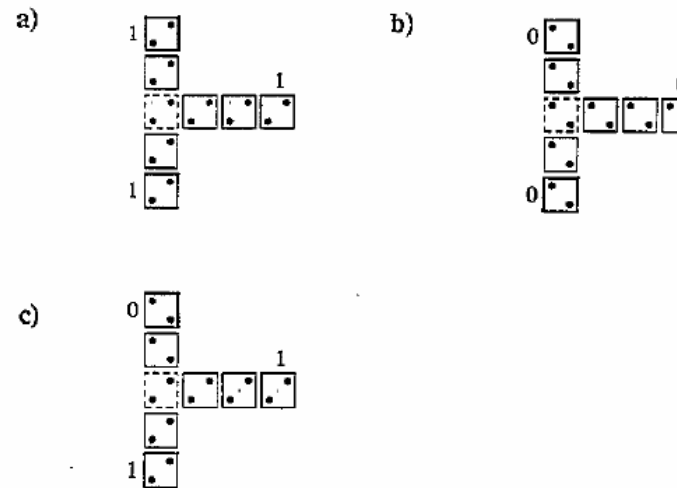


Figure 7. An OR gate. The cells in darker squares are fixed to the input states. The cell in the dashed square is biased slightly toward the '1' state.

Les réseaux de neurones artificiels

Semble résoudre le problème de la tolérance aux fautes

- Peu de réalisations « hardware »
- Quelques essais pour réaliser des fonctions logiques tolérantes

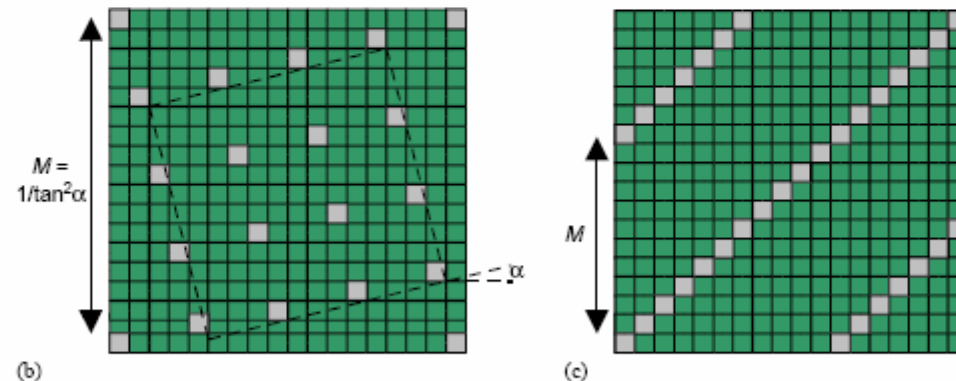
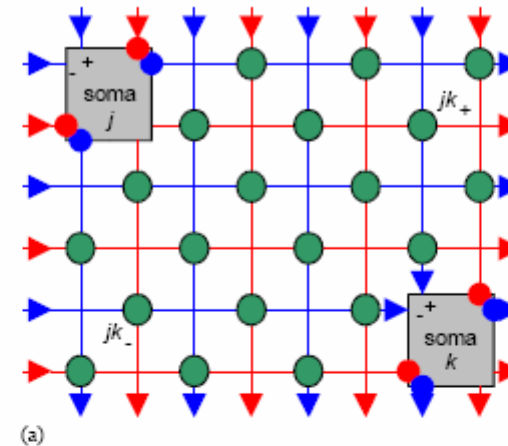
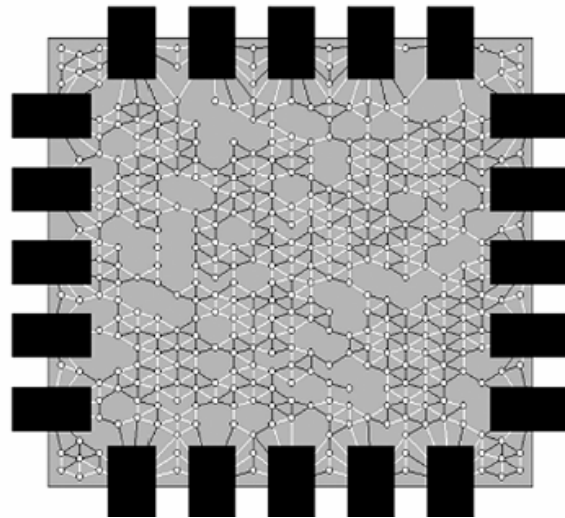
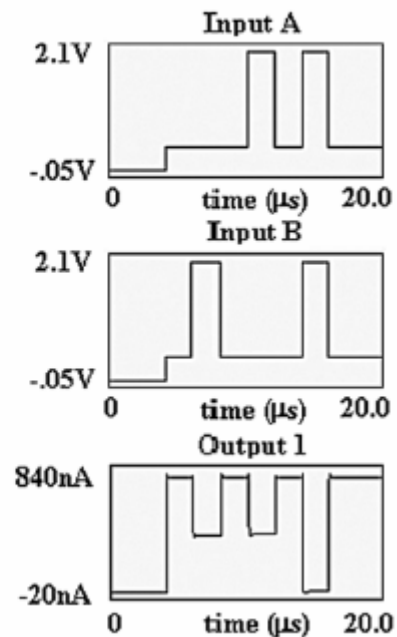
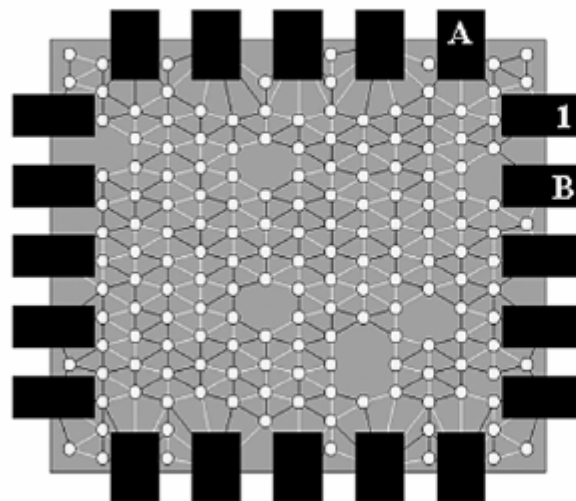


