

micro et nanoélectronique  
microsystèmes  
intelligence ambiante  
biologie et santé chaîne de l'image



2008

## Nano-électronique

## Impact sur les architectures et le logiciel

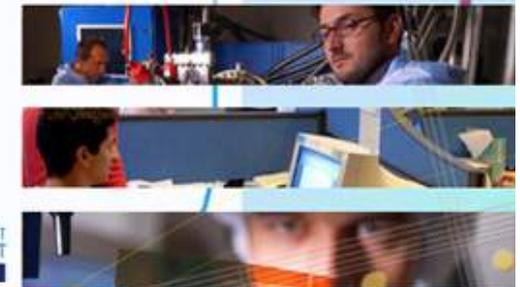
Hervé Fanet

leti

MINATEC®

INSTITUT  
CARNOT  
CEA LETI

cea



- 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

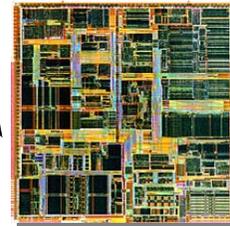
# Les deux secrets de la micro-électronique

## 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 !



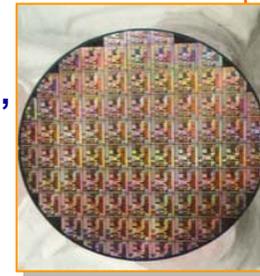
1 lot de fabrication à ST Crolles : = 100 milliards de transistors !!!

## 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

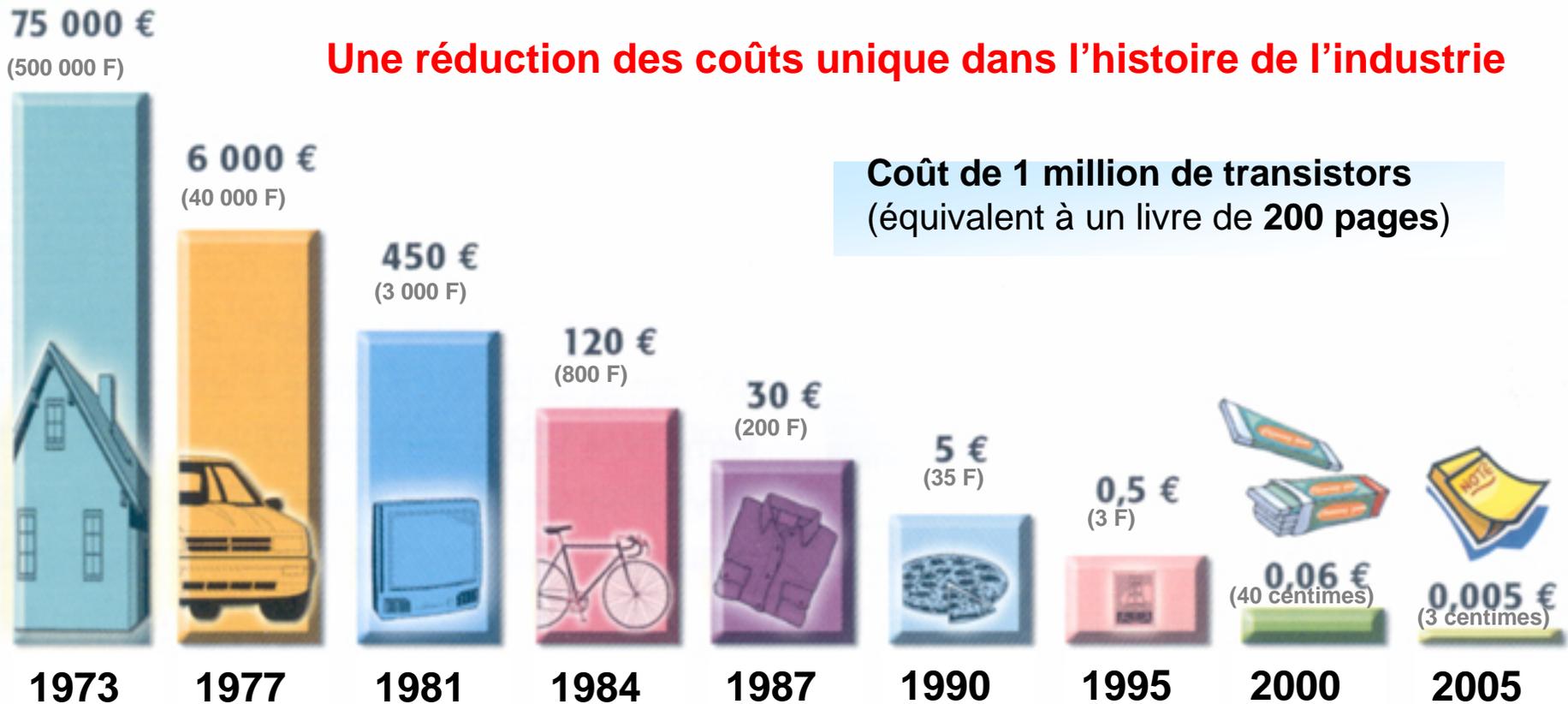


totale ou partielle sur quelque support que ce soit ou utilisation du contenu de ce document sans l'autorisation écrite préalable du CEA. Any reproduction in whole or in part on any medium or use of the information contained herein without the prior written consent of CEA. Tous droits réservés.

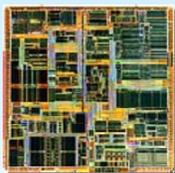
# La micro-électronique et la réduction des coûts

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

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



## Miniaturisation



De plus en plus  
de transistors par  
mm<sup>2</sup> de silicium

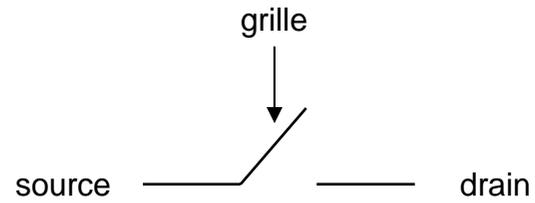
## Traitement collectif



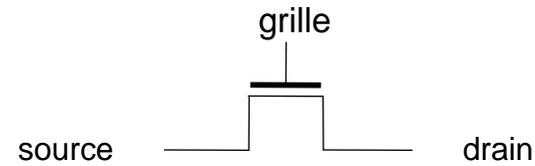
De plus en plus de mm<sup>2</sup>  
de silicium fabriqués  
simultanément

# Une brique de base : le transistor MOS

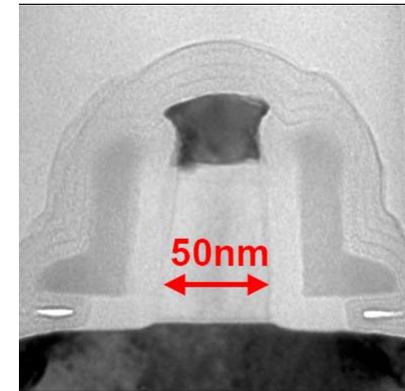
switch



transistor

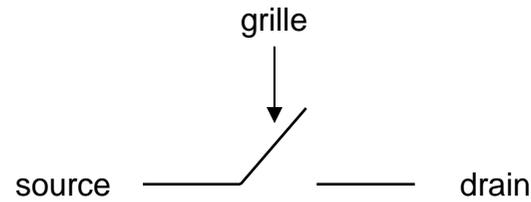


Transistor

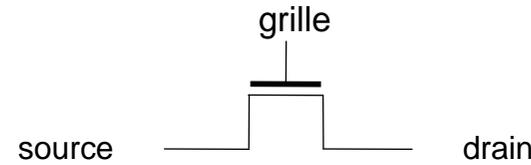


# Une brique de base : le transistor MOS

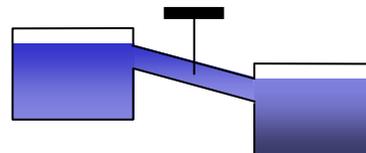
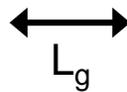
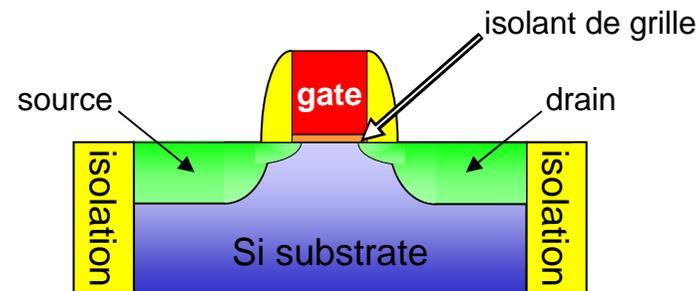
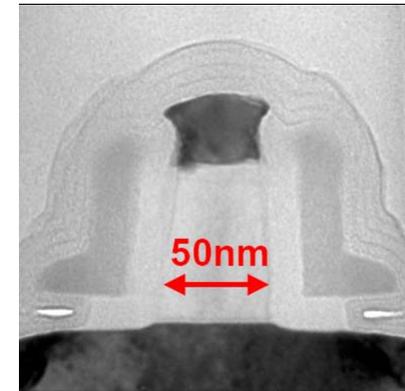
switch



transistor



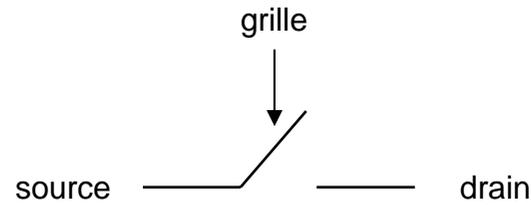
Transistor



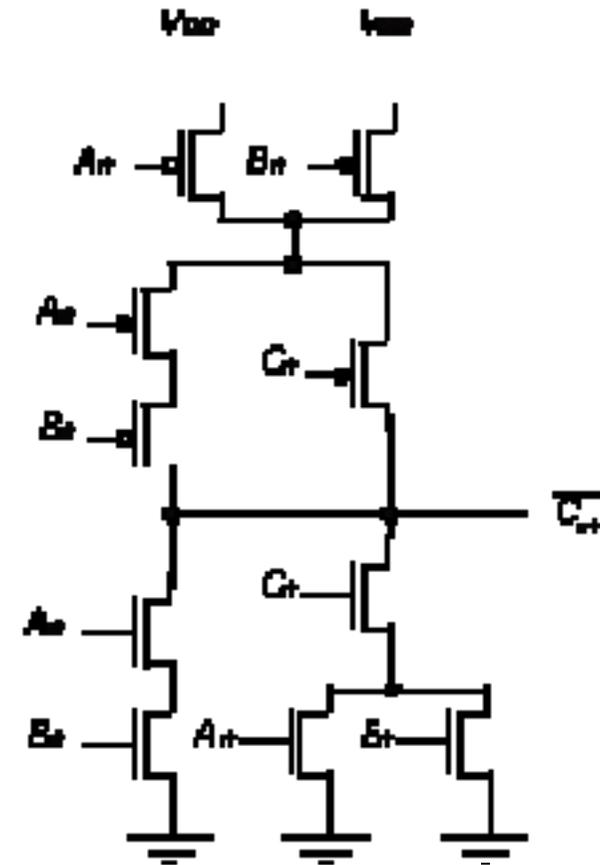
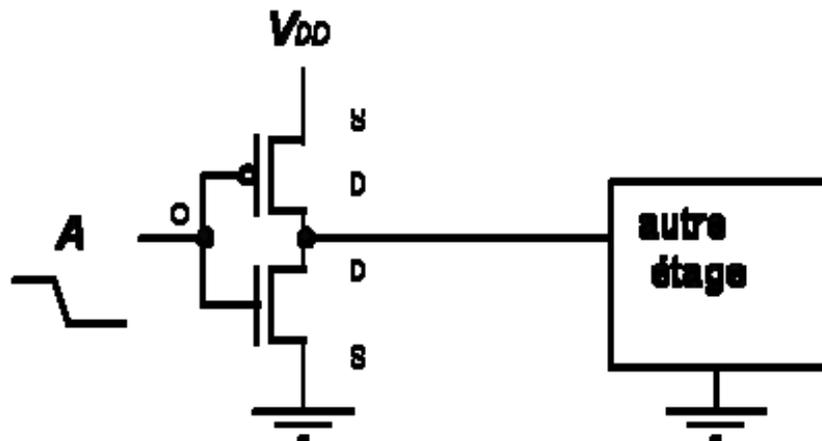
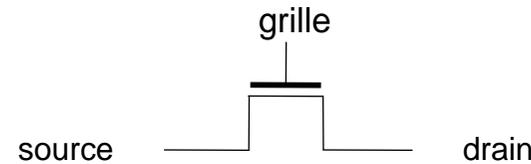
Analogie hydraulique

# Une brique de base : le transistor MOS

switch

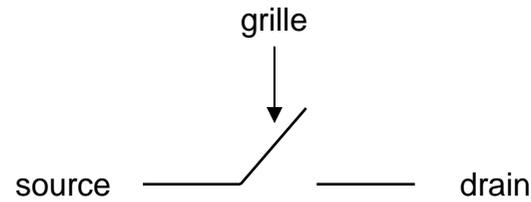


transistor

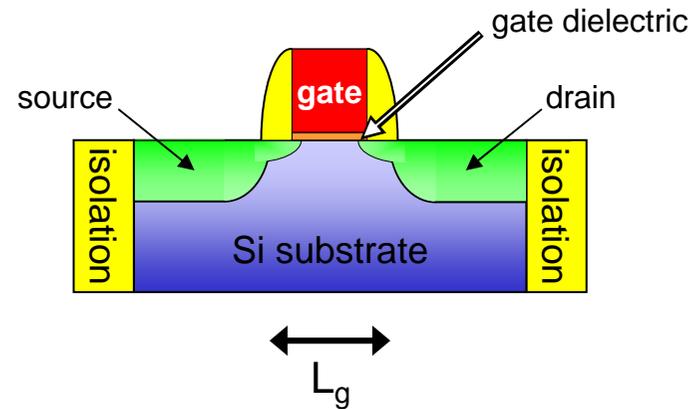
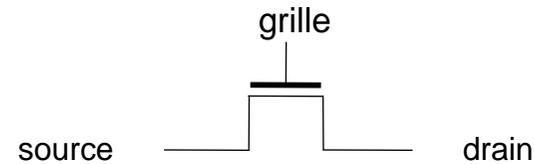


