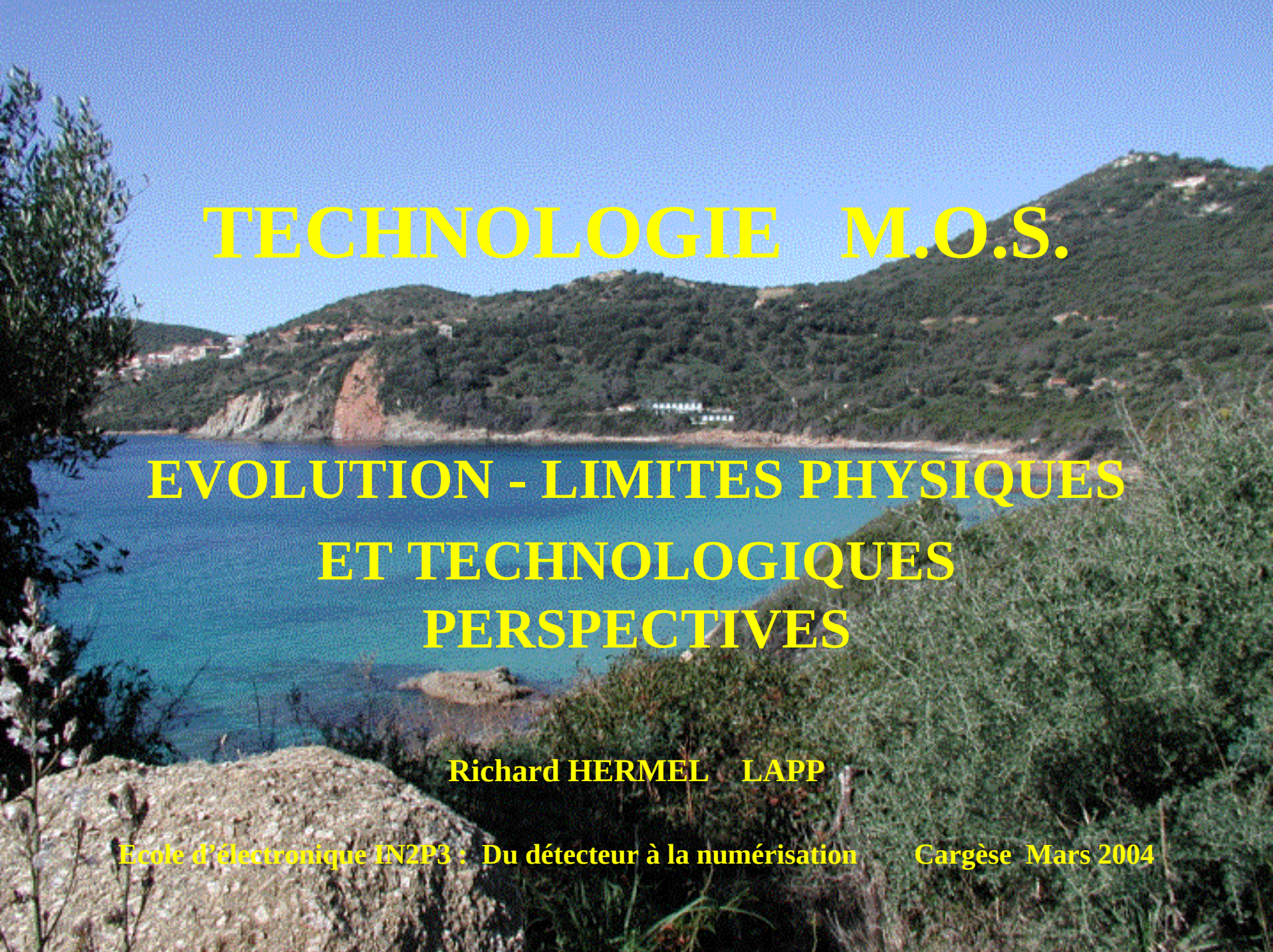


TECHNOLOGIE M.O.S.

EVOLUTION - LIMITES PHYSIQUES ET TECHNOLOGIQUES PERSPECTIVES

Richard HERMEL LAPP

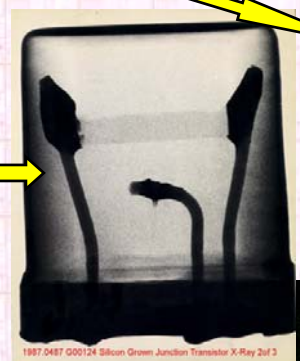
Ecole d'électronique IN2P3 : Du détecteur à la numérisation Cargèse Mars 2004

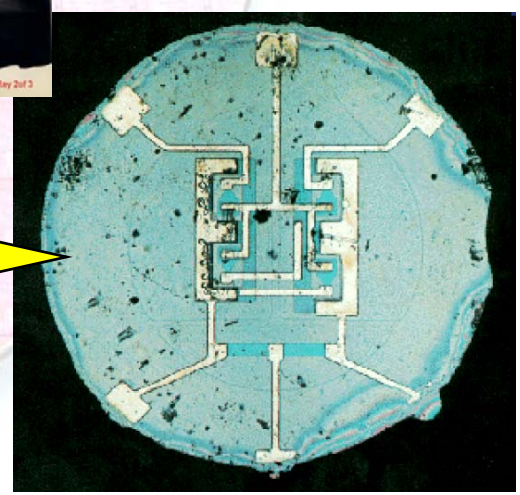
Sommaire

- Introduction, historique
- Le transistor M.O.S.
- Limites physiques et technologiques
- technologies alternatives
- perspectives

Historique (1)

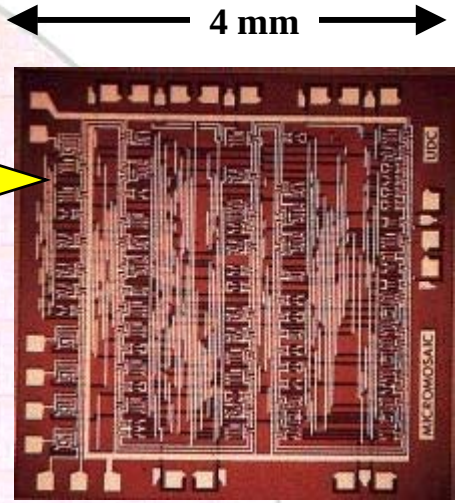
- 1940 : Ohl ⇒ jonction PN
- 1949 : Transistor à pointes à 2 jonctions
  Brattain, Bardeen **(Nobel 56)**
- 1951 : Shockley Transistor à jonctions « industriel » **(Nobel 56)**
- 1954 : 1er récepteur à transistor
  1er transistor silicium (TI, 2.5$)
- 1958 : Technique de la diffusion + SiO_2
- 1959 : 1er circuit intégré monolithique
  Kilby **(Nobel 2000)**, Noyce
- 1961 : 1er circuit intégré logique (TI, Fairchild) : double bascule à 4 transistors (50$, 1,5x1,5 mm^2)





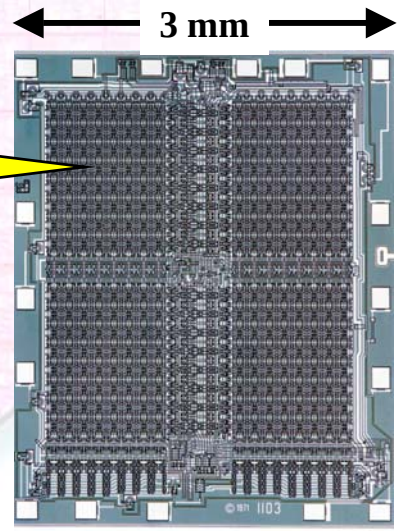
Historique (2)

- 1967 : 1er « semi-custom » (pré diffusé)
  2 niveaux d'interconnections, 150 portes



- 1968 : Création d'Intel (Moore et Noyce)

- 1970 : Mémoire RAM 1 kBits

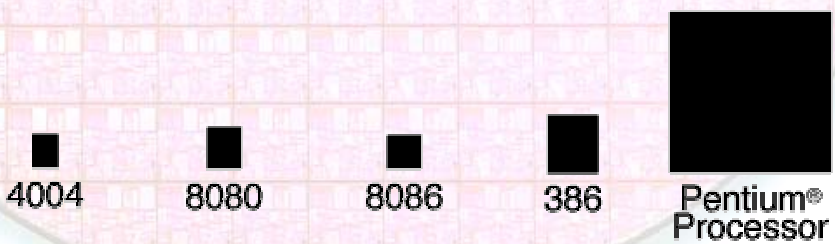


- 1971 : 1er microprocesseur : Intel 4004

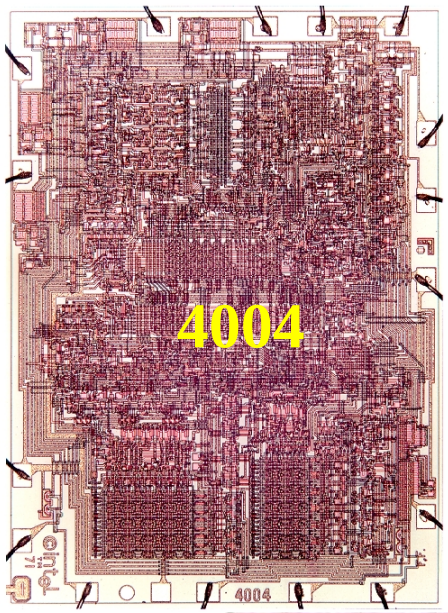
Historique des microprocesseurs INTEL performances - tailles respectives du chip

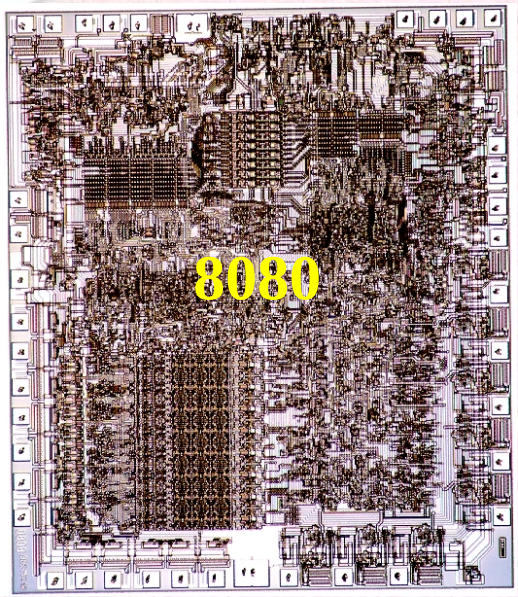
Microprocesseur	4004	8080	8086	i386	Pentium	Pentium 4
Année	1971	1974	1978	1985	1993	2000
Nb. Bits	4	8	16	32	64	64
Horloge (Hz)	108k	2M	10M	33M	66M	1.5G
Mémoire adressable (bytes)	640	64K	1M	16M	4G	64G
Technologie (µm)	10	6	3	1	0.8	0.18
Nb transistors	2300	6000	29000	275000	3.1M	42M
Tension alim (V)	12	12	5	5	5/3.3	1.3 interne