Fig. 1. CrossNets: (a) general structure, and two major species: (b) InBar and (c) FlossBar [1,9,10,5]. Red lines are axonic, and blue lines are dendritic nanowires. Gray squares show interfaces between nanowires and CMOS-based cell bodies (somas). Signs show the dendritic/axonic input/output polarities. Green circles represent latching switches; marked are the two switches connecting cells j and k . Bold red and blue points are open-circuit terminations of the wires, that do not allow direct interactions of the somas and limit cell connectivity. In recurrent CrossNets, similar feedback connections are added.

La « nanocell » ou comment fonctionnaliser le désordre

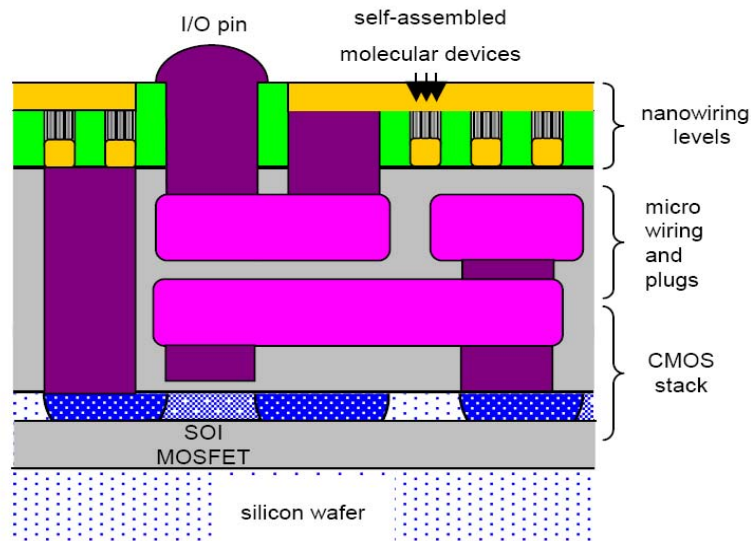


Un réseau désordonné de molécules et de nanoparticules d'or



James Tour,
Summer M
Husband, 2002

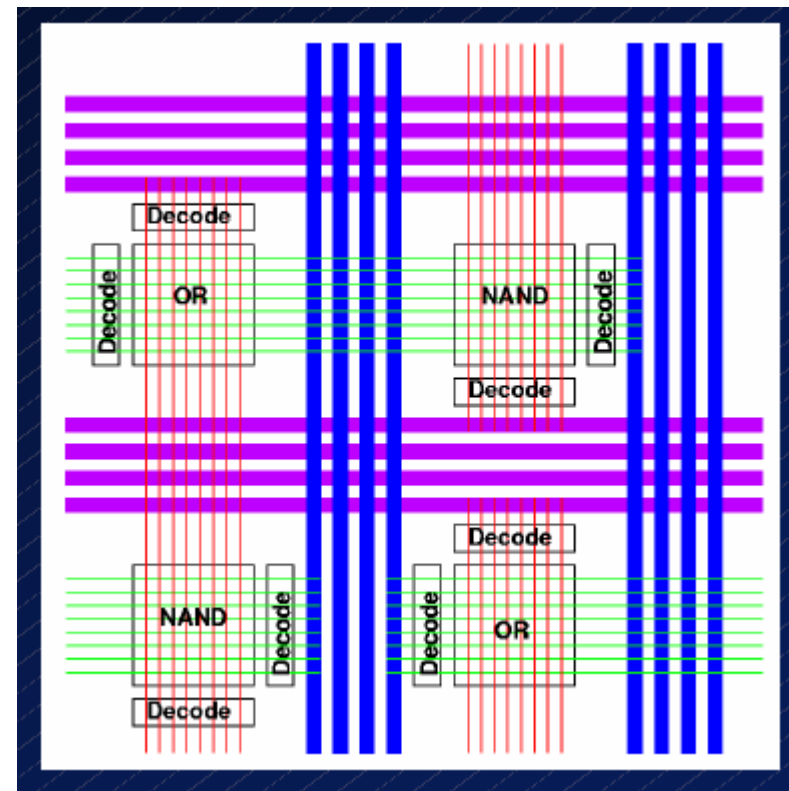
L'approche « nano inside »



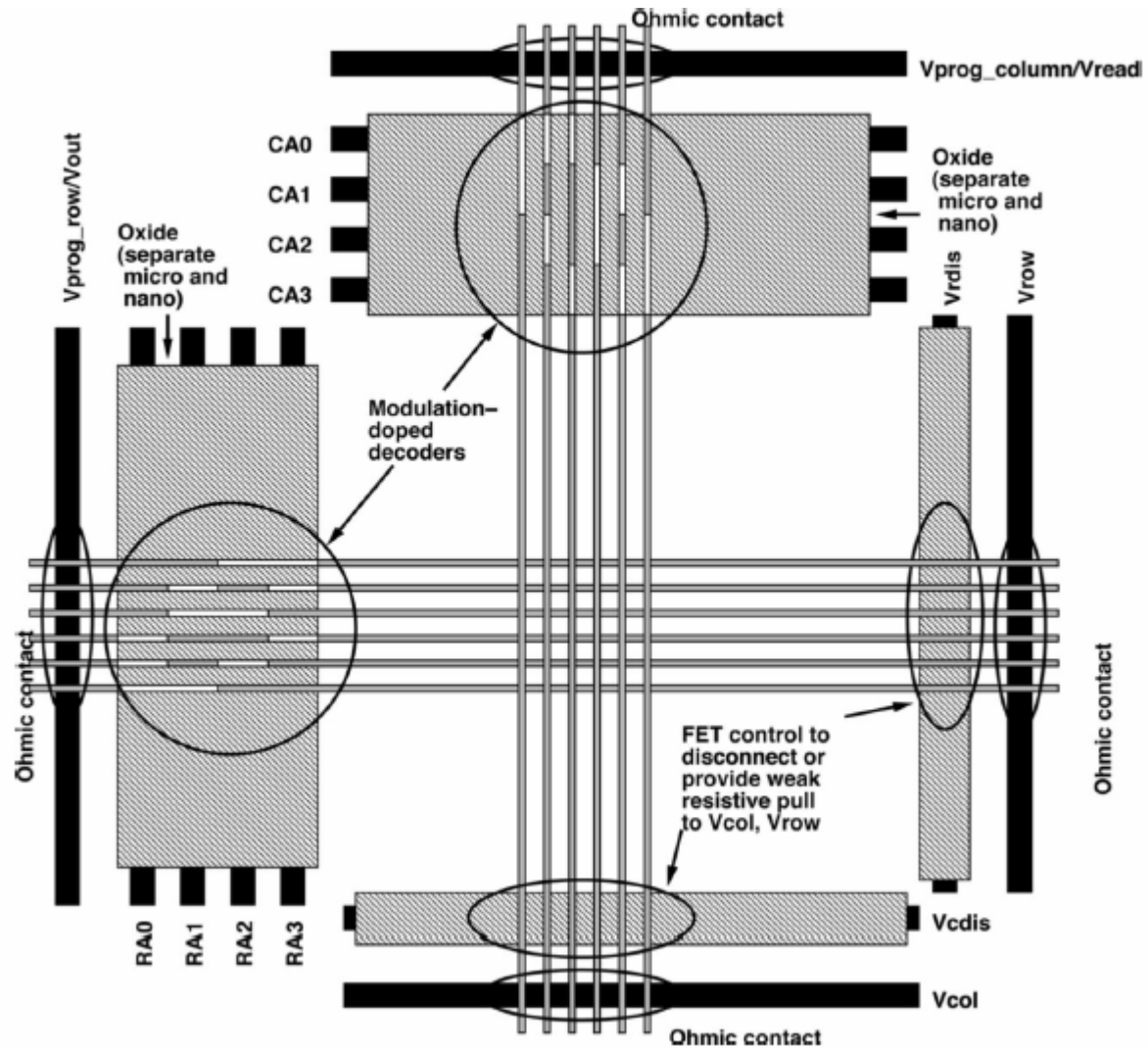
- Blocs mémoire
- Blocs logiques programmables
- Coprocesseurs d'interface avec environnement