# Une brique de base : le transistor MOS

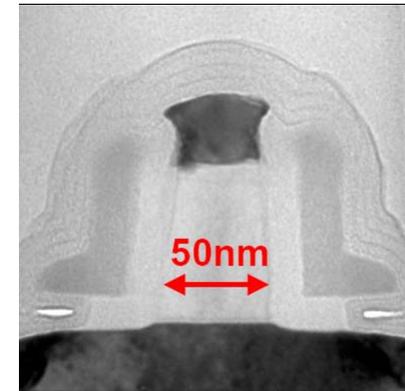
switch



transistor



Transistor

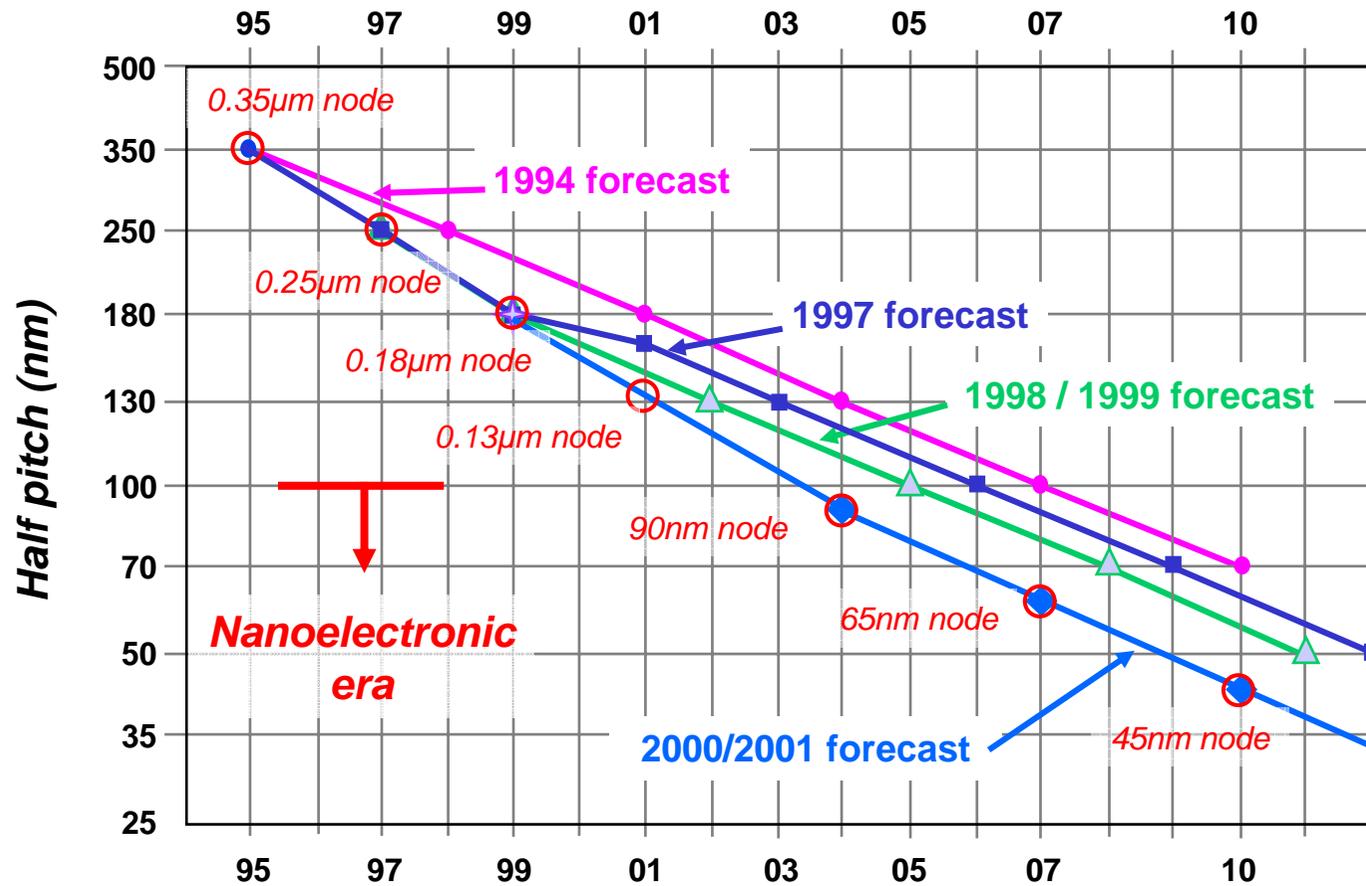


Deux paramètres décisifs :

$I_{off}$

$I_{on}$

# La loi de Moore



La loi de Moore mais....

- **Limites théoriques et pratiques**

Les nouvelles architectures

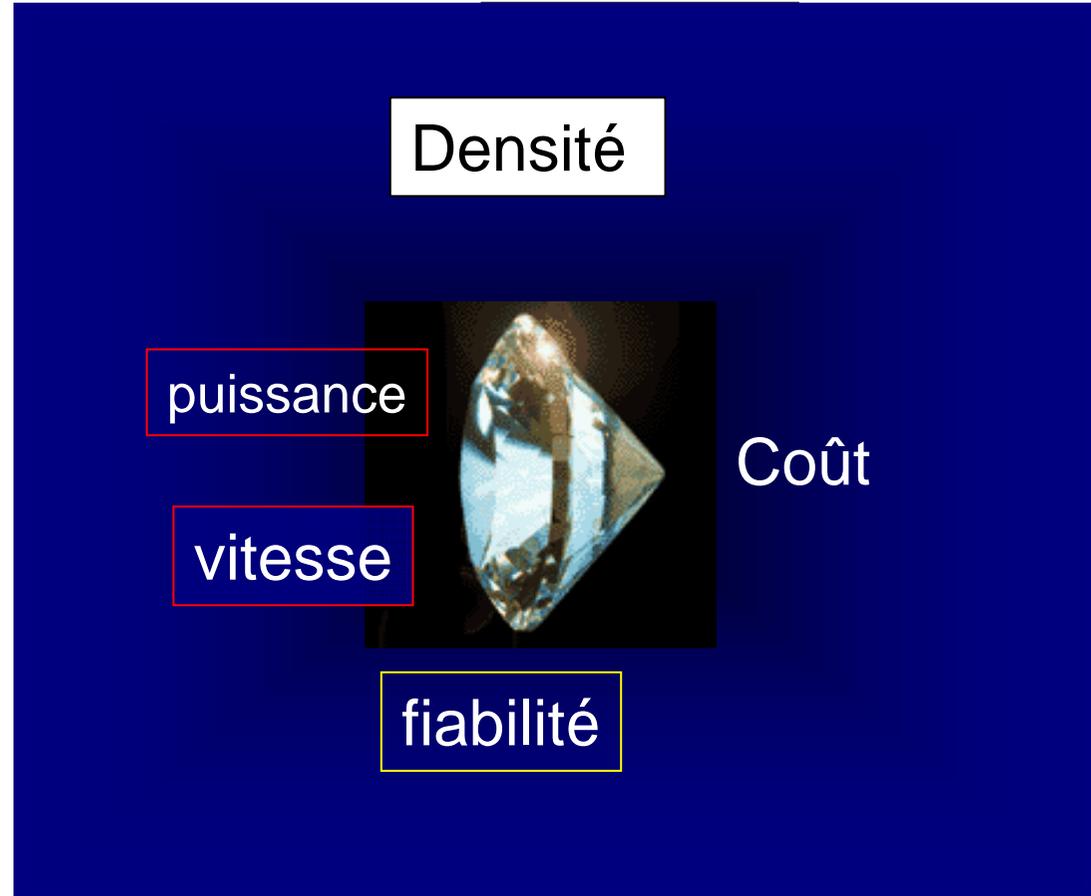
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- Tentative de conclusion

# Conséquences de la miniaturisation



R. Dennard © IEEE

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



**Smaller = better**

Michel Brillouët

# Les limites théoriques

- Boltzmann + Shannon : Energie minimale de commutation

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

- Heisenberg : Longueur minimale du canal

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

- Avec la prise en compte de l'effet tunnel (Zhirnov-Cavin)

$$E \succ k_B T \ln 2 + \frac{\hbar^2 (\ln 2)^2}{8mL^2}$$

# 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

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

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

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

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

Oui mais....

Logique réversible

Hypothèses physiques

.....

# Limites théoriques

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

With duty cycle ~1 %

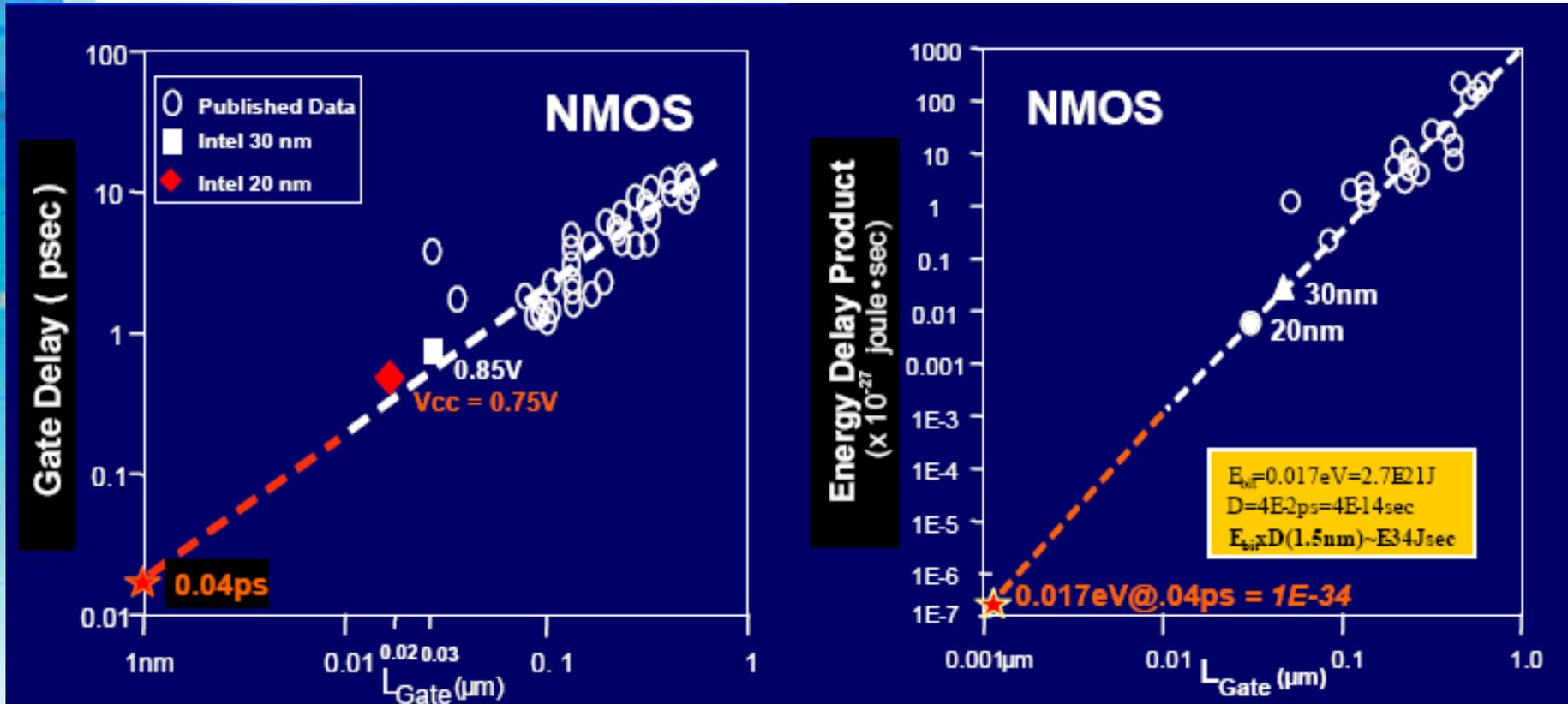
**$3.7 \times 10^4 \text{ W/cm}^2$**

Active transistors ~1 %

**$3.7 \times 10^2 \text{ W/cm}^2$**

Cell area ~ (3-4)x device area

**~100 W/cm<sup>2</sup>**



Source : INTEL 2007

# Energy efficiency of CMOS



## How far from fundamental limits CMOS operates?

- 2011 ITRS 22nm node

- Channel length  $\sim 9\text{nm}$
- Switching energy  $\sim 2 \times 10^{-18} \text{ J}$
- S-Ch Barrier height  $\sim 0.4\text{eV}$
- Electrons/switching event  $\sim 50$
- Energy/electron  $\sim 4 \times 10^{-20} \sim 12\text{kT}$

### Fundamental limits

**1.5nm**

**$kT\ln 2 = 3 \times 10^{-21} \text{ J}$**

**$kT\ln 2 = 0.017 \text{ eV}$**

**1**

**$3 \times 10^{-21} \sim kT$**

Scaling to atomic dimensions may not give performance increase: optimal dimensions could range between 5 and 20nm!