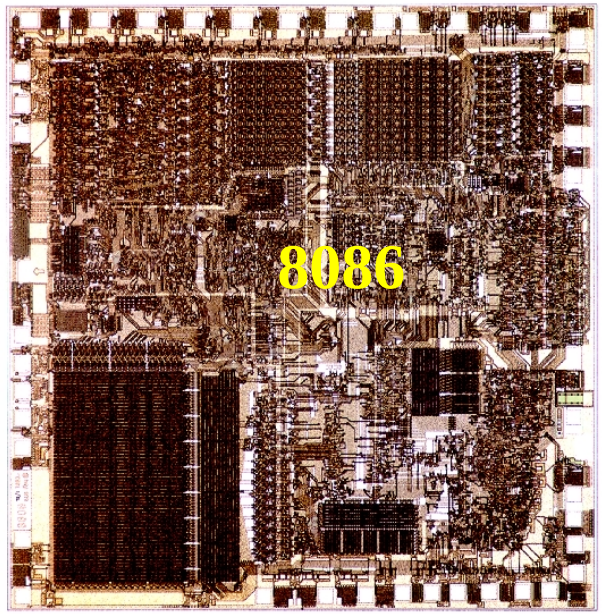
Source : www.intel.com

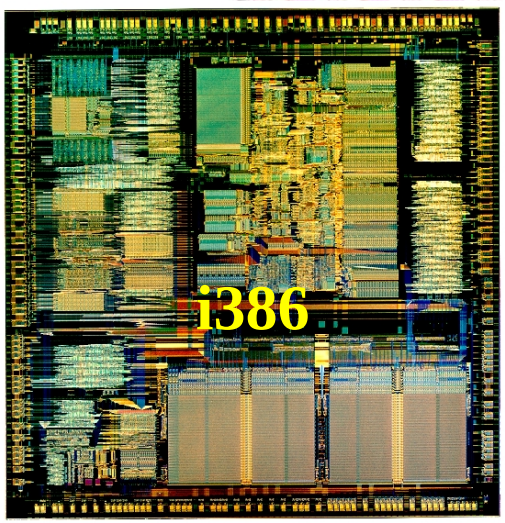


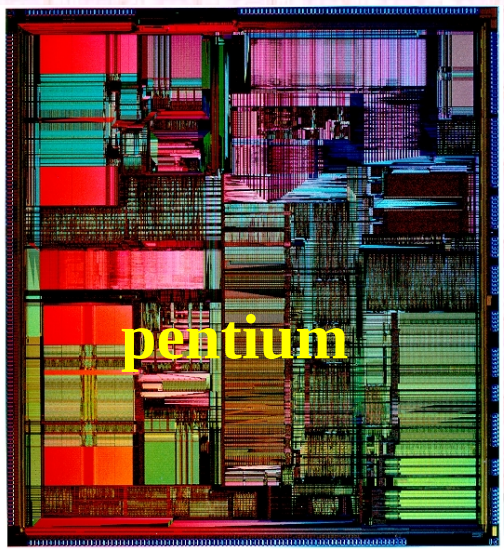
Microprocesseurs INTEL

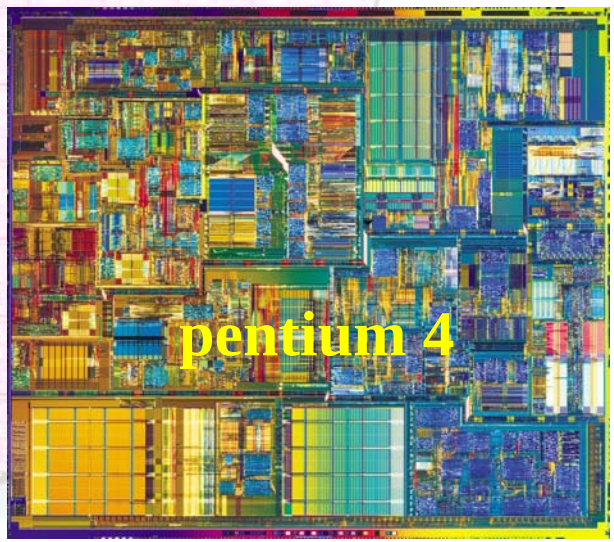




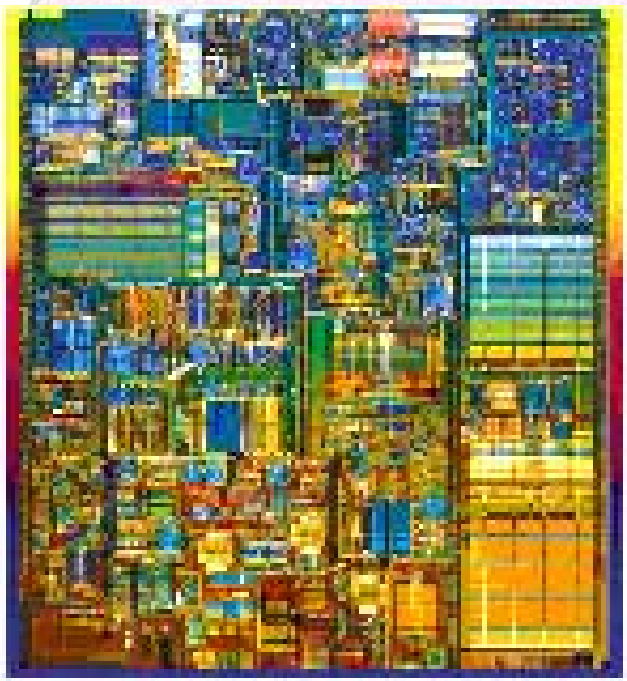








Le même circuit (pentium 4) en 2 générations technologiques successives



180 nm Technology

130 nm Technology

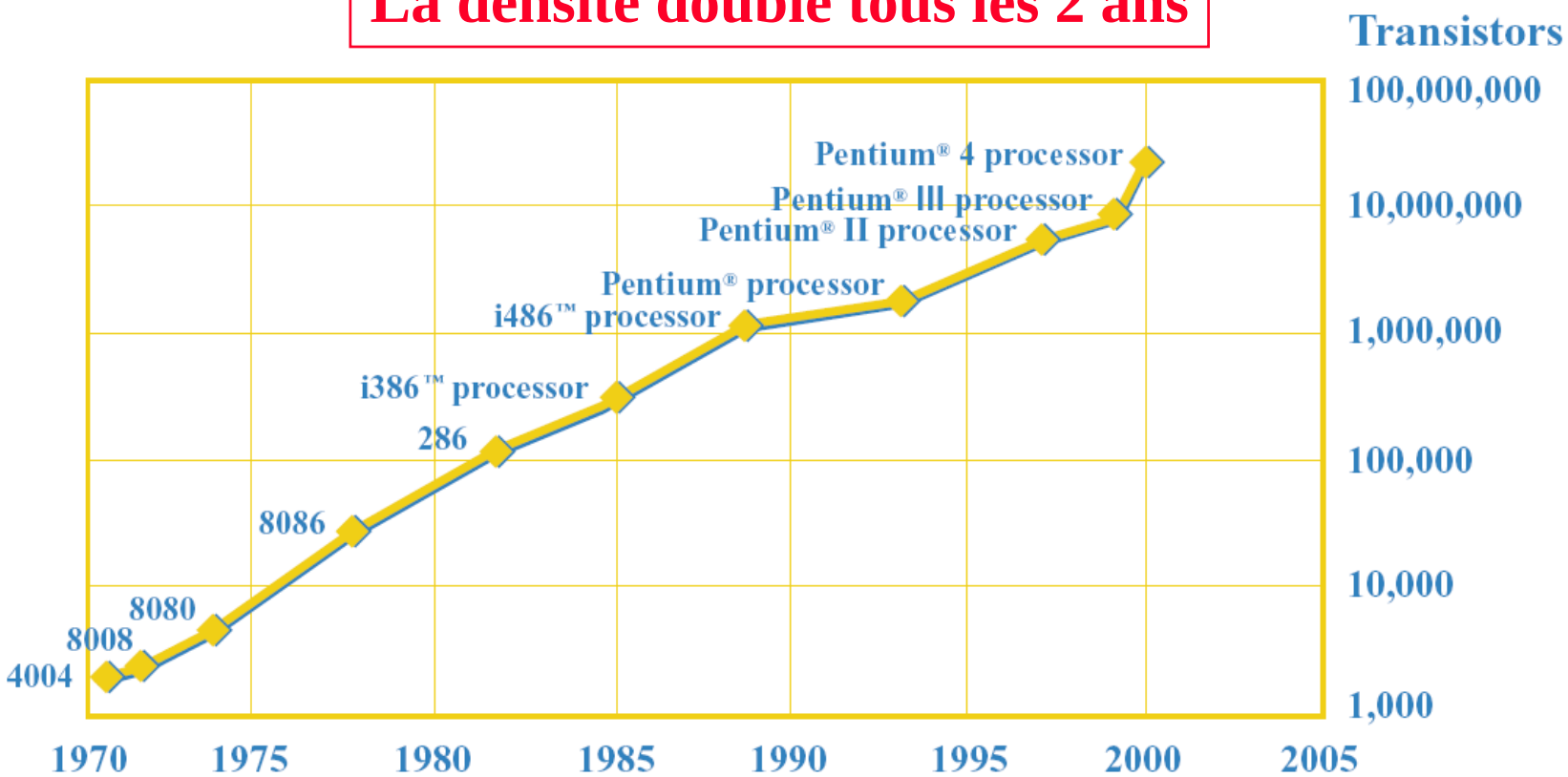
Evolution depuis 1970 : résumé

Grandeurs	1970	2000
Inflation en Francs	1	5.5
Prix de 10^6 transistors	10 000 $	1 $
Marché (y compris discrets)	2 G$	226 G$
Taille d'un MPU	12 mm^2	200 mm^2
Prix d'un MPU	125 $	200 $
Taille de gravure	0.01 mm	0.15 µm
Nombre de couches	1	8
Transistors/puce	2 300	64 000 000
Fréquence	200 kHz	1 GHz
Coût d'une usine	10 M$	1 500 M$
Diamètre de wafer	1 pouce	12 pouces
Nombre d'opérations de fabrication	20	800