A Déhon
2002

S Likharev
2003



L'approche « nano inside », bloc mémoire

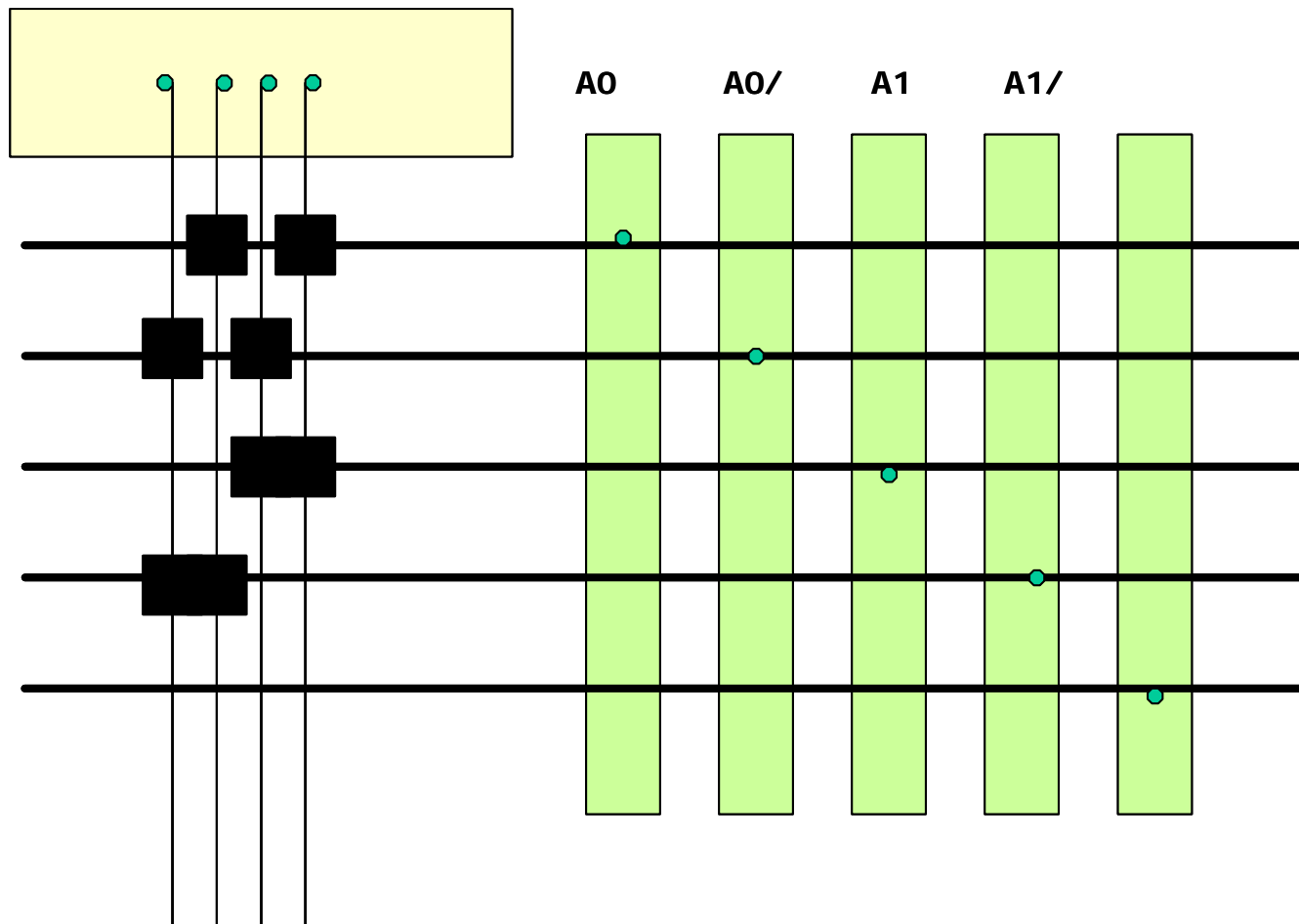


A Déhon 2003

Réduire le nombre
de fils d'adressage

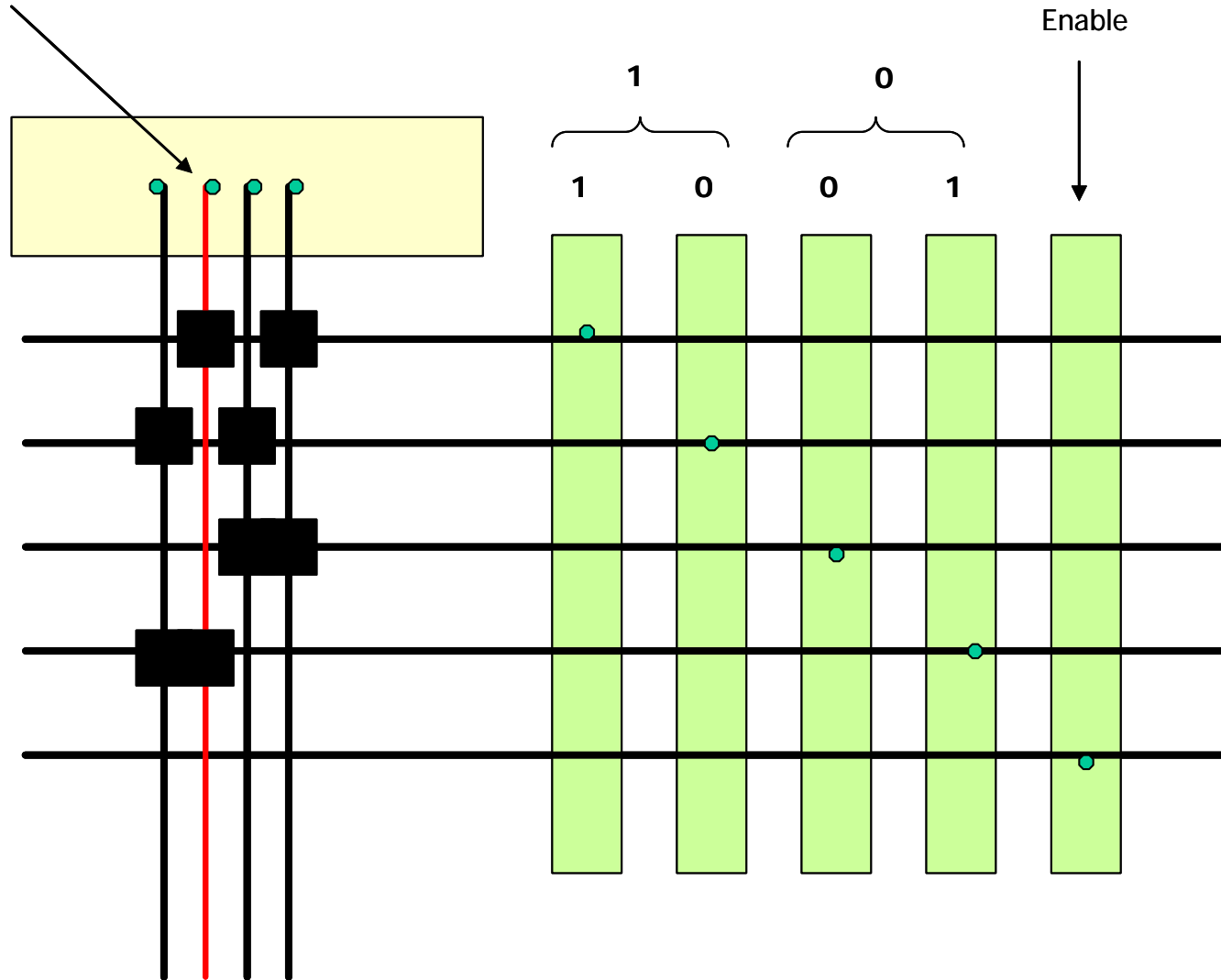
Tolérer les défauts

Du nano au micro



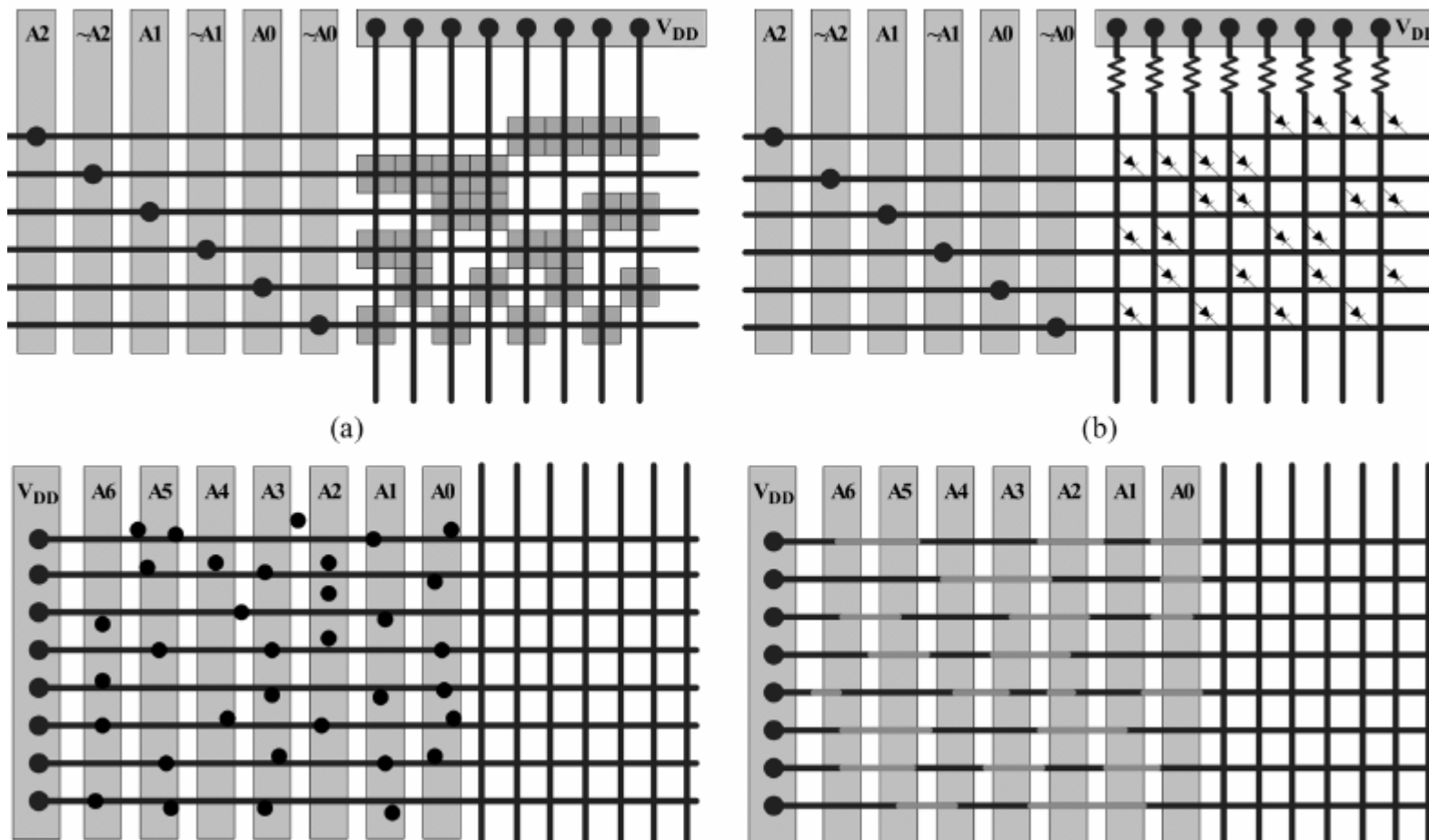
Du nano au micro

Ce fil conduit

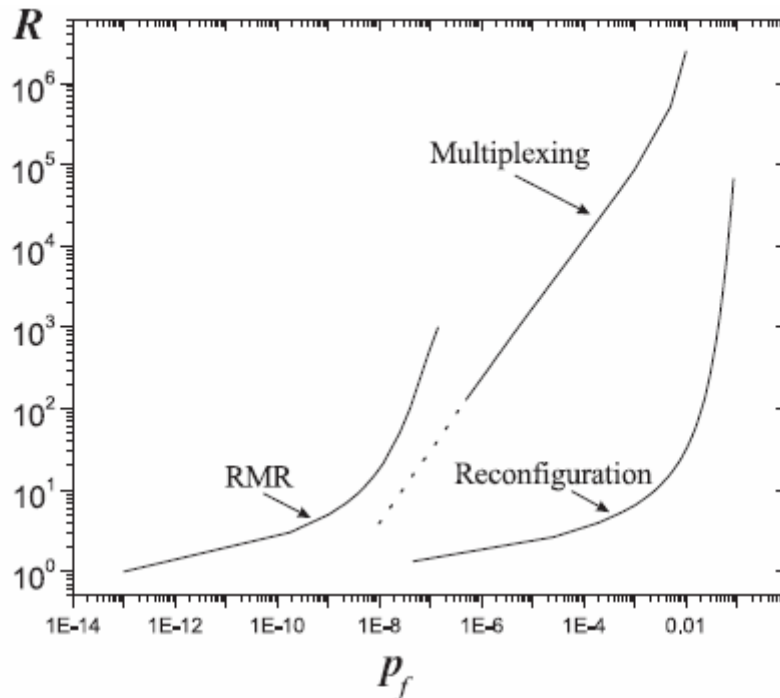