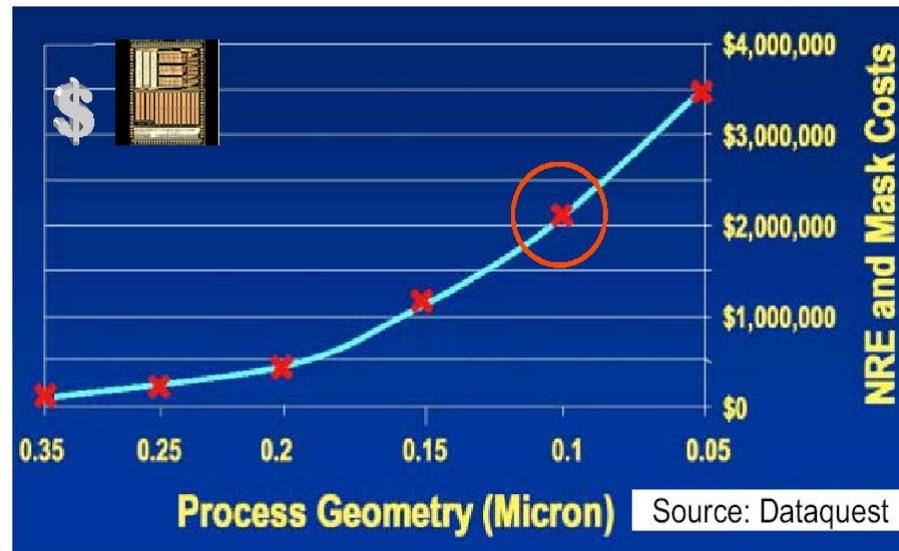
Source: Thomas Skotnicki and J.D. Meindl

# Limites pratiques

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

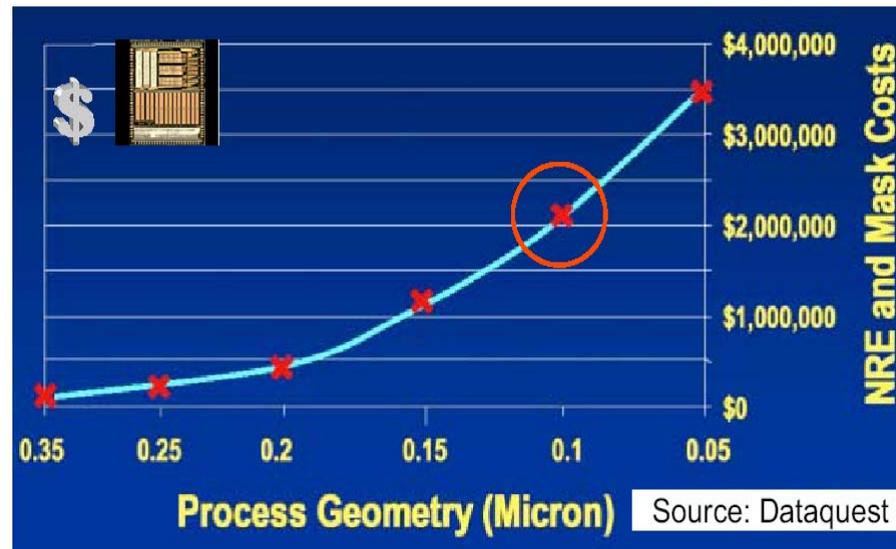
# Limites pratiques

- Le coût des masques



# Limites pratiques

- Le coût des masques

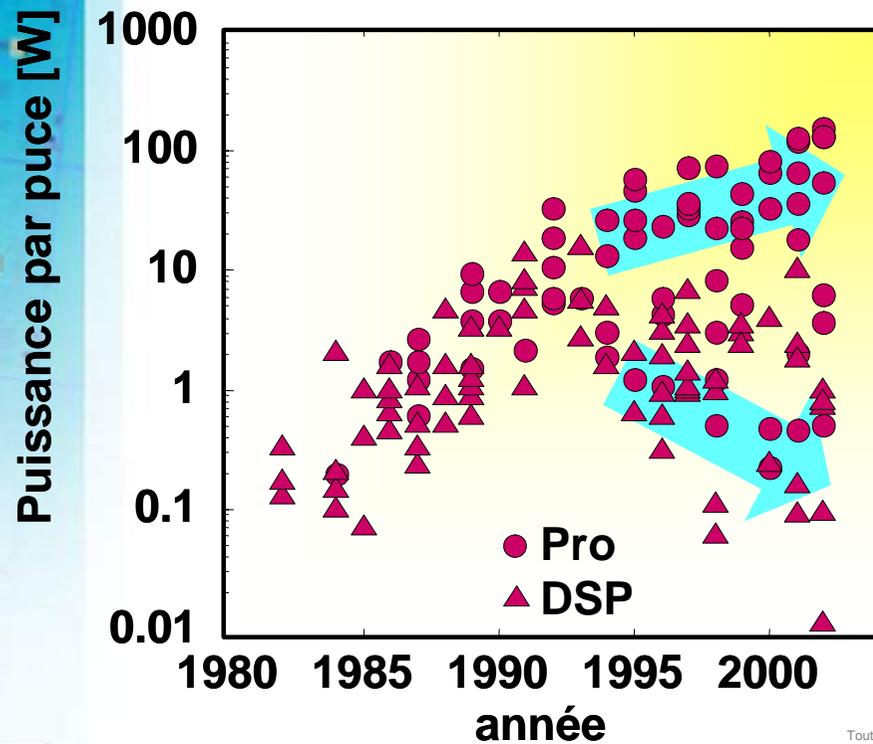


## Solutions

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

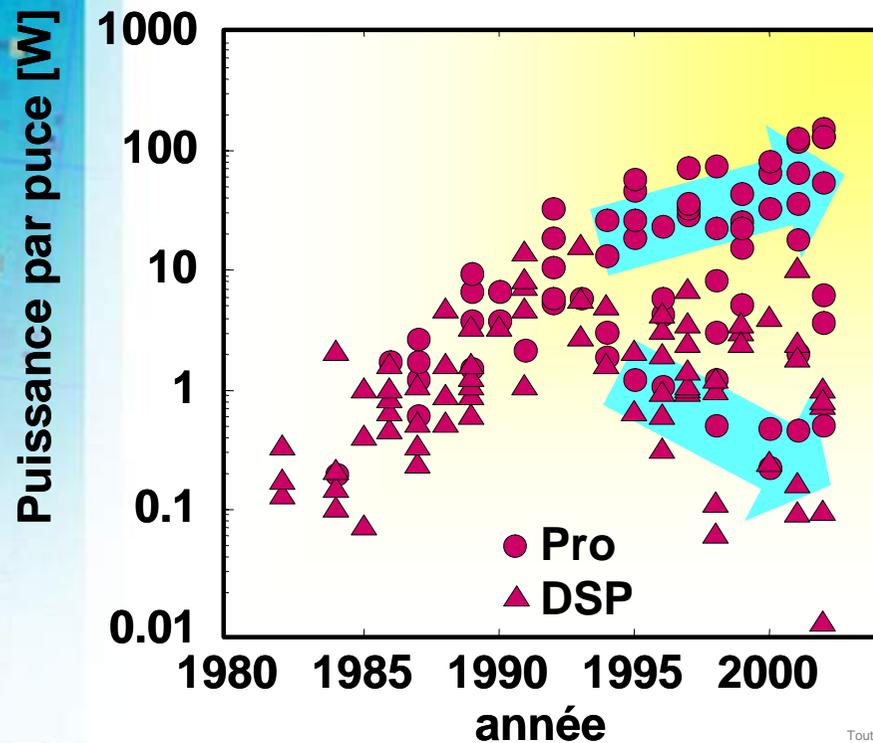
# Limites pratiques

- La puissance dissipée



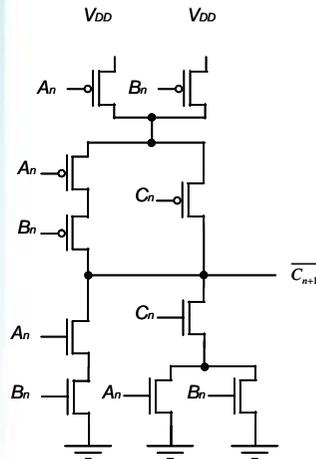
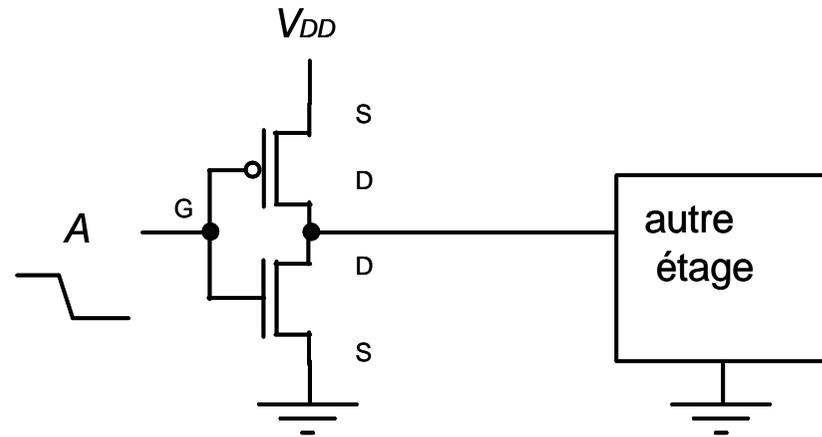
# Limites pratiques

- La puissance dissipée



## Solutions

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



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



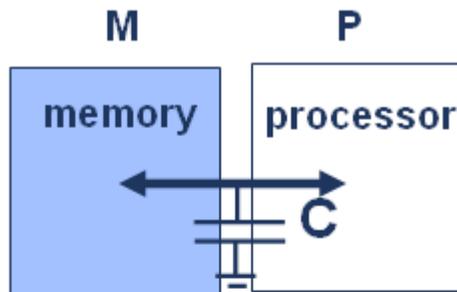
$$P = a f C V_{DD}^2 + I_0 V_{DD} \exp^{-\frac{V_T}{S}} + V_{DD} I_t$$