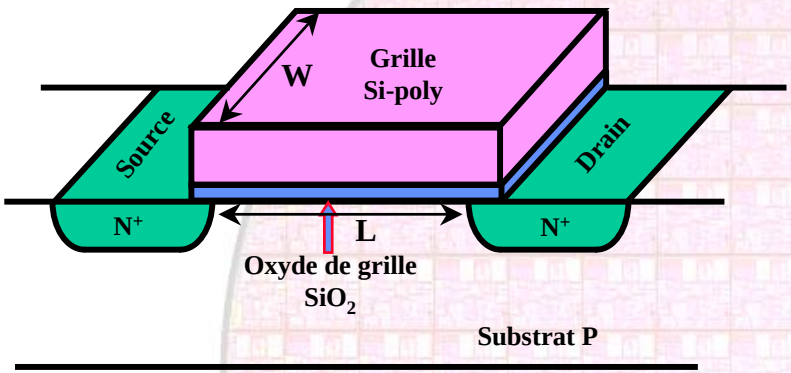
Source : Rapport microélectronique B. Legait et al. Avril 2001

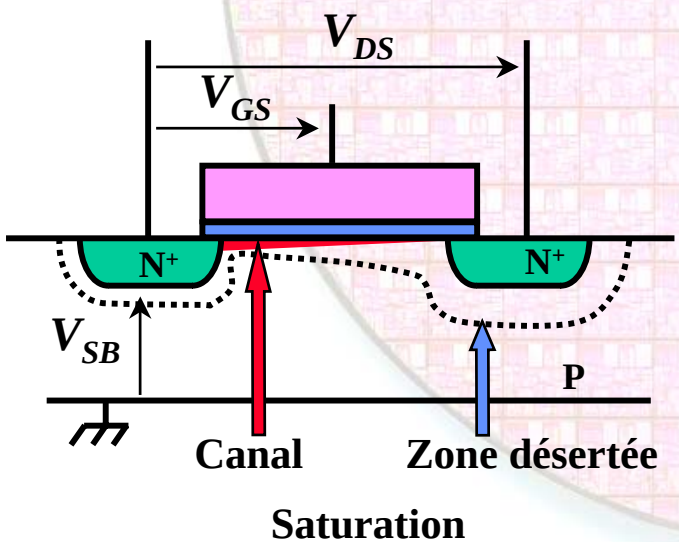
Evolution depuis 1970 : « loi » de Moore

La densité double tous les 2 ans



Le transistor M.O.S. : fonctionnement





Saturation

- *W, L* : largeur et longueur du canal
- Création du canal :

  injection d'électrons sous la grille par la mise en direct *localement* de la jonction source-substrat, contrôle de la quantité de porteurs par le *champ électrique vertical* créé par la tension V_{GS}

- Mise en mouvement des porteurs :

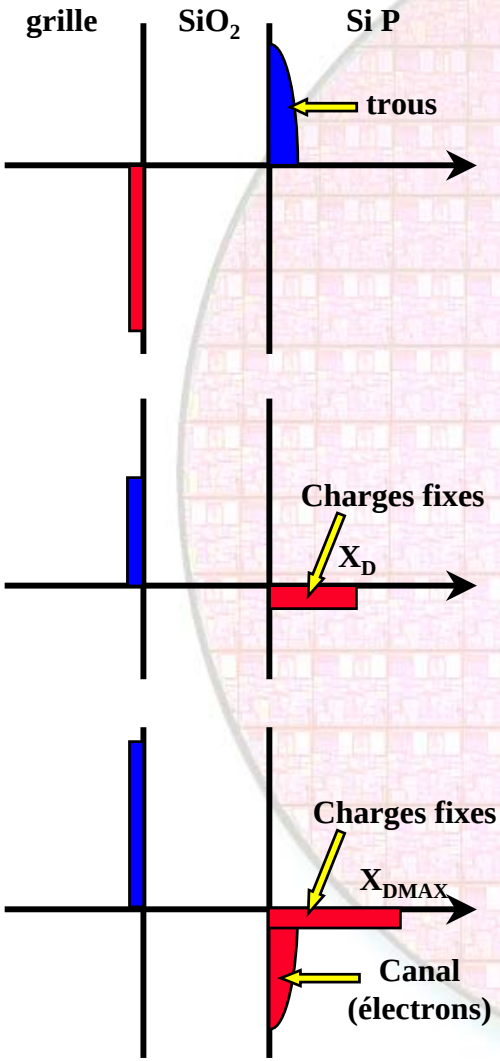
  par le *champ électrique longitudinal* créé par la tension V_{DS}

- V_T : Tension de seuil

  ⇩

  limite de l 'inversion :
  concentration des porteurs libres dans le canal = concentration de dopant du substrat

Tension de seuil (canal N)



Répartition des charges dans une structure MOS

charges positives

charges négatives

$$V_{T0} = V_{FB} + 2\Phi_{FP} + \gamma\sqrt{2\Phi_{FP}}$$

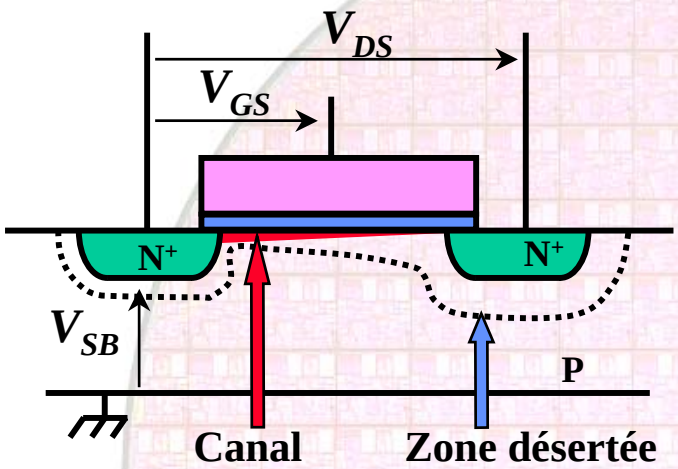
V_{FB} = **tension de « bandes plates »**

Φ_{FP} = **potentiel de Fermi des porteurs libres du substrat**

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_B}}{C_{OX}}$$

Coefficient d'effet de substrat

Le transistor M.O.S. : courant



μ : mobilité des porteurs

C_{OX} : Capacité d'oxyde par unité de surface

N_B : Dopage substrat

ε_{Si} : permittivité Si

q : charge élémentaire

$$I_{DS} = \mu\ C_{OX}\ \frac{W}{L}\left[(V_{GS} - V_T)\ V_{DS} - \frac{V_{DS}^{\ 2}}{2}\right]$$

en inversion forte

$$\beta = \mu\, C_{OX}\ \frac{W}{L}$$

$$V_T = V_{T0} + \gamma\left(\sqrt{V_{SB} + 2\Phi_{FP}} - \sqrt{2\Phi_{FP}}\right)$$

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_B}}{C_{OX}}$$

Coefficient d'effet de substrat

$$I_{DSAT} = \frac{\beta}{2}(V_{GS} - V_T)^n \qquad 1 < n < 2$$

en saturation ⇨ utilisation dans la circuiterie analogique

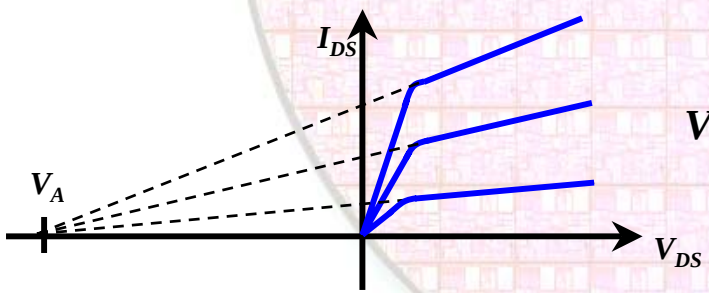
Le transistor M.O.S. : petits signaux

$$I_{DS} = f(V_{GS}, V_{DS}, V_{SB})$$

$$\frac{\partial I_{DS}}{\partial V_{GS}} = g_m = \sqrt{2\beta I_{DS0}} \quad : \text{transconductance}$$

$$\frac{\partial I_{DS}}{\partial V_{DS}} = g_{ds} = \frac{I_{DS0}}{V_A} \quad : \text{conductance drain/source,} \quad V_A = \text{tension d'Early}$$

$$\frac{\partial I_{DS}}{\partial V_{SB}} = g_{mb} = -g_m \frac{\gamma}{2\sqrt{V_{SB} + 2\Phi_{FP}}} \quad : \text{transconductance due à l'effet de substrat}$$



Effet de la modulation de longueur de canal

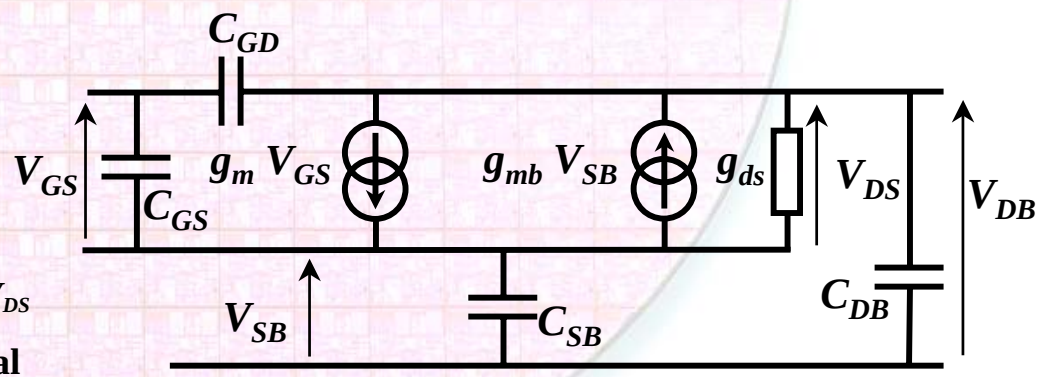
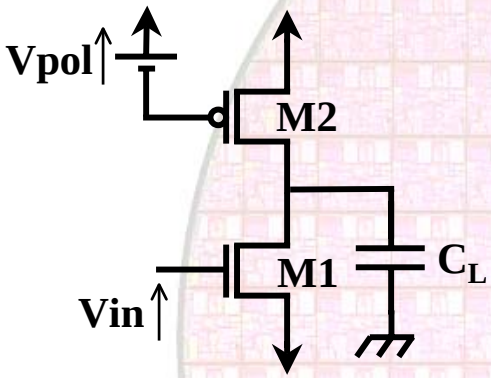