L'approche « nano inside », le problème de l'interface

Démultiplexage déterministe ou aléatoire



Les architectures tolérantes aux fautes



Forshaw 2001

Intérêt du reconfigurable

Figure 6. These curves compare three different fault-tolerant strategies (RMR, von Neumann multiplexing and a reconfigurable computer technique), applied to a hypothetical chip with 10^{12} devices (perhaps the ultimate number for a 1 cm^2 chip). The curves show the necessary level of redundancy R , as a function of the failure rate per device p_f , which ensures that the whole chip will work with a 90% probability. Starting points are at $R = 1$ (RMR), $R = 100$ (multiplexing) and $R = 1.33$ (reconfiguration). Results for the multiplexing are extrapolated for small R ($R < 100$) and presented with a broken curve, as the von Neumann's formula is not accurate in this region.

Application aux mémoires

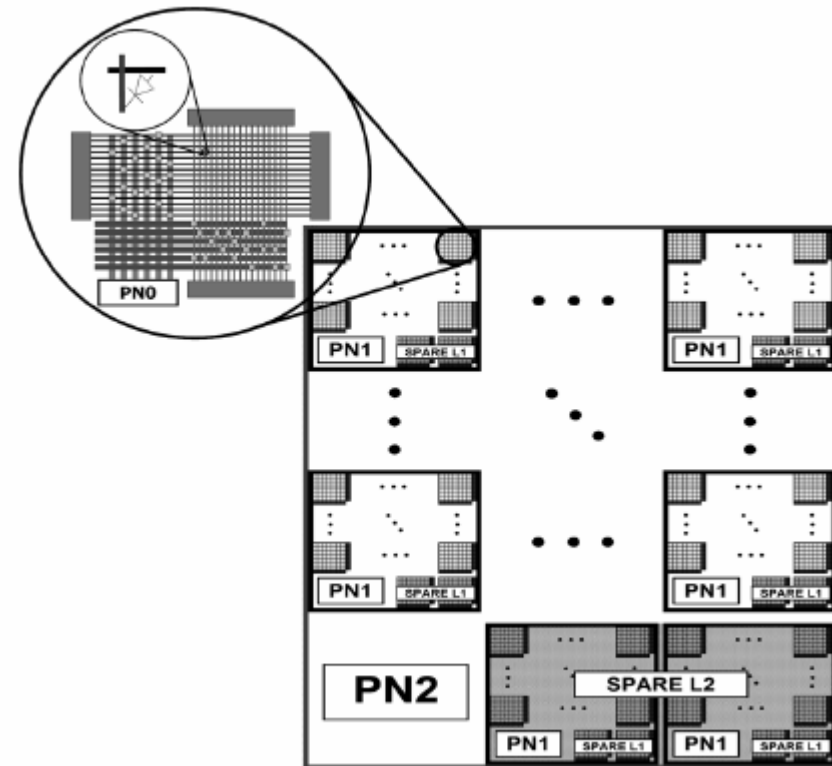
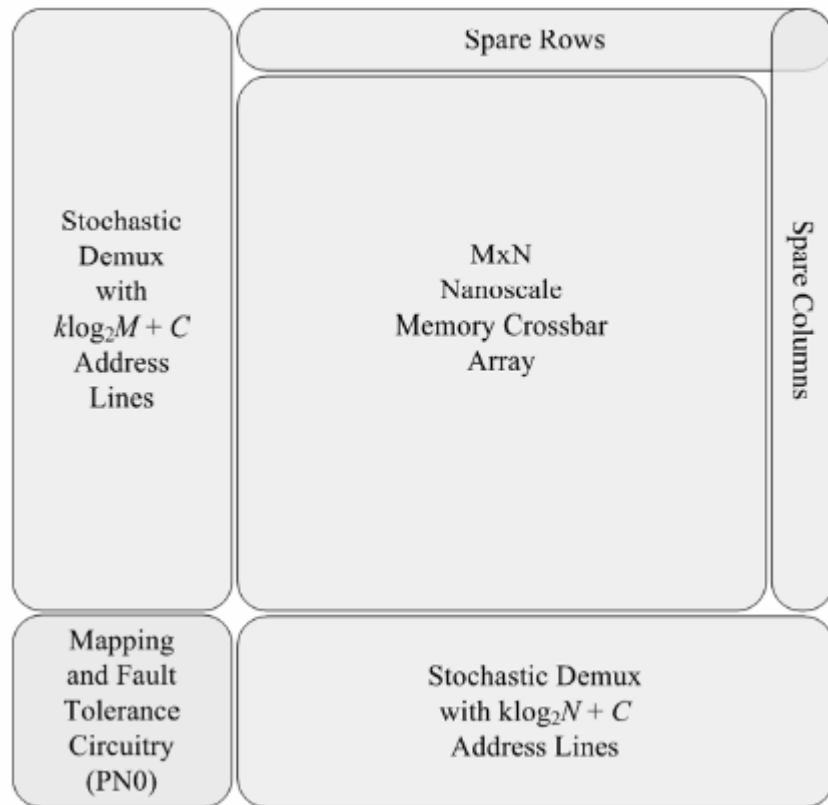
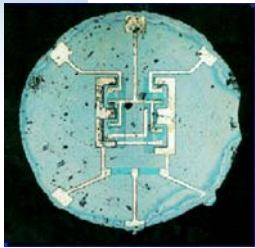
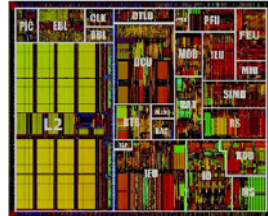