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

Minimiser les distances d'échange

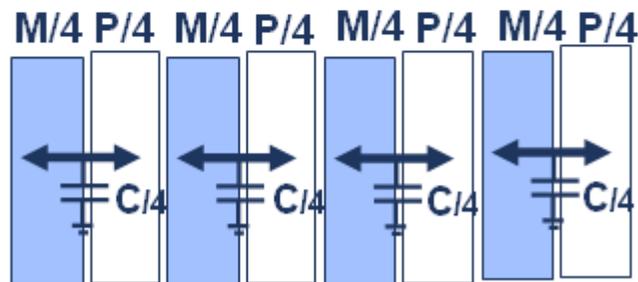
Contrôler les tensions de seuil

Limiter la miniaturisation



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

Intérêt du  
parallélisme pour la  
consommation  
dynamique



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

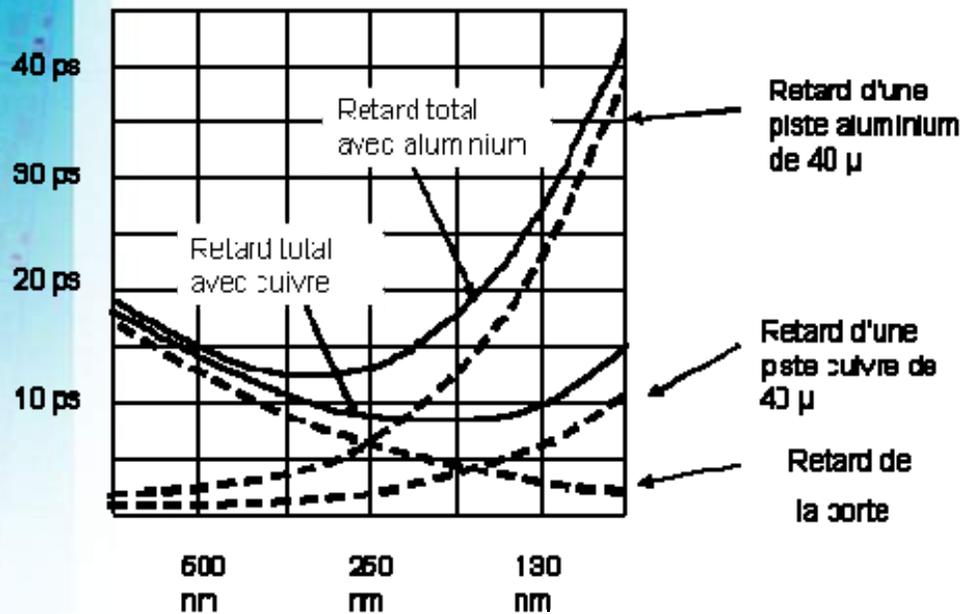
But if  $f \sim V$

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

Attention pour la  
consommation  
statique

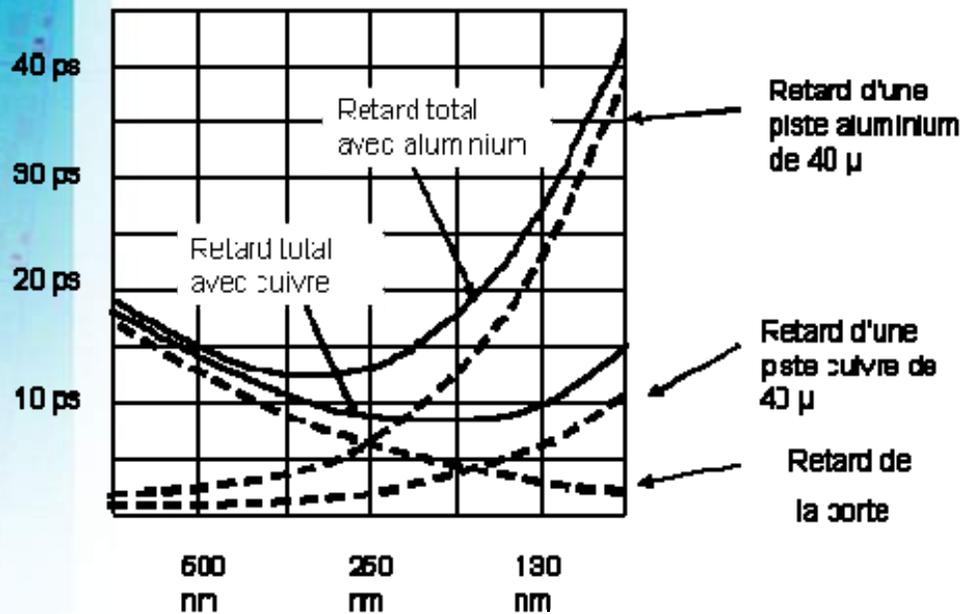
# Limites pratiques

- Le délai des interconnexions



# Limites pratiques

- Le délai des interconnexions

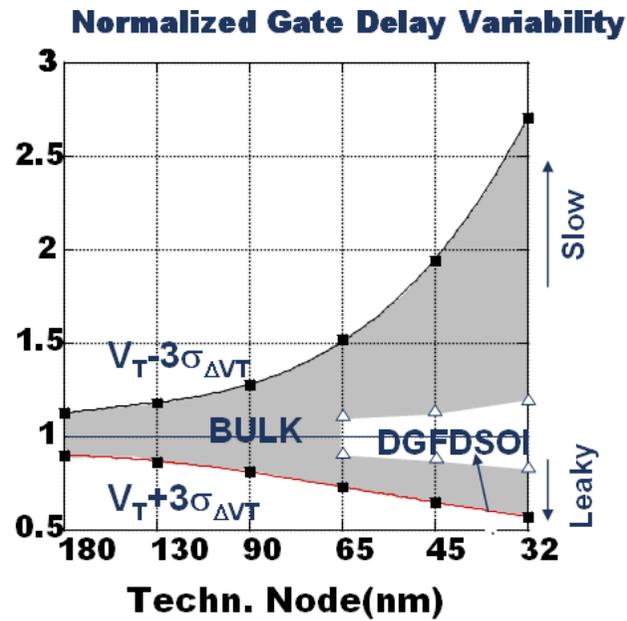
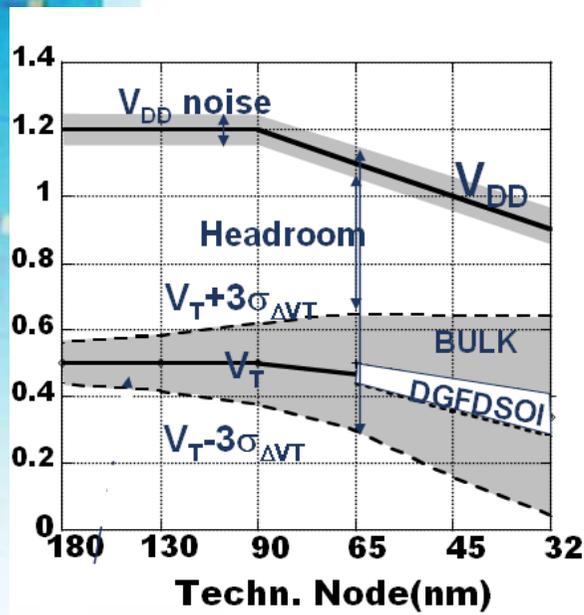


## Solutions

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

# Limites pratiques

- Augmentation des dispersions

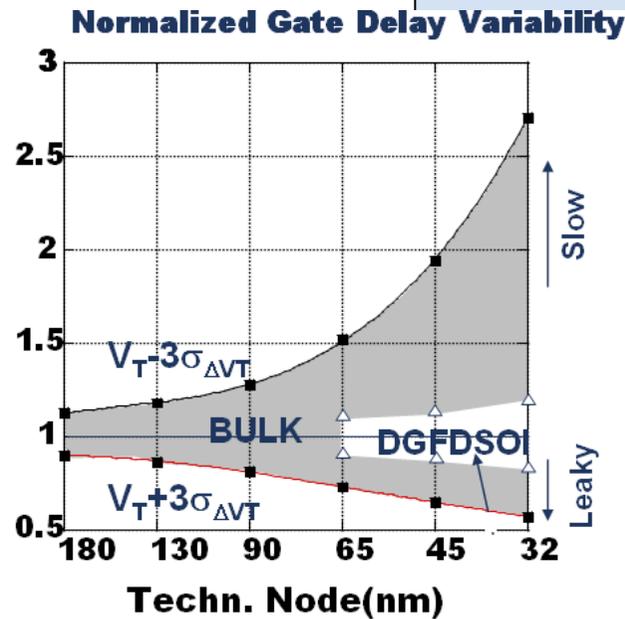
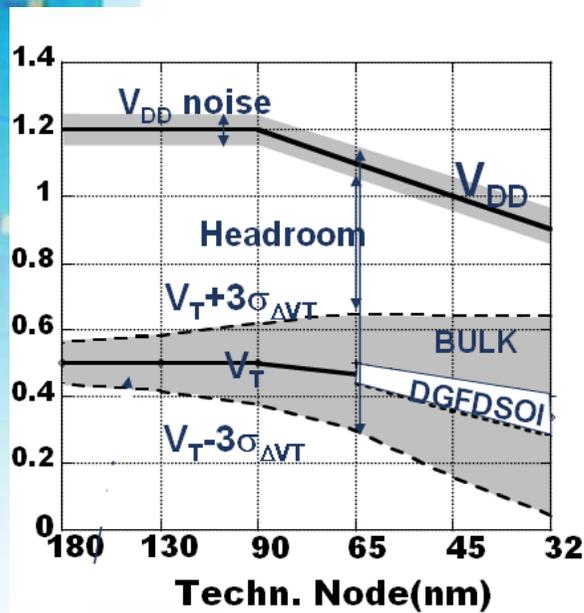


# Limites pratiques

- Augmentation des dispersions

## Solutions

- Architectures tolérantes
- Nouveaux composants



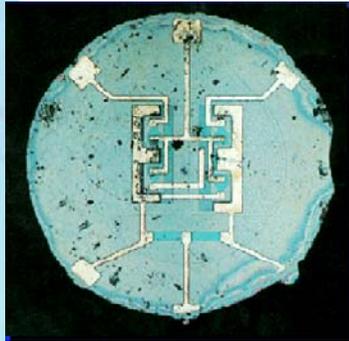
La loi de Moore mais....

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

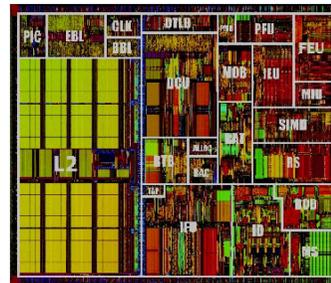
Un grand nombre de nouveaux dispositifs

- Quelles architectures pour les nanocomposants
- Tentative de conclusion

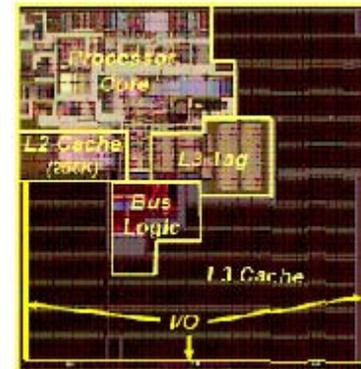
# Evolution des architectures



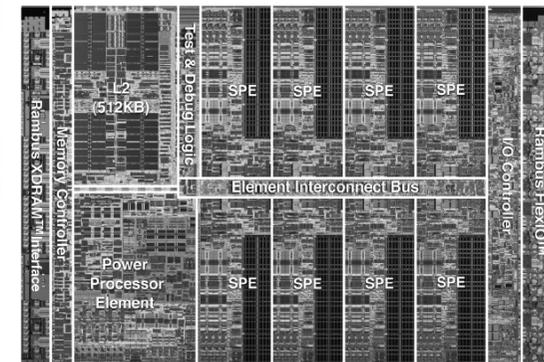
Le premier circuit intégré



Le pentium 4



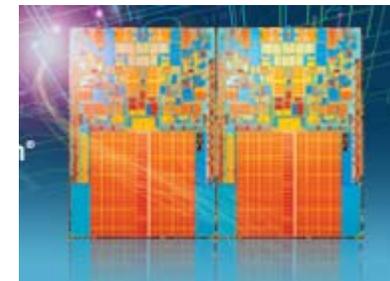
Le processeur Itanium



Le processeur Cell

- Plus de transistors
- Plus de mémoire
- Parallélisme

Le processeur 4 cœurs INTEL



# De nouvelles architectures

Ease of programming  
Low power  
Flexibility  
High performance

Dynamic behavior

Few applications  
Pseudo static behavior



1 algorithm

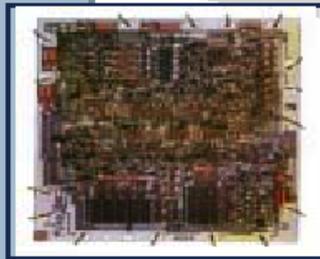
1 application



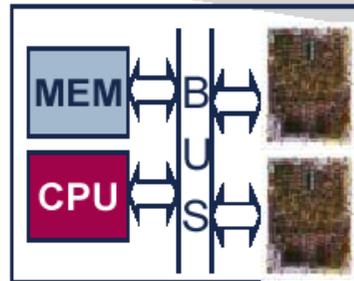
Middleware

Operating system

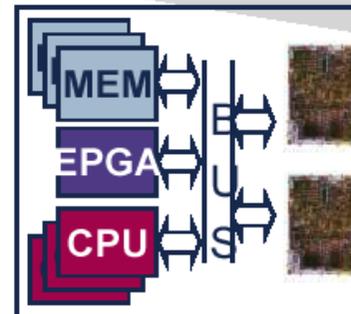
SW & HW multi-tasking



ASIC



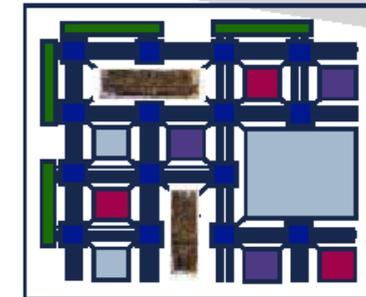
SoC



Multi-processor  
SoC platform

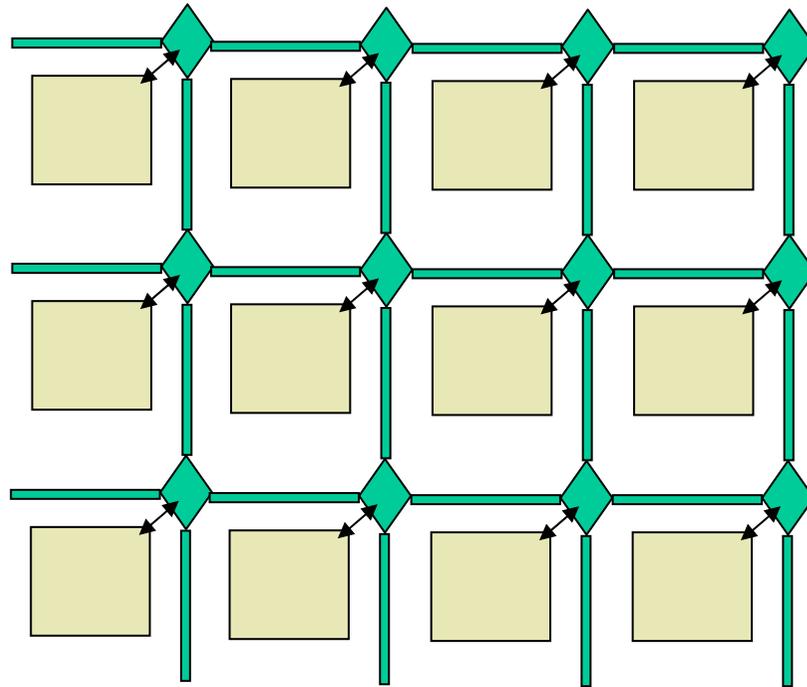
Re-configurable hardware  
On-chip communication network

Memory hierarchy



Flexible SoC platform

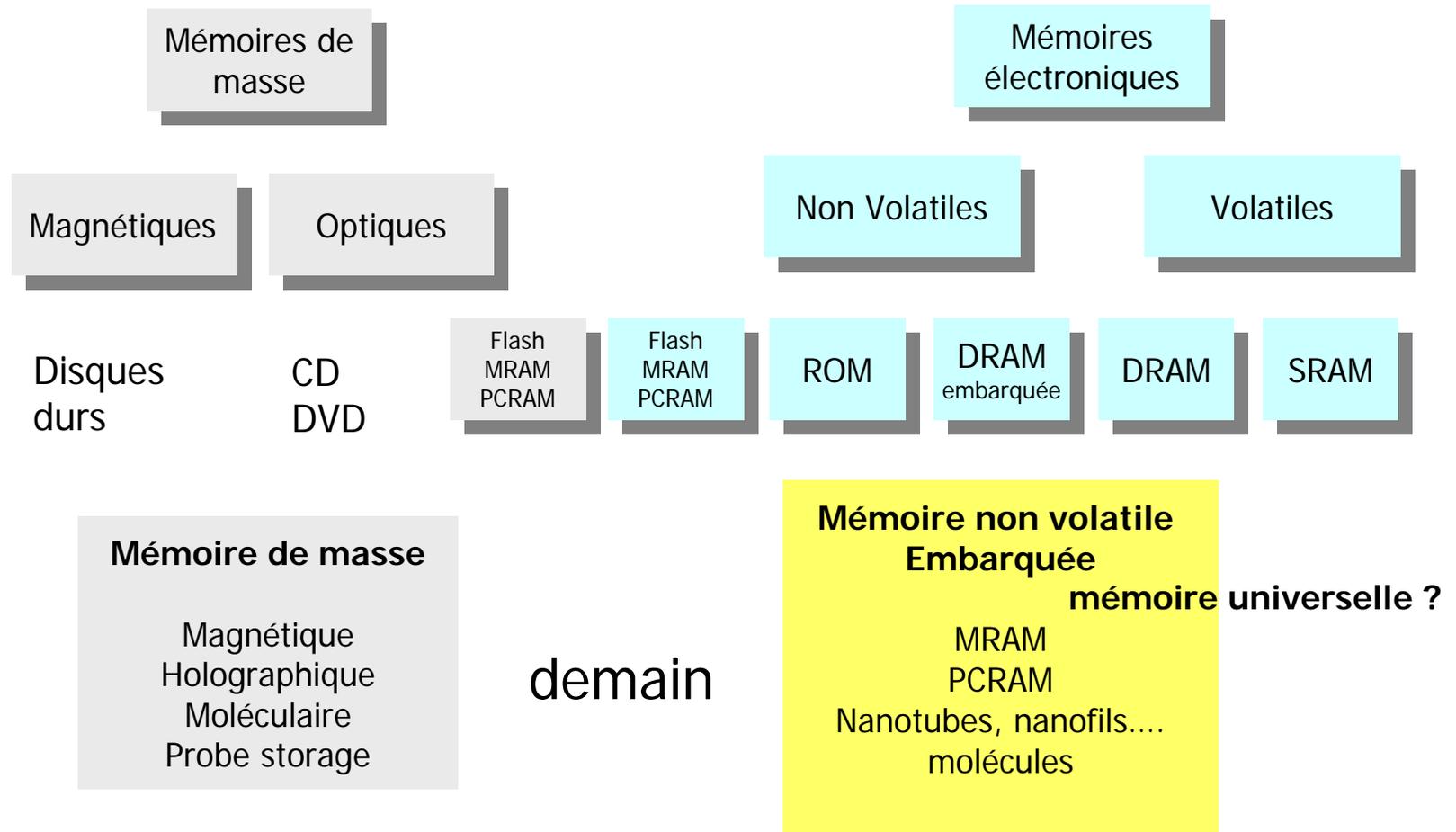
# L'architecture du futur ?