Schéma équivalent linéaire

Vitesse des circuits CMOS

Analogique

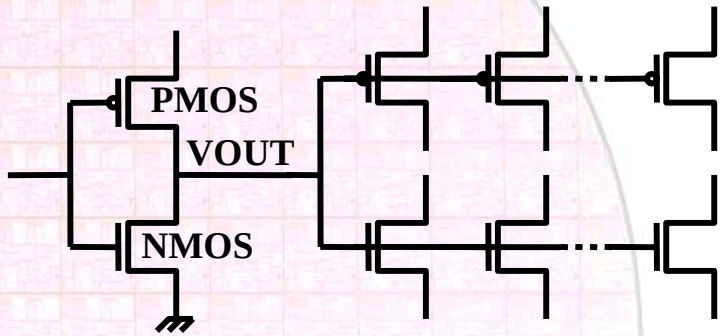


$$A_V = -\frac{g_{m1}}{g_{s1}+g_{s2}} \qquad f_T = \frac{g_{m1}}{2\pi C_L}$$

$$A_V \propto V_A(L) \qquad f_T \propto \frac{W}{L}\frac{1}{C_L}$$

Attention !
Diminuer *L* entraîne la diminution de V_A

Digital

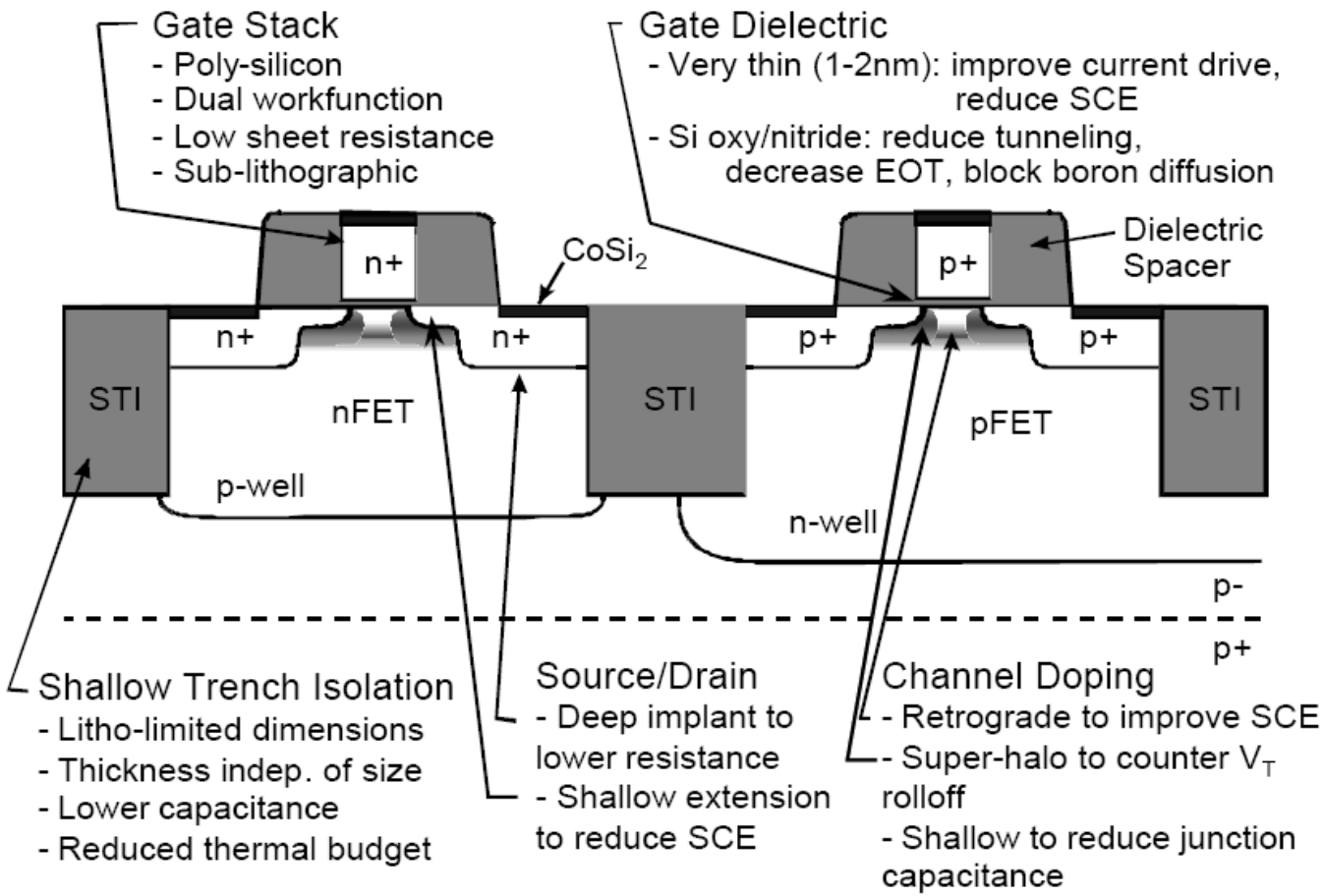


N étages

$$\frac{dV_{OUT}}{dt} = \frac{I_{DS0}}{C} \propto \frac{\frac{W}{L}}{WL} = \frac{1}{L^2}$$

La vitesse d'une technologie varie comme le *carré* de la longueur de canal

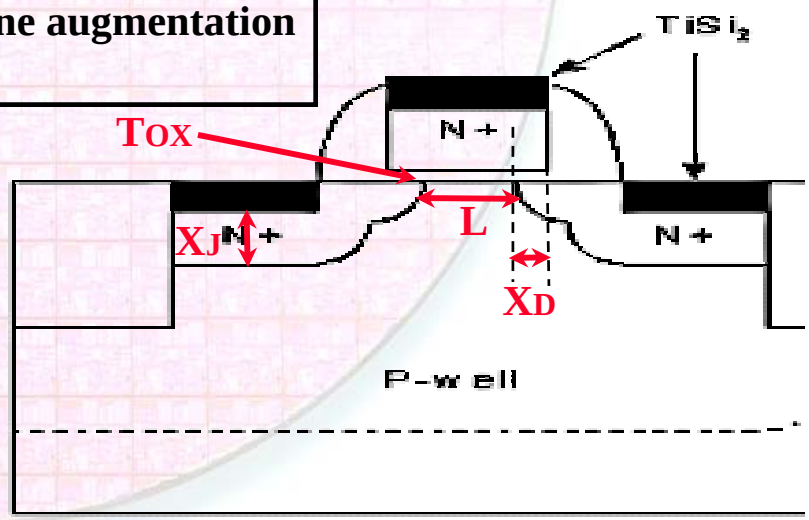
Technologie CMOS : IBM



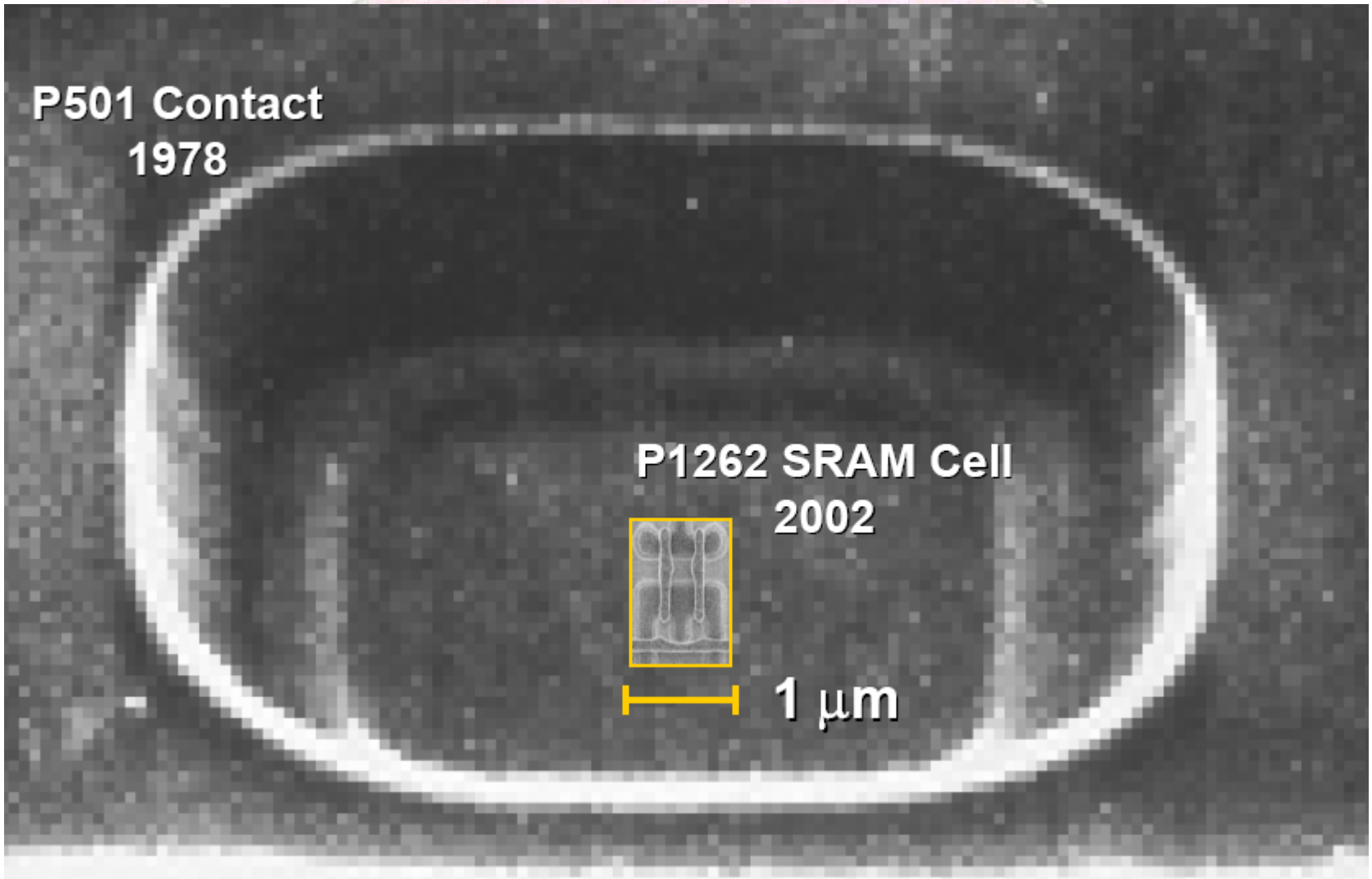
[H.-S. P. Wong, et al., Proc. IEEE, **87**, p.537, 1999 and S.-F. Huang, et al., IEDM Tech. Dig., p.237, 2001.]