Fig. 3. Hierarchical memory architecture with spare devices and processing

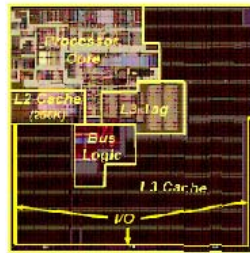
Evolution des architectures



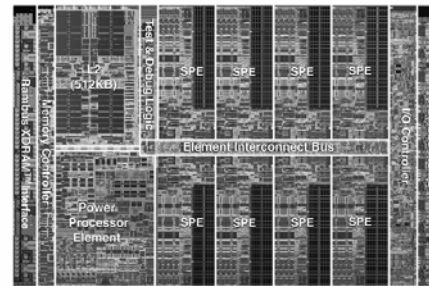
Le premier circuit intégré



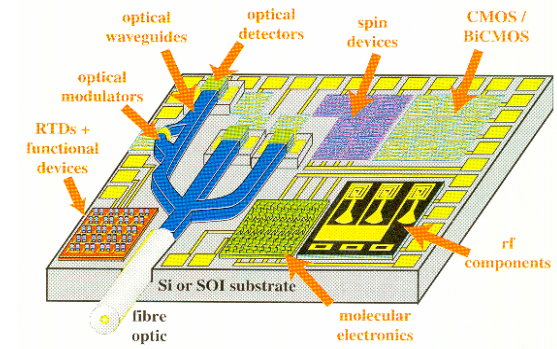
Le pentium 4



Le processeur Itanium



Le processeur Cell



L'approche nano inside

Parallélisme

Hétérogénéité

Mémoire embarquée



La loi de Moore mais....

- Limites théoriques et pratiques
- Les nouvelles architectures
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- **Tentative de conclusion**

- A moyen terme, introduction du parallélisme pour les applications demandant de la puissance de calcul et nécessité du développement des outils logiciels associés

- A moyen terme, introduction du parallélisme pour les applications demandant de la puissance de calcul et nécessité du développement des outils logiciels associés
- A plus long terme, intérêt de l'approche « nano inside » en particulier pour la fonction mémoire et peut-être pour la fonction logique programmable

- A moyen terme, introduction du parallélisme pour les applications demandant de la puissance de calcul et nécessité du développement des outils logiciels associés
- A plus long terme, intérêt de l'approche « nano inside » en particulier pour la fonction mémoire et peut-être pour la fonction logique programmable
- Dans tous les cas (MOSFET ou autres), importance de la gestion des fautes et des dispersions