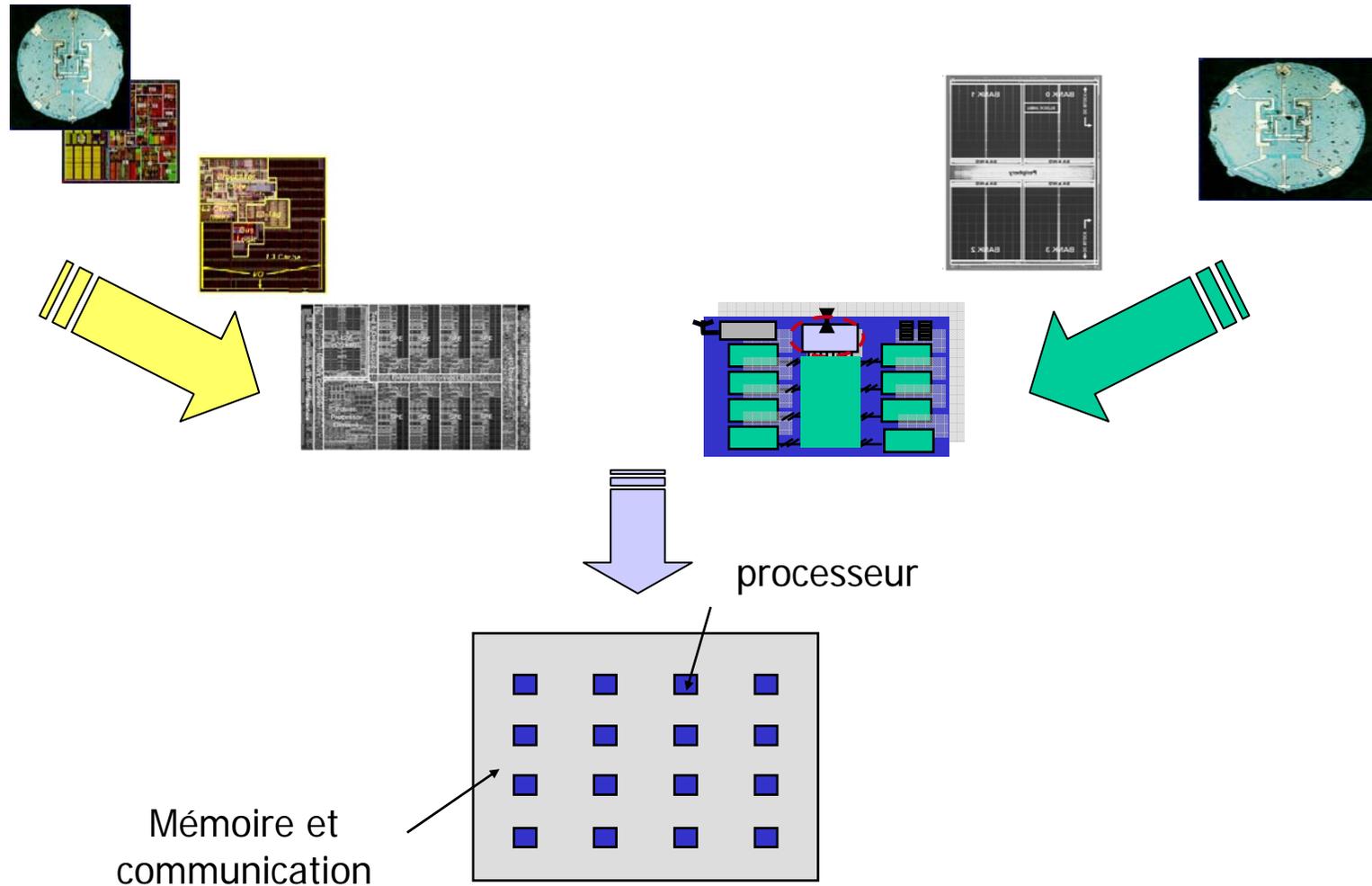
- Calcul ou mémoire
- 30 millions de transistors
- 1 GHz

## Le Network On Chip de type GAL

# Les Mémoires ne sont pas en reste



# Convergence processeur-mémoire vive



## Old principles

- Power is free but transistors are expensive
- The only concern is dynamic power
- Processors are reliable with errors occurring at the pins
- Multiply is slow but load and store is fast
- Increasing clock frequency is the primary method to improve performances
- Large memories are out of processing chips

From landscape of computing research :  
a view from Berkeley

## New principles

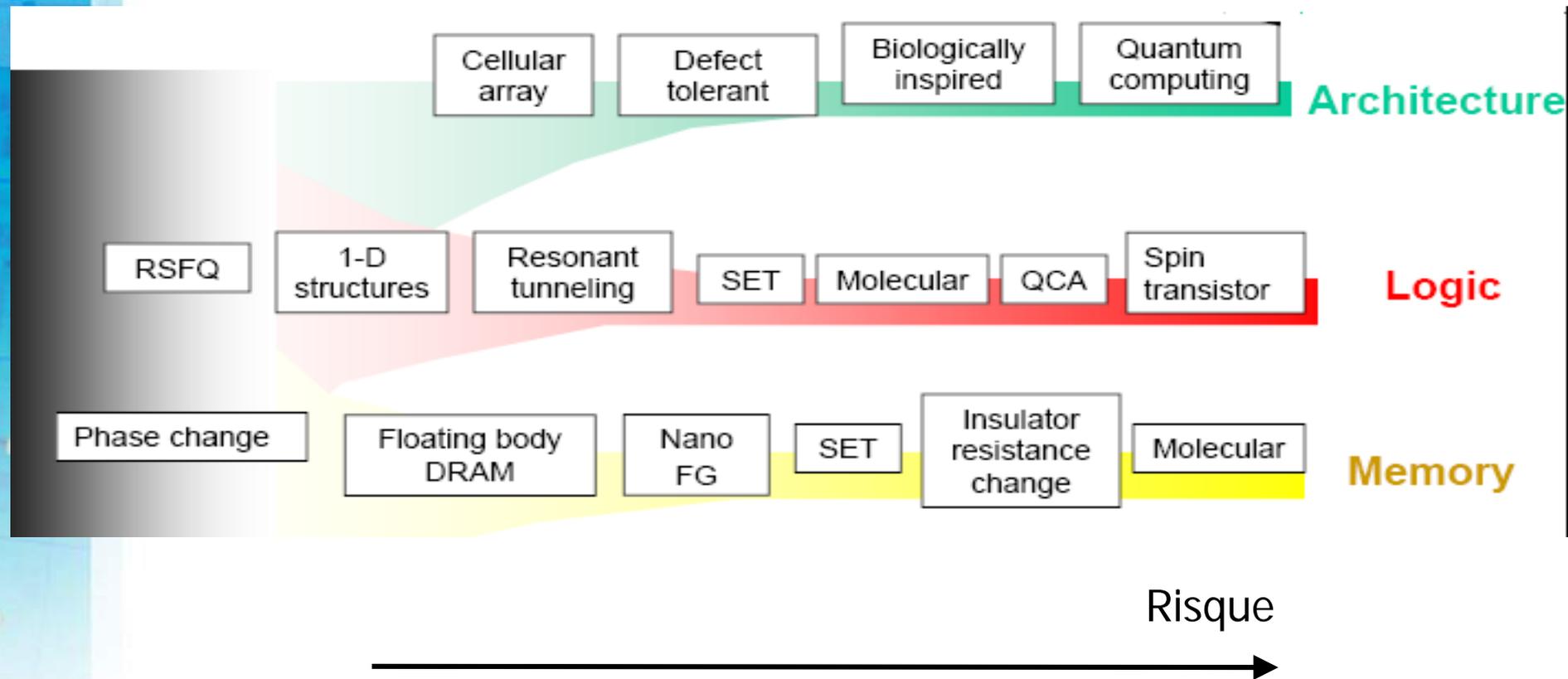
- Transistor is free but power is expensive
- Static power can be 40% of power
- Chips below 65 nm will have soft and hard error rates
- Multiply is fast but load and store is slow
- Increasing parallelism is the primary method to improve performances
- Large memories are if possible embedded

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

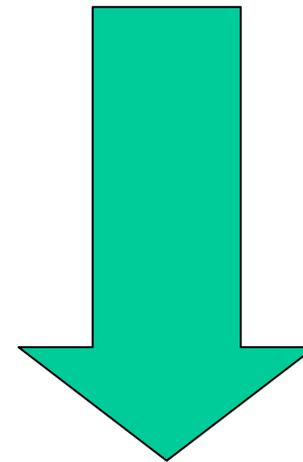
## La loi de Moore mais....

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

# Vision de l'ITRS : dispositifs émergents

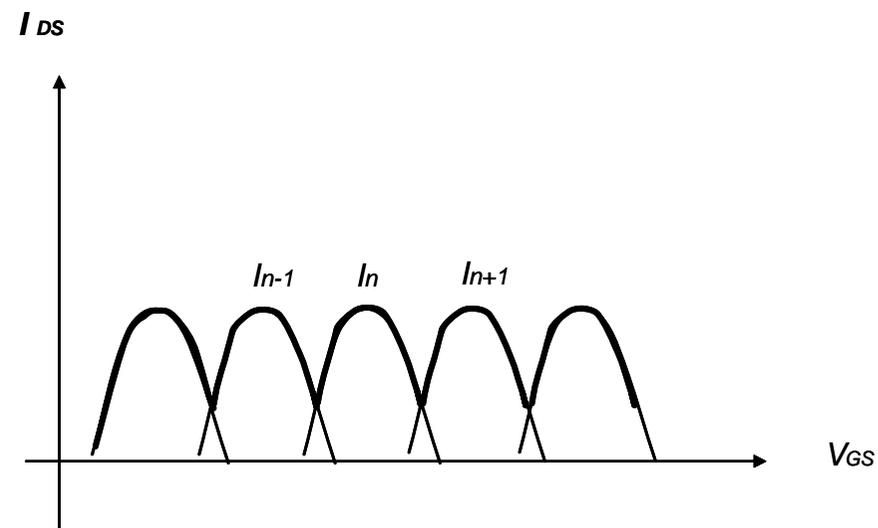
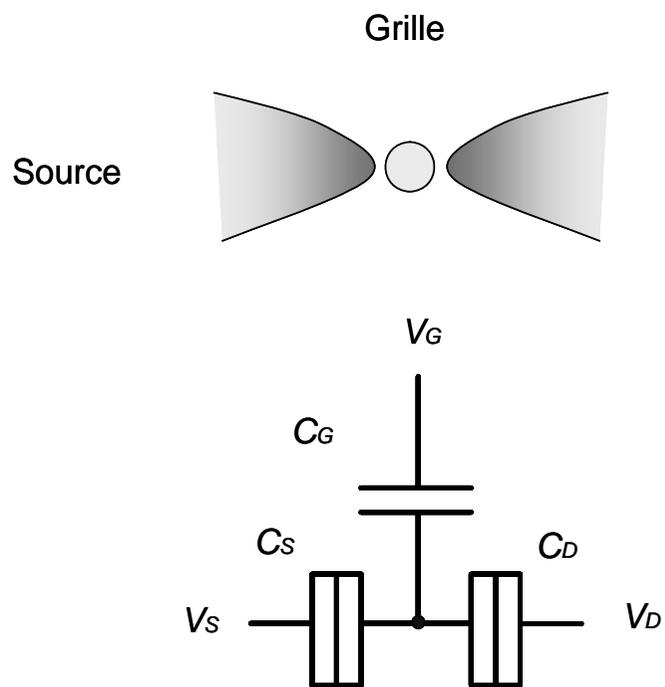


- Transistor à électron unique
- SpinFET
- Nanotubes de carbone et nanofils
- Nanocristaux
- Molécules et atomes