Dimensions intéressantes à réduire

Grandeur	Effet
Epaisseur d'oxyde (Tox)	Augmentation de la vitesse Augmentation de C_{ox} Augmentation de la transconductance
Longueur de canal (L)	
Profondeur de jonction (X_J)	Réduction des éléments parasites
Extension source/drain (X_D)	
La réduction de toutes les dimensions entraîne une augmentation de la densité d'intégration	



Réduction des dimensions : 1978……2002



Limites physiques et technologiques

Grandeur	limite	raison
Epaisseur d'oxyde	23Å	Effet tunnel
Longueur de canal	0.06µm	Courant de fuite
Longueur de grille	0.1µm	Courant de fuite
Dopage (tension de seuil)	V_T=0.25V	Courant de fuite
Profondeur de jonction	30nm	résistance
Extension source/drain	15nm	résistance

Réduction de l'épaisseur de l'oxyde de grille (1)

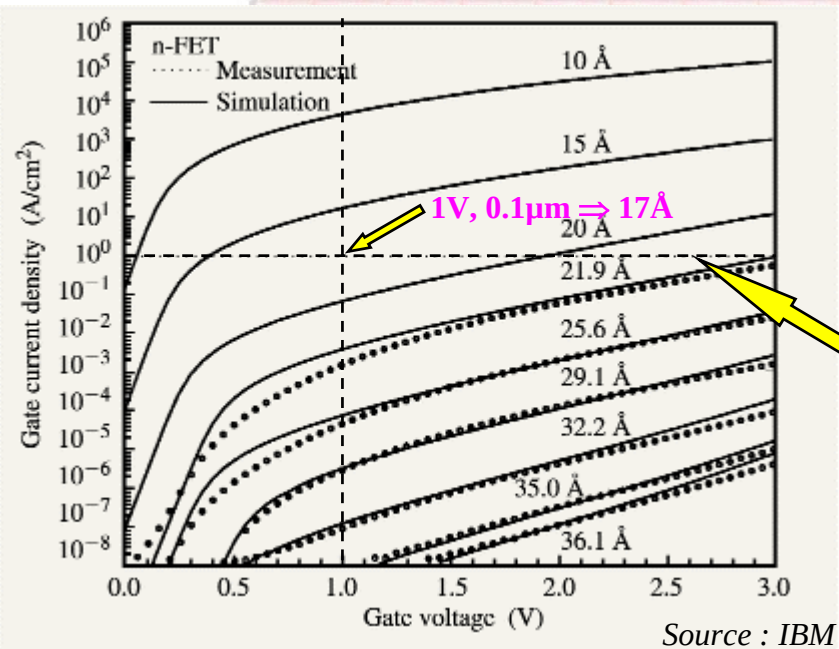
Limite : courant de grille dû à l'effet tunnel à travers l'oxyde de grille

le courant varie de façon exponentielle avec l'épaisseur d'oxyde :

1 pA/cm² à 35Å et 1 A/cm² à 17Å :

12 ordres de grandeur de variation

du courant pour une variation d'épaisseur d'un facteur 2 !!!



Source : IBM

On considère que la limite acceptable est atteinte lorsque le courant de grille est égal au courant de fuite.
La valeur de 1nA/µm de largeur est une moyenne pour les technologies actuelles

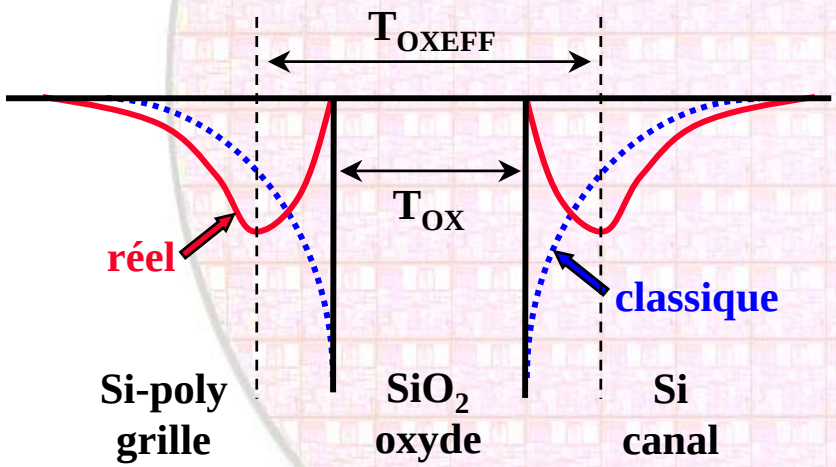
1nA/µm = 1A/cm²
pour L = 0.1µm

Réduction de l'épaisseur de l'oxyde de grille (2) épaisseur d'oxyde effective

Effets quantiques et effets de désertion dans le Si-poly de la grille

⇩

augmentation de l'épaisseur effective par rapport à l'épaisseur physique



Répartition des charges dans une structure MOS en inversion

L'augmentation de T_{OX} due à ces 2 effets est de 0.6 nm environ

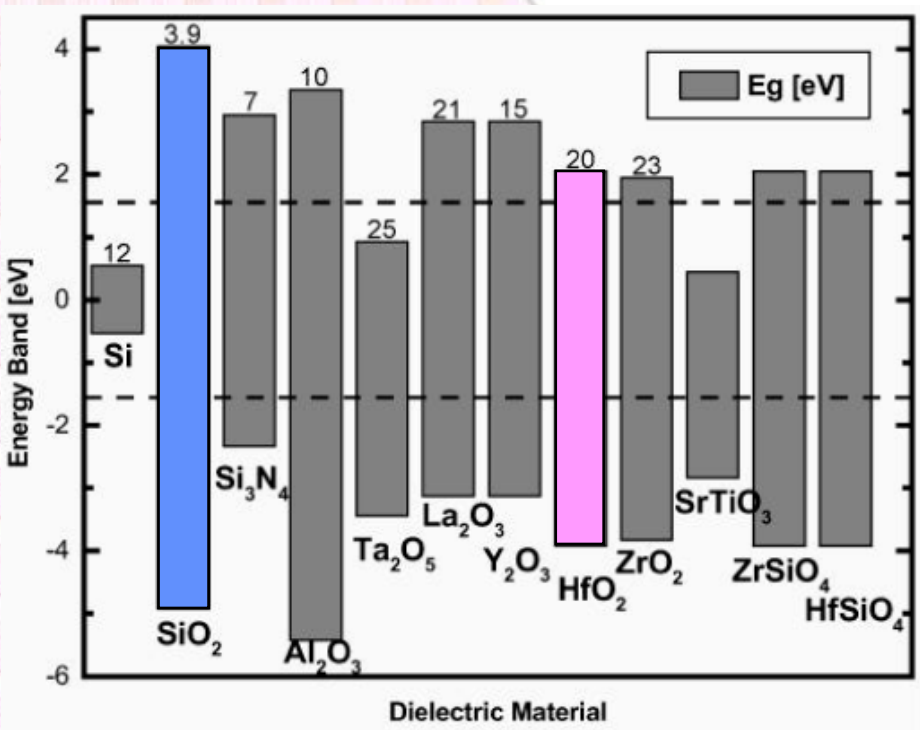
⇩

T_{OX} min ≈ 23 Å

Réduction de l'épaisseur de l'oxyde de grille (3) perspectives

Année de production	Longueur de grille (µm)	Epaisseur d'oxyde effective (Å)
1997	0.25	40-50
1999	0.18	30-40
2001	0.15	20-30
2003	0.13	20-30
2006	0.10	15-20
2009	0.07	<15
2012	0.05	<10

Source : SIA

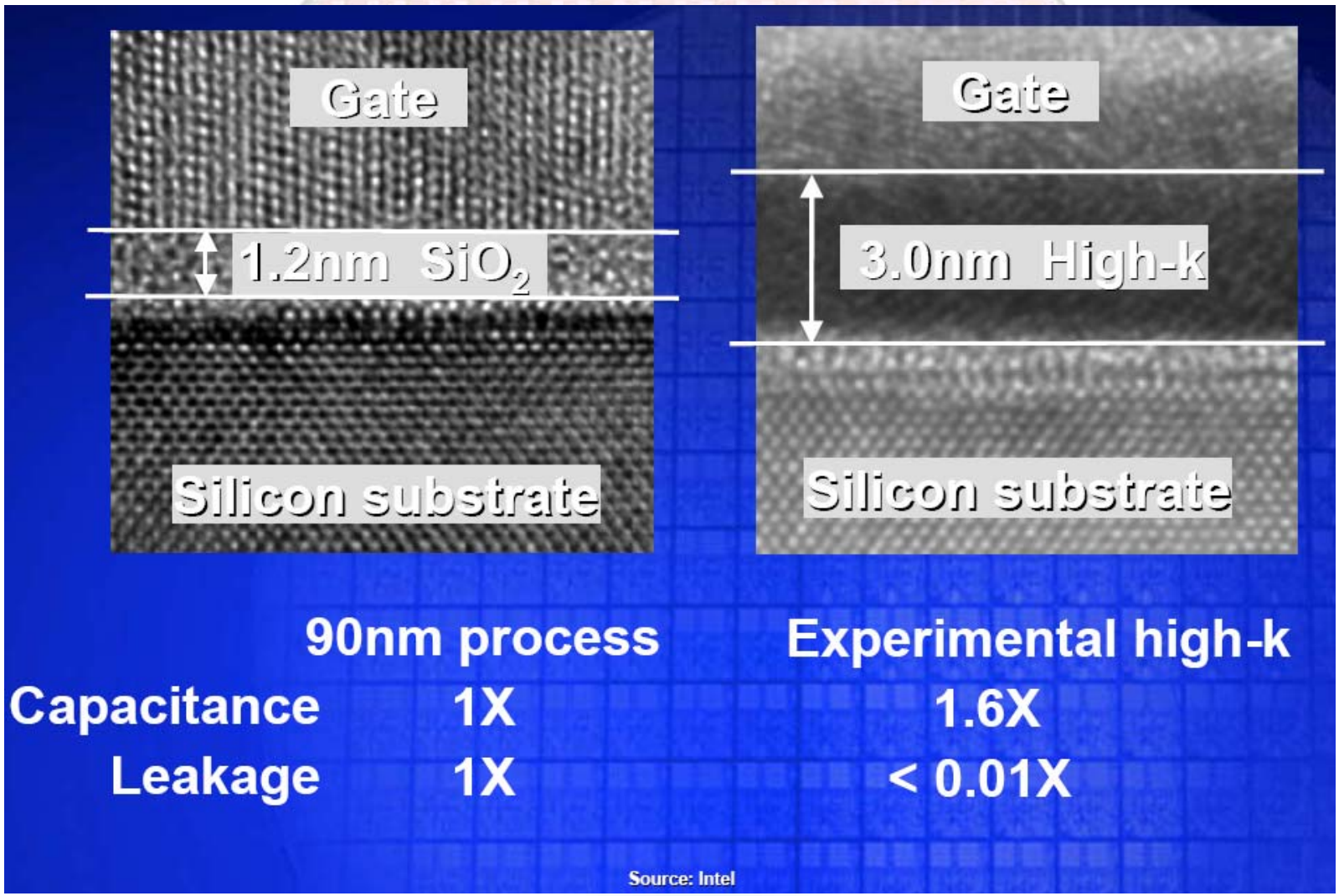


[H.-S. P. Wong, IBM J. Res. Devel., **46**, March/May 2002.]

Les matériaux utilisés doivent être compatibles avec la filière technologique et conserver un bon comportement en hautes fréquences

On peut espérer gagner 5 ordres de grandeur sur le courant de fuite en gardant les mêmes performances au niveau du canal

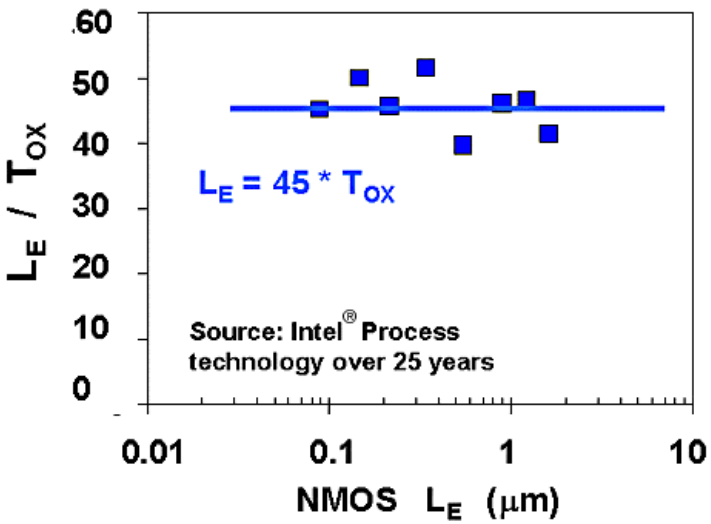
Utilisation d'isolants à haute permittivité

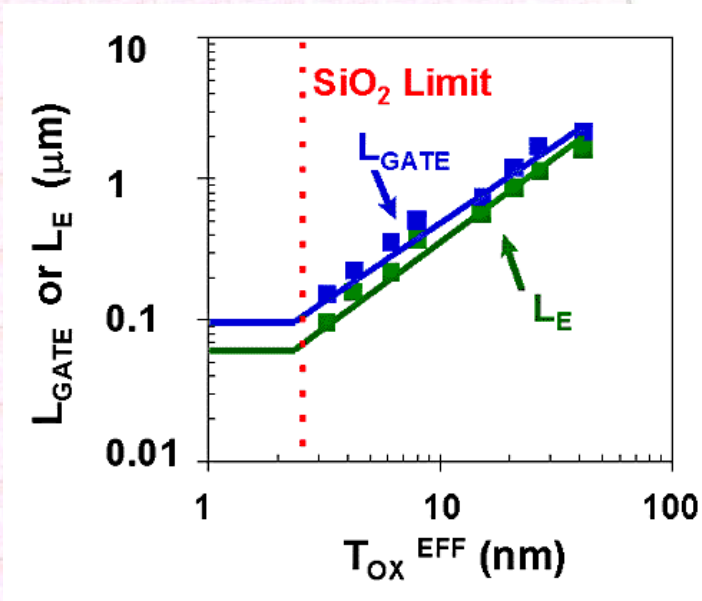


Couplage L/T_{OX}

- **L ↘ ⇒ T_{OX} doit diminuer pour garder un bon contrôle du canal par la grille :**

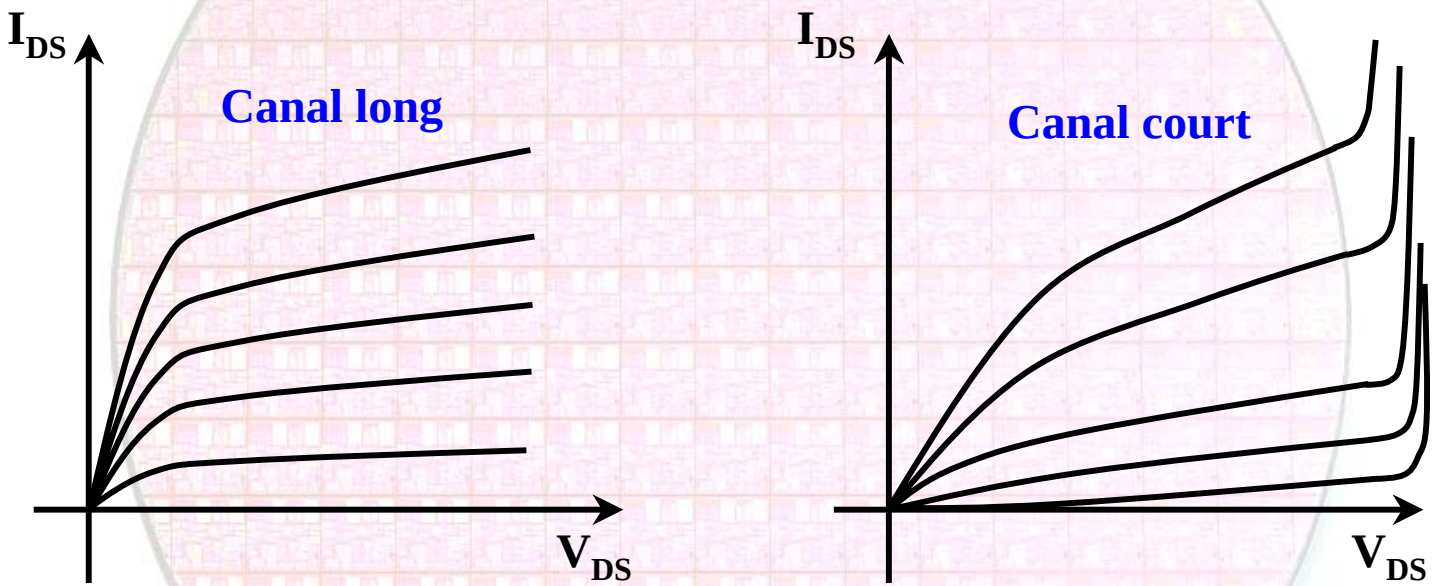
  L'épaisseur du diélectrique du condensateur grille/oxyde/canal doit être petite devant les autres dimensions (W,L)





Le rapport L/T_{OX} est « à peu près » constant pour des technologies apparues au cours des 25 dernières années

Caractéristiques IDS =f(VDS)
canaux longs - canaux courts



Le comportement de type « canal long » ou « canal court » est un comportement électrique et ne dépend pas uniquement de la longueur physique du canal

Réduction de la longueur de canal

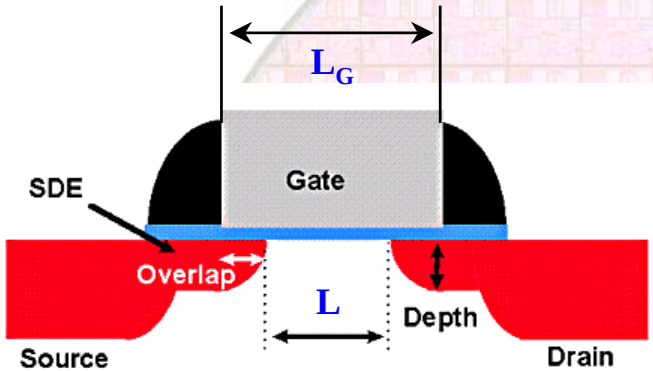
- Effets indésirables des canaux courts
  - Saturation de la vitesse des porteurs

    $ID_{SAT} \propto V_{DSAT}$ au lieu de V_{DSAT}^2
  - Augmentation du courant de blocage

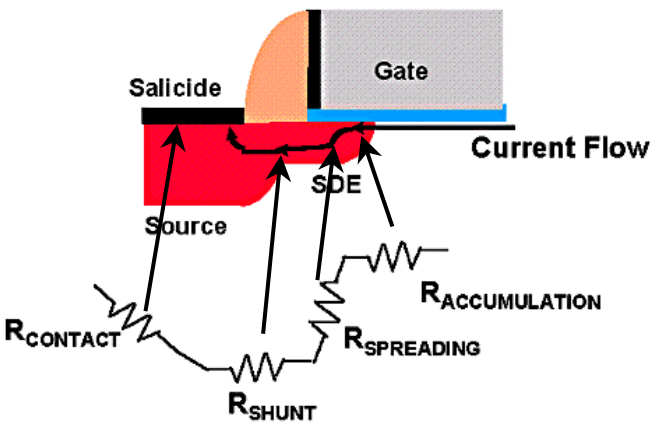
    remède : optimisation du dopage du substrat
  - Augmentation de la tension de seuil
- Les effets dus aux canaux courts sont gouvernés par le rapport entre l'épaisseur de la zone désertée sous le canal et la longueur du canal

On a intérêt à avoir la zone désertée la plus mince possible pour conserver un comportement électrique de type « canal long » tout en réduisant sa longueur physique

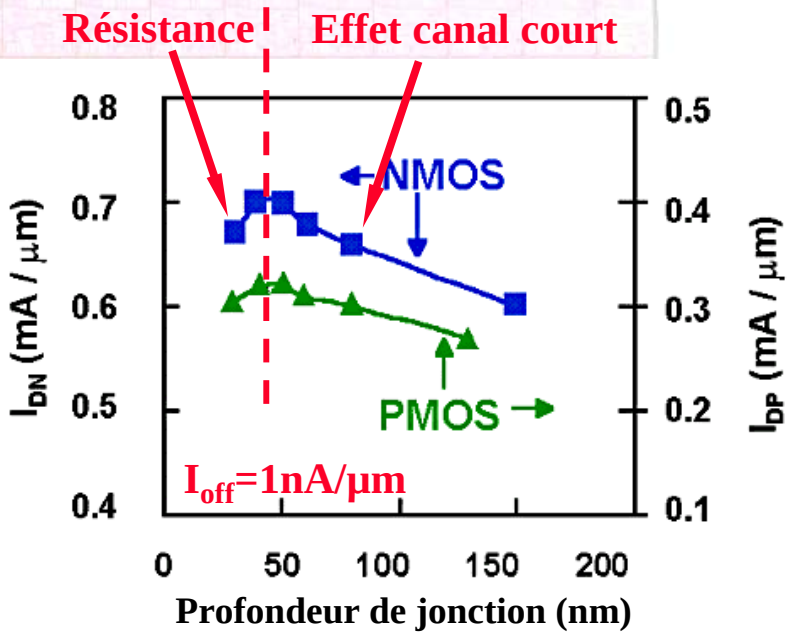
Réduction de la profondeur de jonction



SDE : Source/drain extension
overlap : recouvrement de la grille au dessus de D/S indispensable
L : longueur de canal
L_G : longueur de grille