Taille du  
dispositif

# Le transistor à électron unique

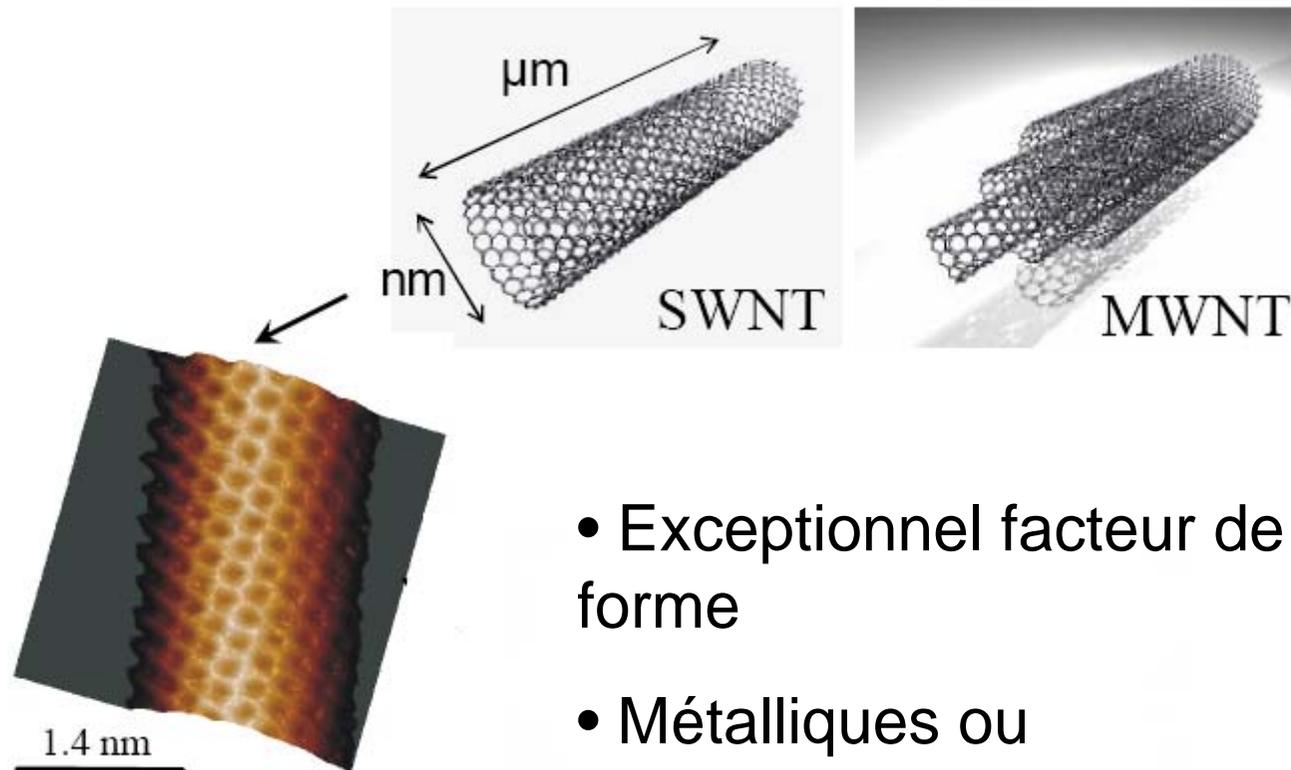


## Le transistor à électron unique

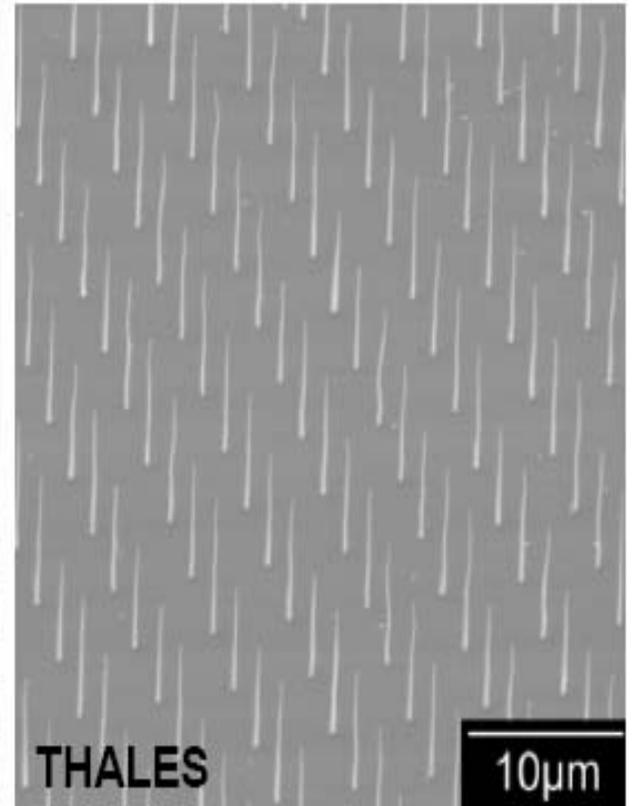
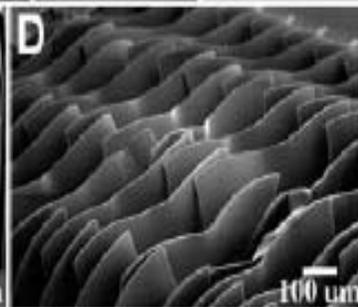
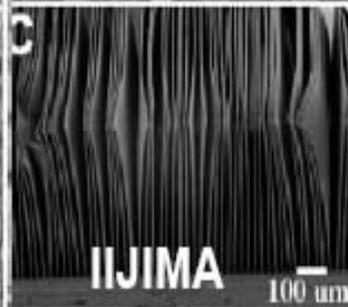
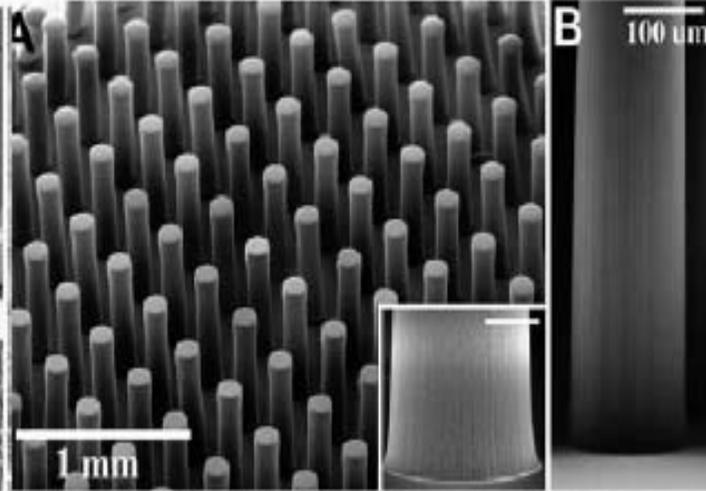
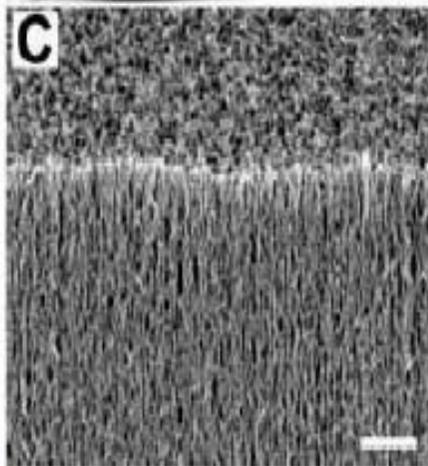
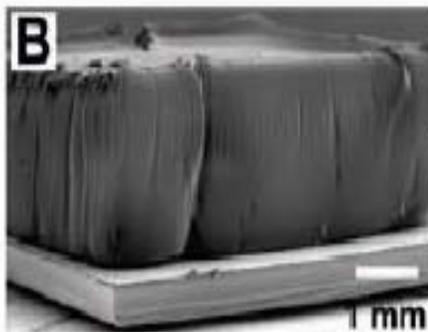
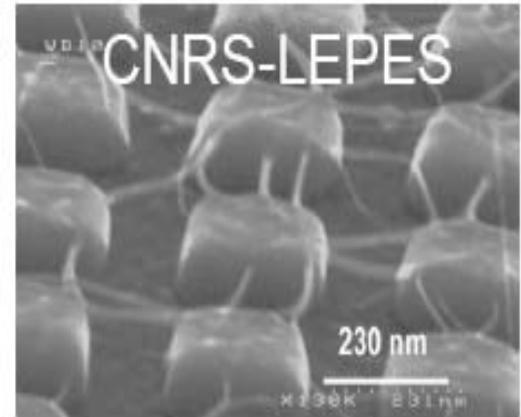
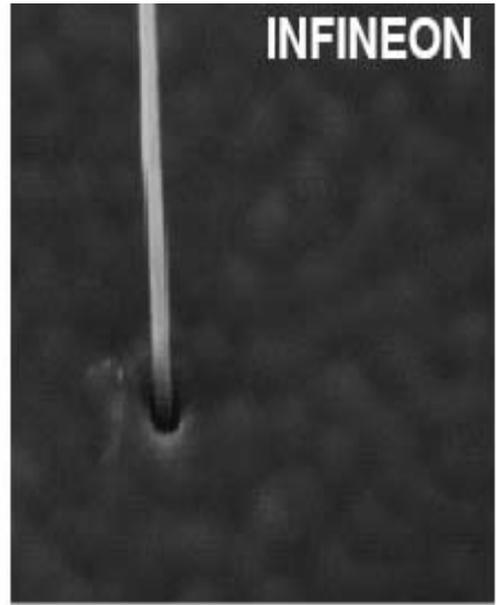
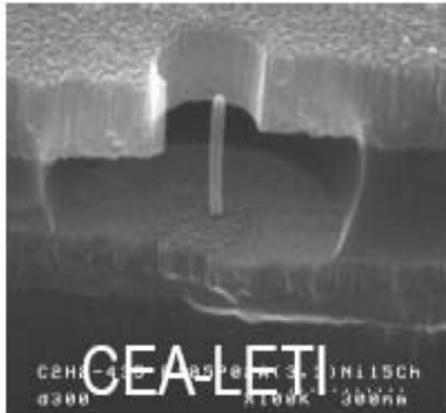
- Fonctionnement à température ambiante ?
- Capacités des interconnexions

Miniaturisation extrême nécessaire (10 nm)

## Les nanotubes de carbone



- Exceptionnel facteur de forme
- Métalliques ou semiconducteurs



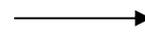
# Comparer les nanocomposants

$CV / I$  versus  $L_g$



vitesse

$CV / I \cdot CV^2$  versus  $L_g$



vitesse-énergie

$I_{on} / I_{off}$  versus  $L_g$

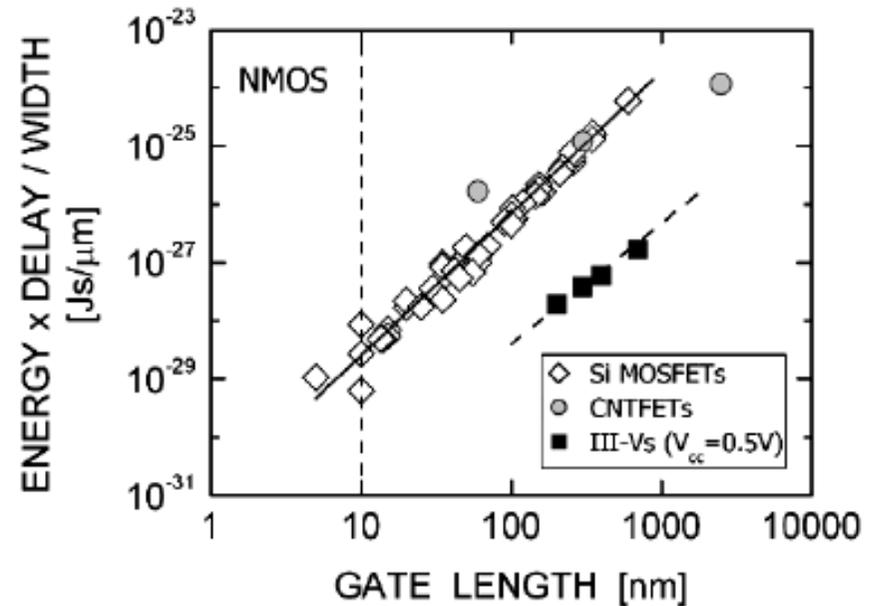
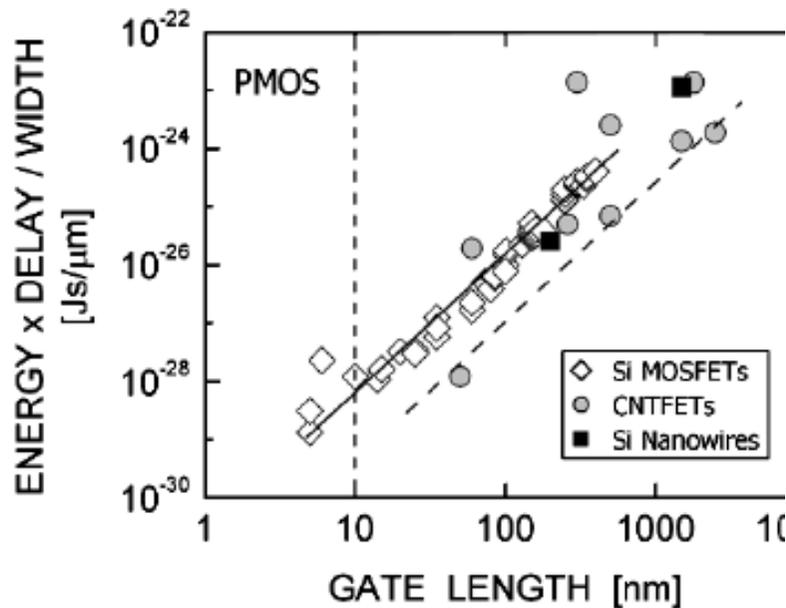