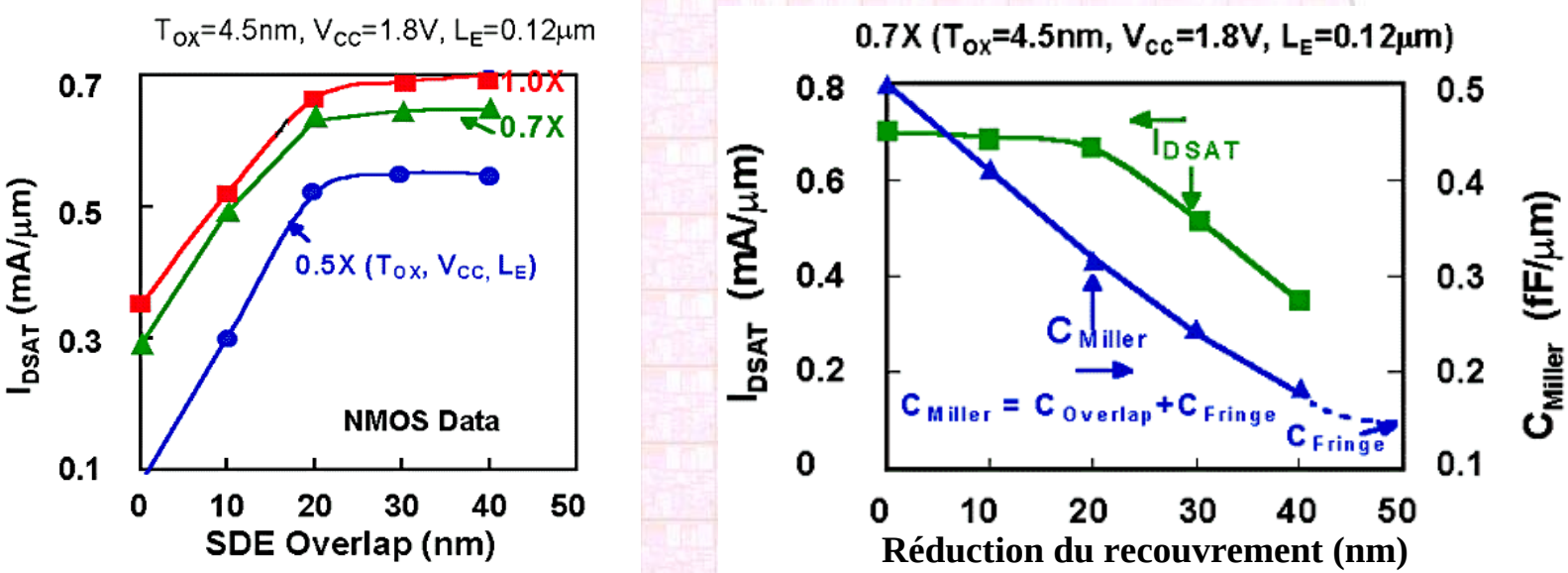


Résistances d'accès au canal



Recouvrement grille/source et grille/drain

- **Le recouvrement de la grille au-dessus du drain et de la source est indispensable au fonctionnement du transistor mais il introduit une capacité parasite.**
- **La réduction de ce recouvrement diminue cette capacité mais augmente la résistance d'accès au canal**

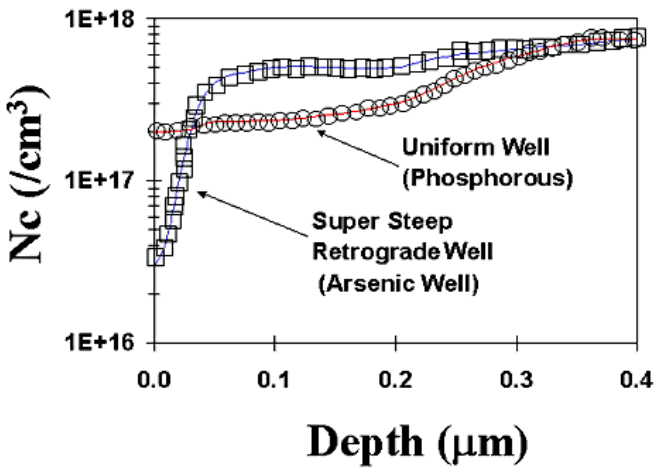


Compromis entre la réduction de la capacité parasite et l'augmentation de la résistance

Optimisation du profil de dopage sous le canal

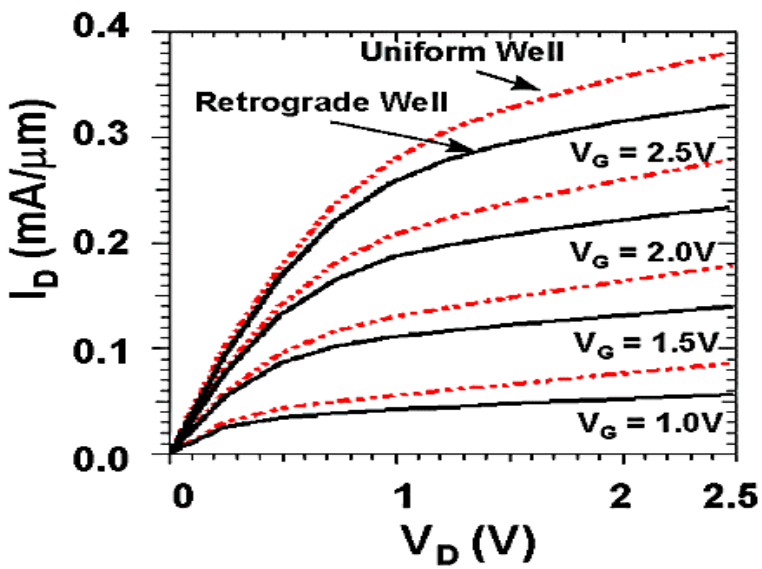
- **But :**

  minimiser le courant de blocage et maximiser le courant de saturation

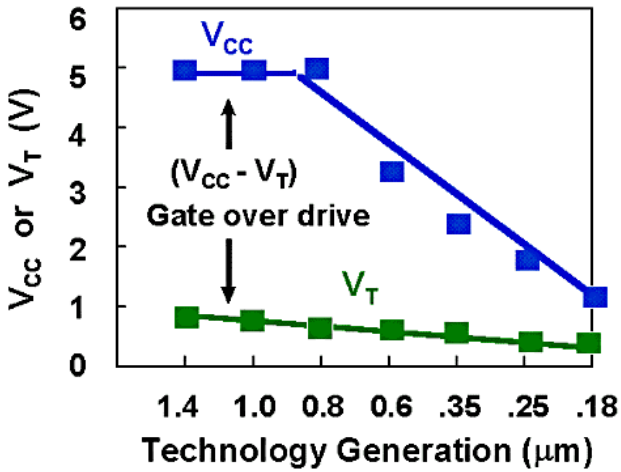


Idée : augmenter le dopage substrat loin de la surface pour améliorer le comportement en canaux courts et diminuer le dopage près de la surface pour augmenter la mobilité des porteurs dans le canal

Effet négatif : augmentation de la tension de seuil
⇒ diminution de ID_{SAT}

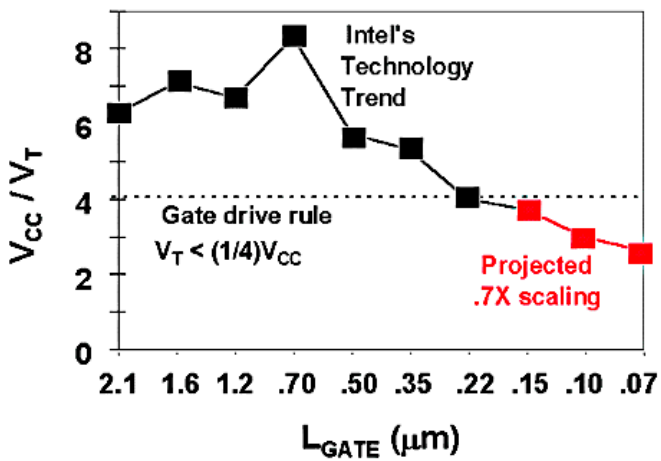


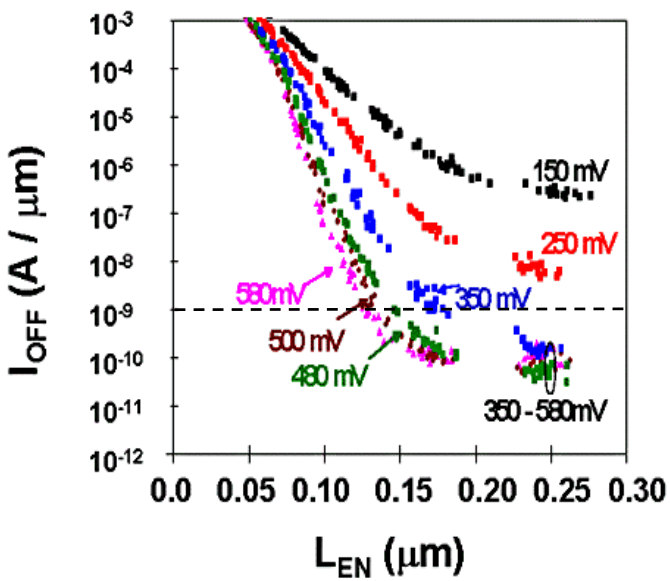
Conséquences pour la circuiterie (1) tension de seuil



$L \searrow \Rightarrow V_{CC} \searrow \Rightarrow V_T$ doit $\searrow$
pour garder I_D le plus élevé possible :
$I_D \propto (V_{GS} - V_T)^n$ $\quad n1<n<2$

mais $V_T \searrow \Rightarrow I_{OFF} \nearrow$
$\Downarrow$
compromis : $V_T > 0.25V$