Scalability

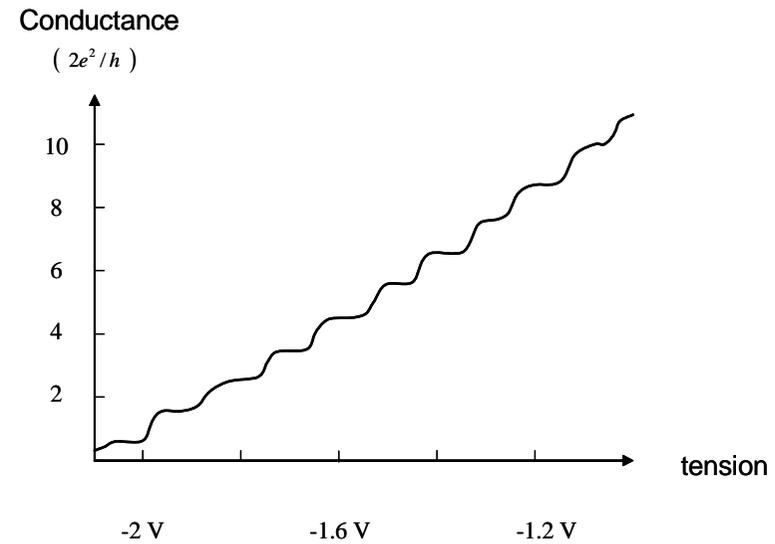
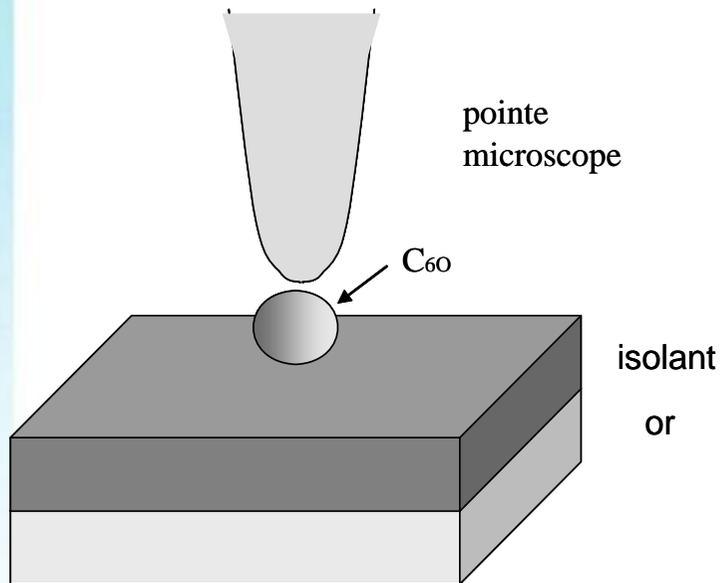
$CV / I$  versus  $I_{on} / I_{off}$



fuites

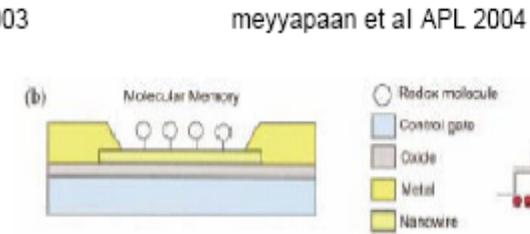
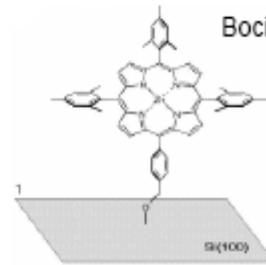


# Propriétés électriques de la molécule unique

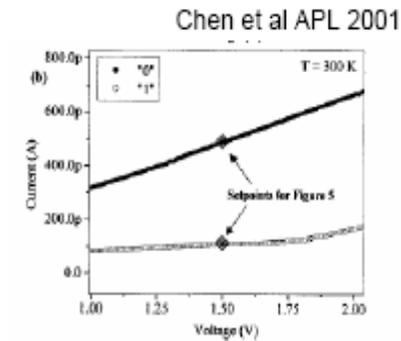
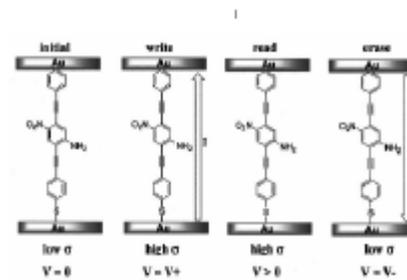


# Applications des molécules

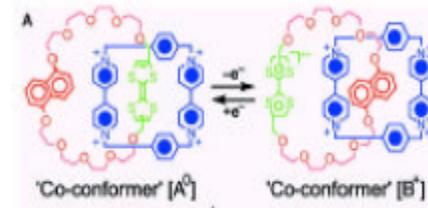
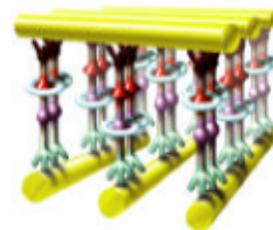
Stockage multivalué



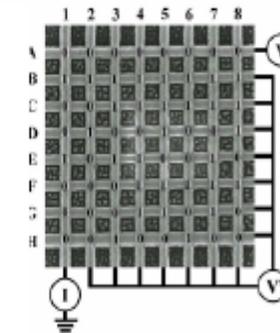
Résistance négative



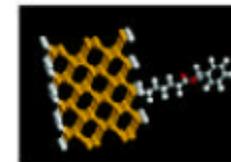
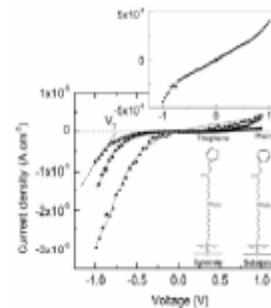
Interrupteurs



Williams, Stoddart, Heath  
UCLA/HP



Diodes

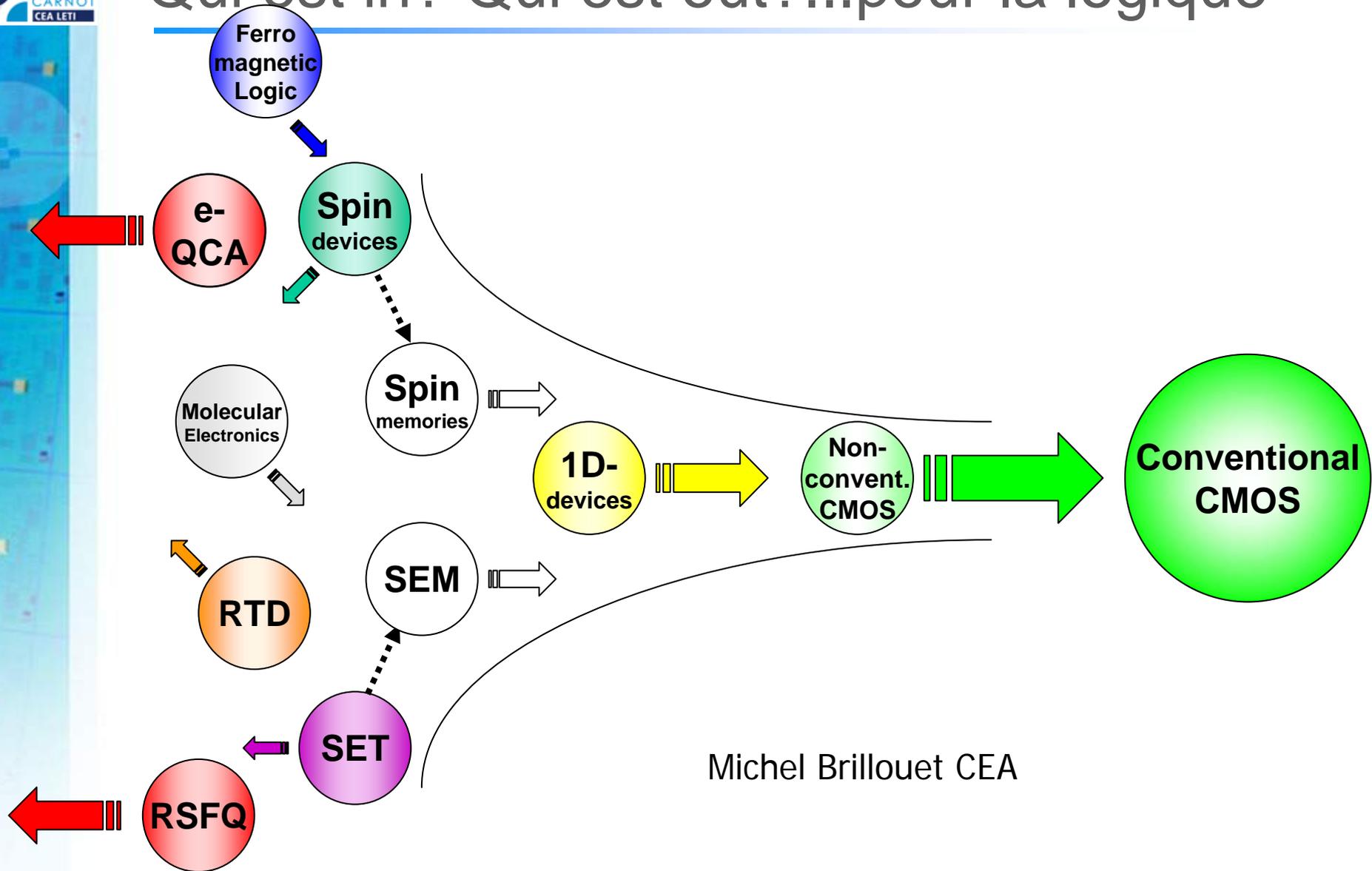


Vuillaume et al Nanolett 2003

JP Bourgoin CEA

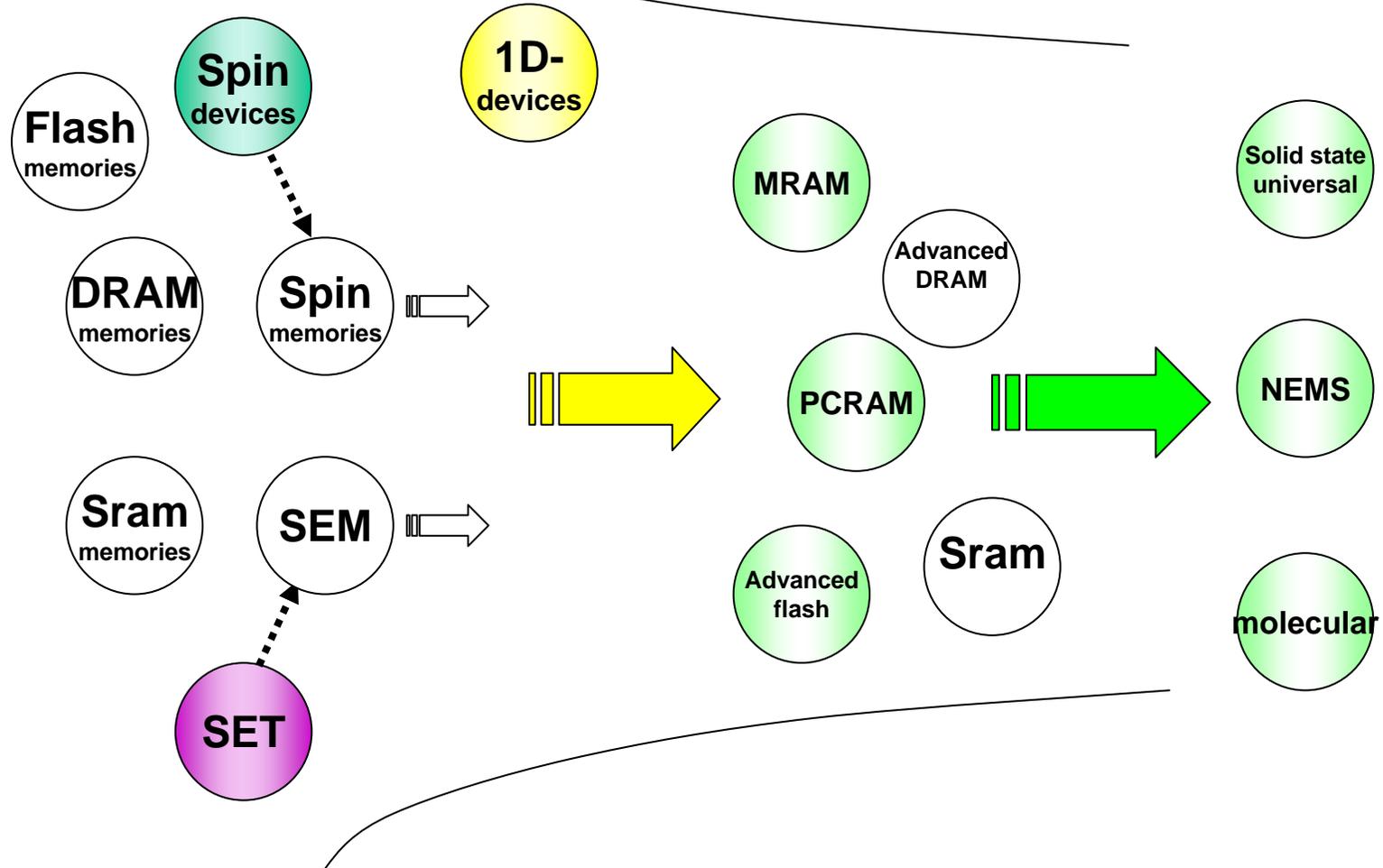
le 30 09 2005

# Qui est in? Qui est out?...pour la logique



Michel Brillouet CEA

# Qui est in? Qui est out?...pour les mémoires



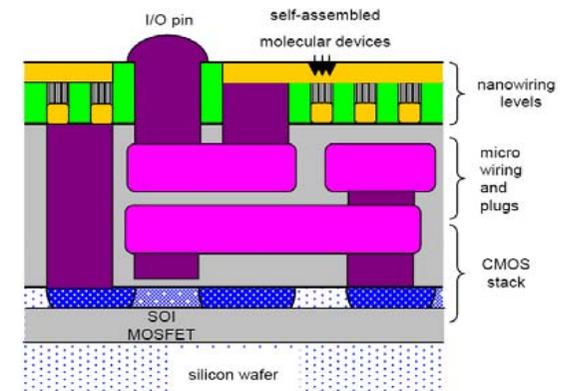
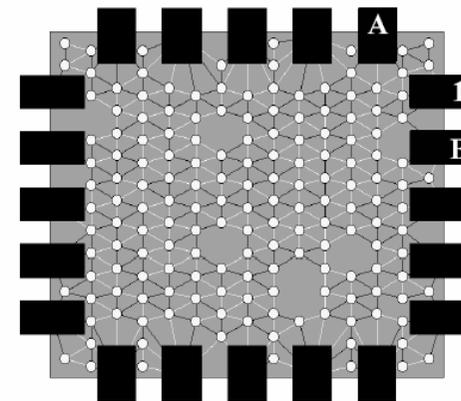
La loi de Moore mais....

- 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 »



## Les règles à prendre en compte

---

- 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 ou graphène
- 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

# Le calcul quantique

---

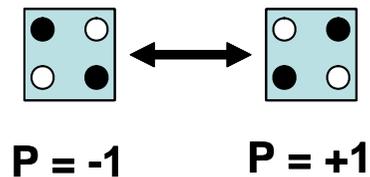
De nombreuses idées associées :

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

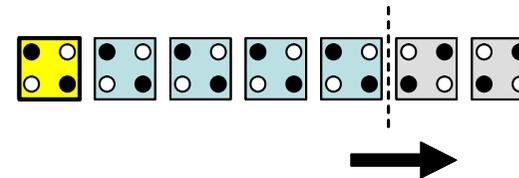
Des difficultés considérables de mise en œuvre pratique et une utilisation dans l'avenir peut-être limitée à quelques algorithmes particuliers.

# Les réseaux d'automates cellulaires

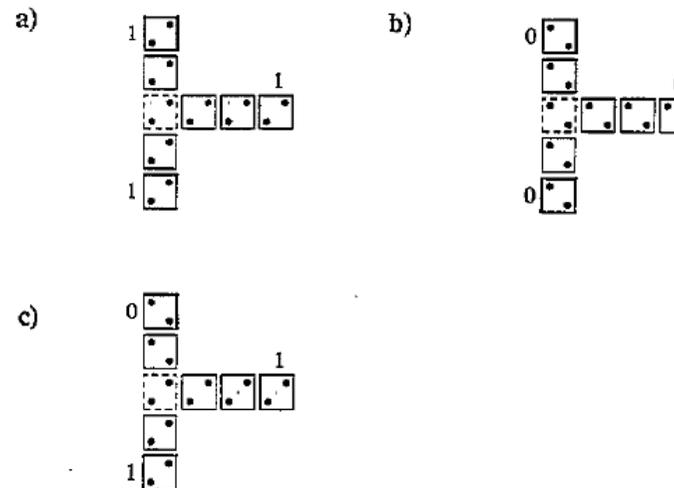
Semble résoudre les difficultés des interconnexions



- 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

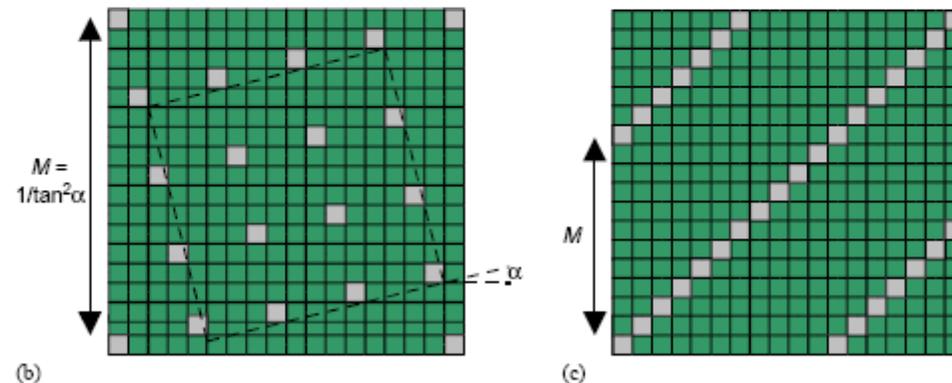
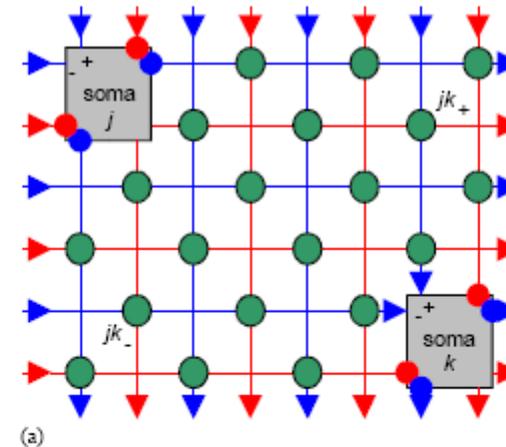
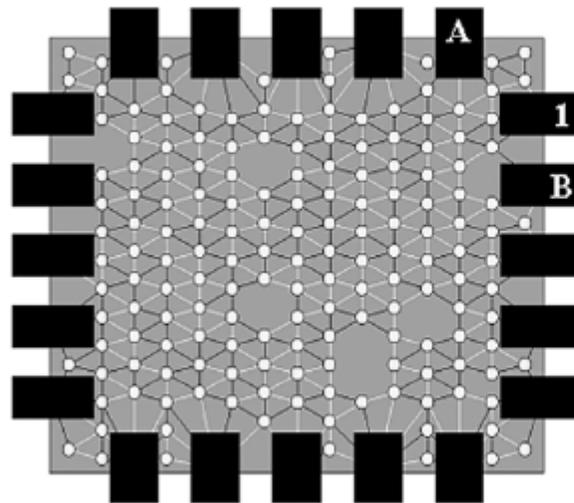
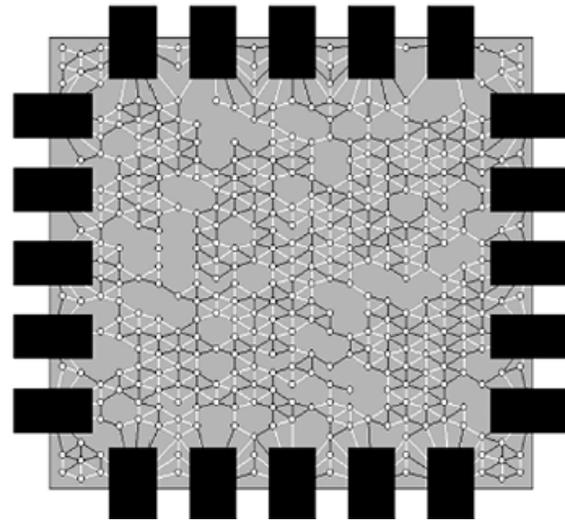
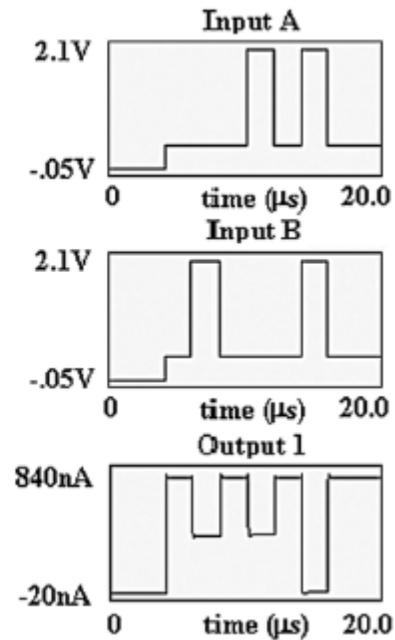


Fig. 1. CrossNets: (a) general structure, and two major species: (b) InBar and (c) FlossBar [1,9,10,5]. Red lines are axonic, and blue lines are dendritic nanowires. Gray squares show interfaces between nanowires and CMOS-based cell bodies (somas). Signs show the dendritic/axonic input/output polarities. 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

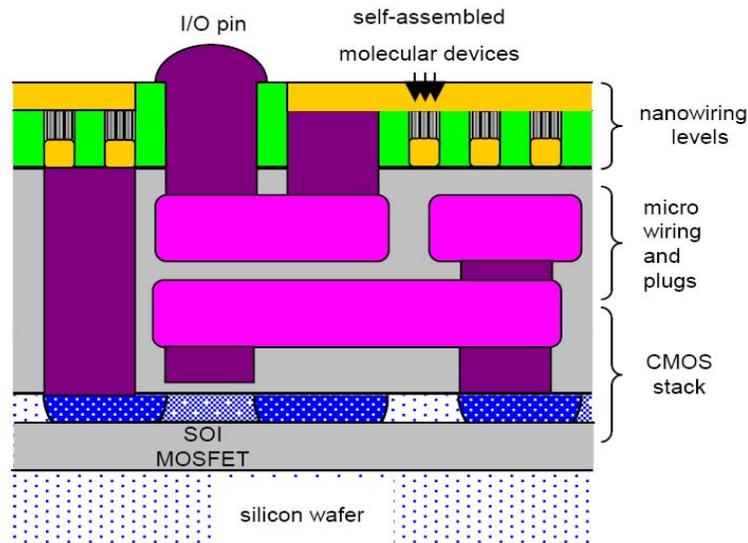


Un réseau désordonné de molécules et de nanoparticules d'or



James Tour,  
Summer M  
Husband, 2002

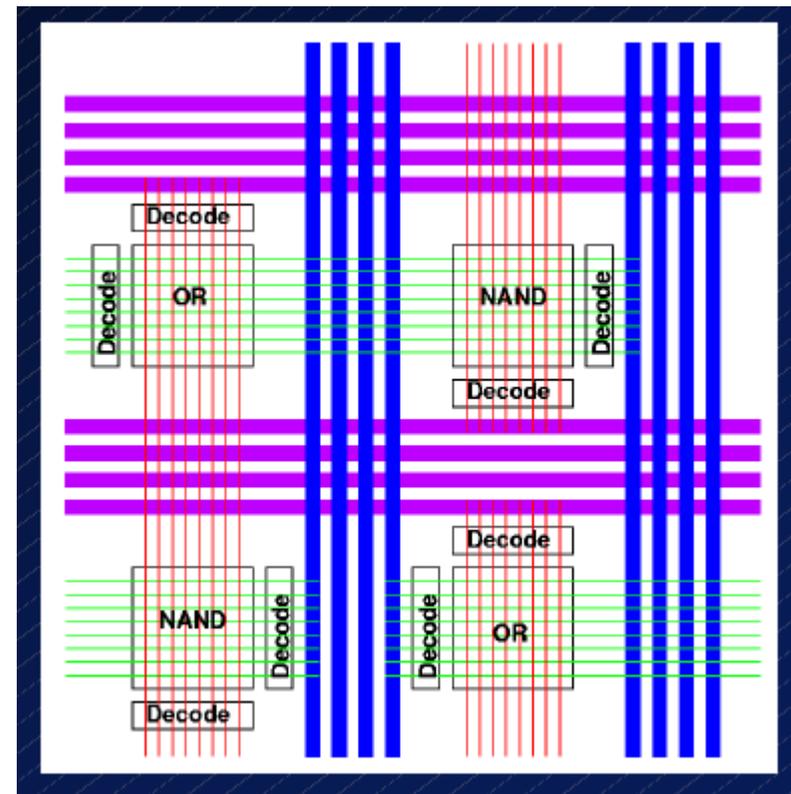
# L'approche « nano inside »



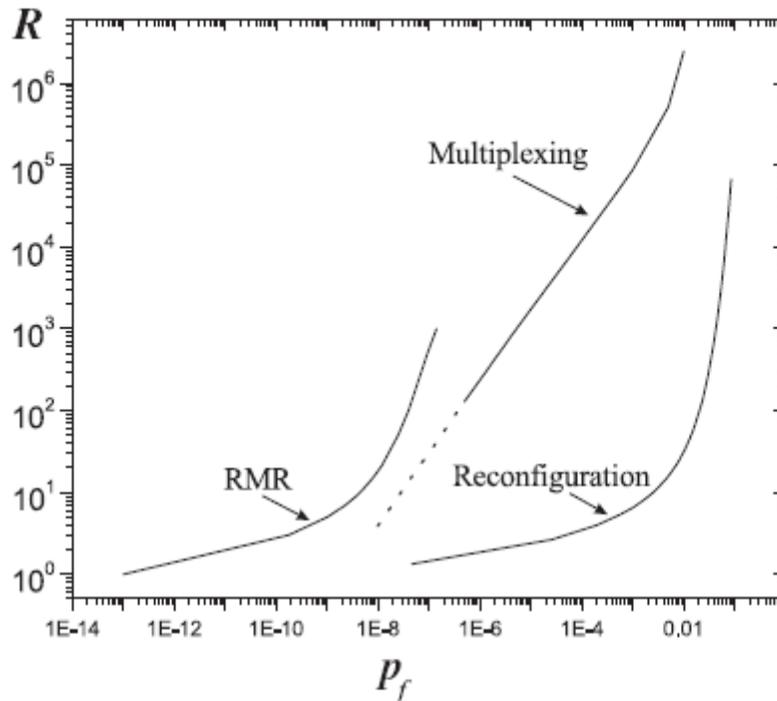
- Blocs mémoire
- Blocs logiques programmables
- Coprocesseurs d'interface avec environnement

A Déhon  
2002

S Likharev  
2003



# 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  $1 \text{ cm}^2$  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.



## La loi de Moore mais....

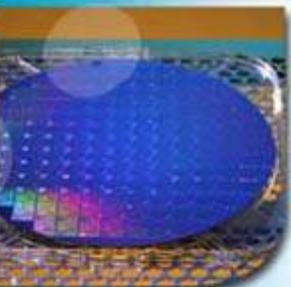
- Limites théoriques et pratiques
- Les nouvelles architectures
- Un grand nombre de nouveaux dispositifs
- Quelles architectures pour les nanocomposants
- **Tentative de conclusion**

- 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

micro et nanoélectronique  
microsystèmes  
intelligence ambiante  
biologie et santé  
chaîne de l'image



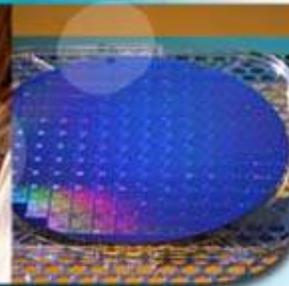
Loyauté  
Envie d'entreprendre  
Travail en équipe  
Loyauté Innovation  
Envie d'entreprendre  
Travail en équipe  
Innovation



leti



micro et nanoélectronique  
microsystèmes  
intelligence ambiante  
biologie et santé  
chaîne de l'image



Prévoyez de nous rejoindre à la  
**10<sup>e</sup> Annual Review**  
du Leti du 24 au 25 juin 2008  
à Minatec

Pour en savoir plus :  
[www.leti.fr](http://www.leti.fr)

leti

MINATEC

INSTITUT  
CARNOT  
CEA LETI

cea