Conséquences pour la circuiterie (2) tension de seuil

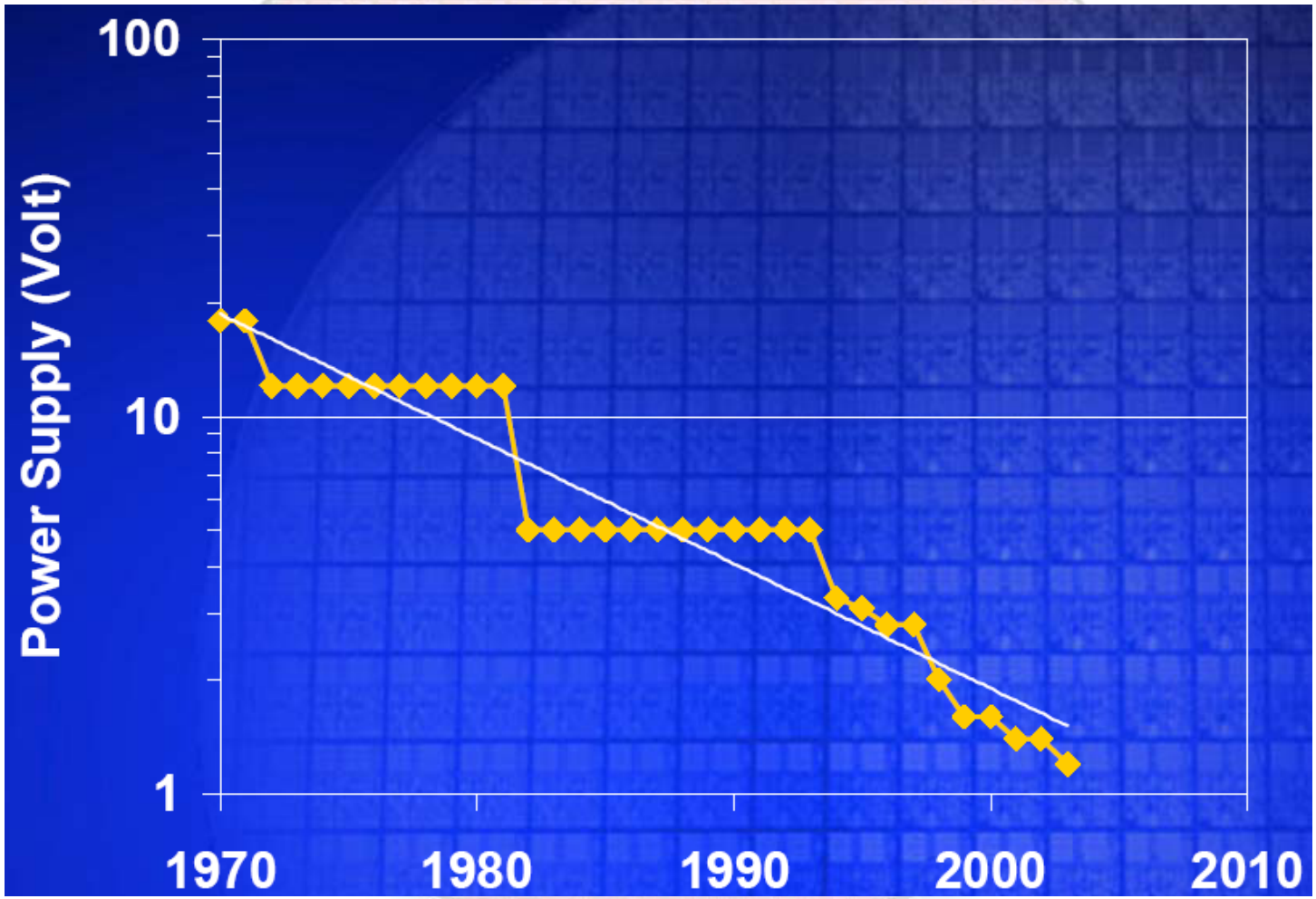
- **Architecture à deux tensions de seuil**

  Sur le même chip :

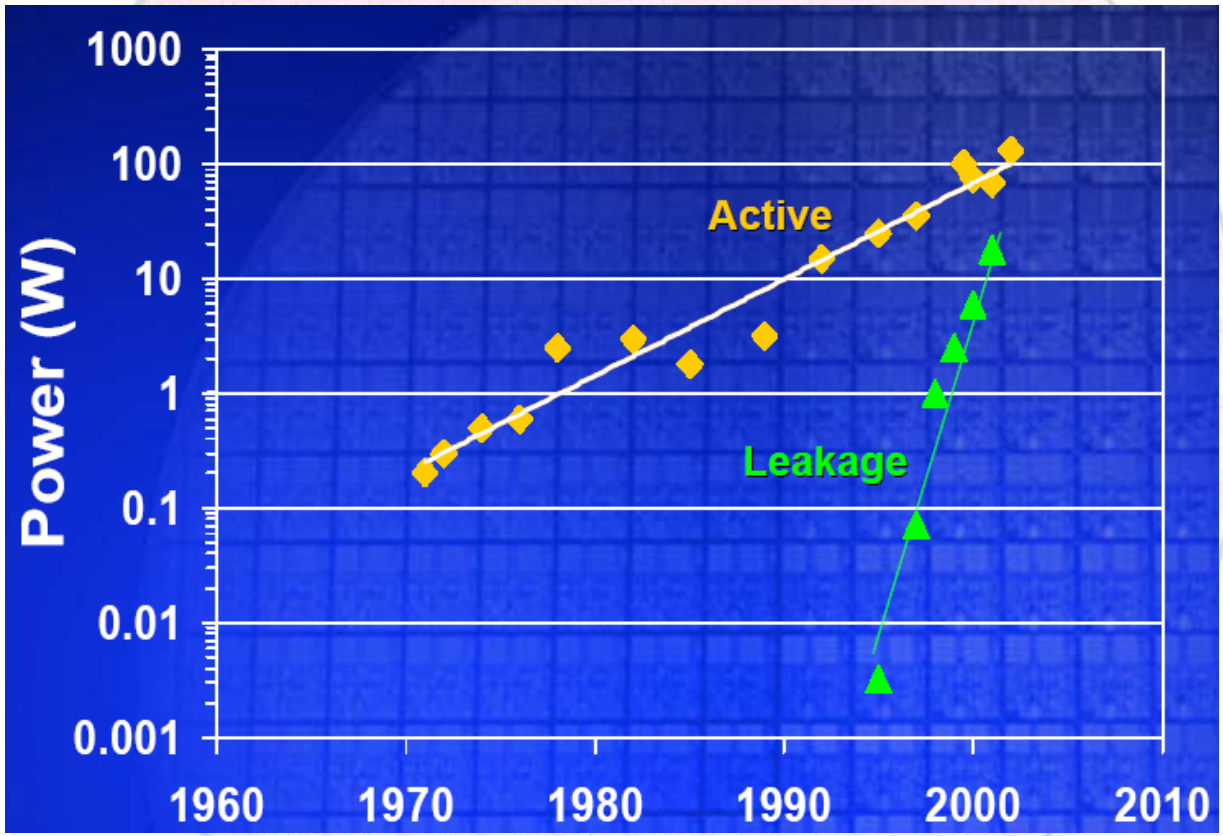
  - **Des transistors à basse tension de seuil, donc à courant de fuite élevé pour les chemins critiques du circuit**
  - **Des transistors à tension de seuil élevée, à bas courant de fuite et basse consommation pour le reste du circuit**

La proportion entre les 2 types de composants dépend fortement de l'architecture du circuit et des contraintes de puissance

Evolution des tensions d'alimentation



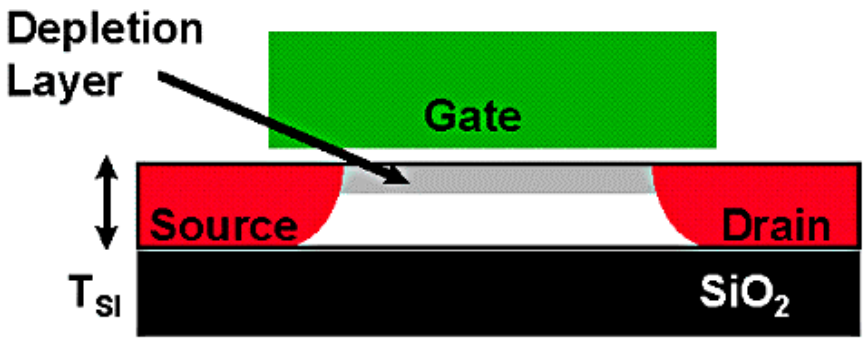
Conséquences pour la circuiterie (3) puissance dissipée



Avec la réduction des dimensions, la puissance dissipée par chip augmente et l'écart avec la puissance au repos se réduit

Technologies alternatives

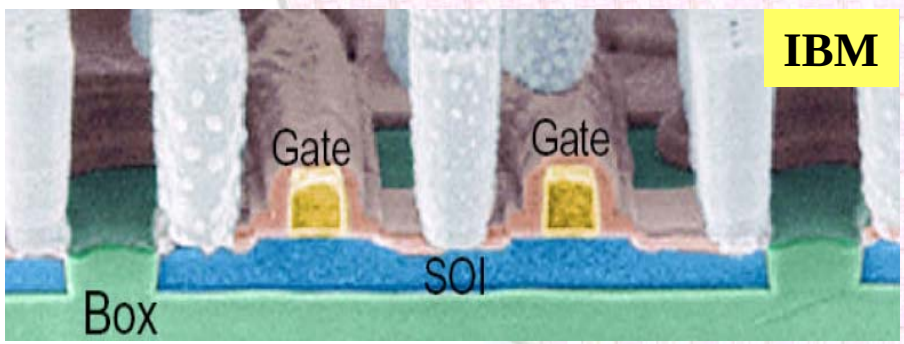
- **Silicium sur isolant (SOI)**



MOS sur isolant
structure partiellement désertée

Parameter	Best Case Gain
Junction Capacitance	12%
Body Factor	3%
Gate-to-Body Coupling	3%
Channel Length	0%
Total	18%

Réduction des éléments parasites



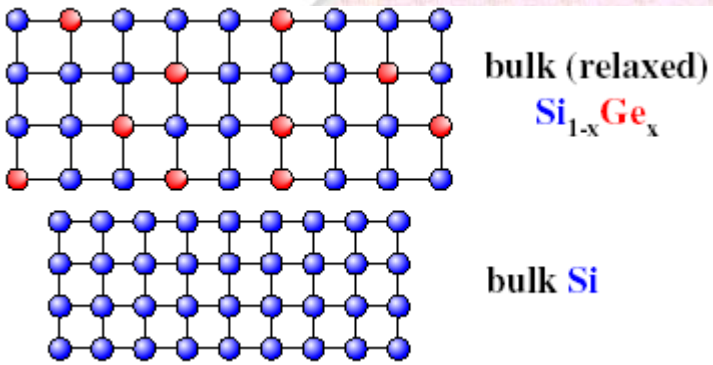
Attention ! La réduction des effets parasites s'accompagne d'effets indésirables dus au substrat flottant, en particulier une augmentation du courant de fuite

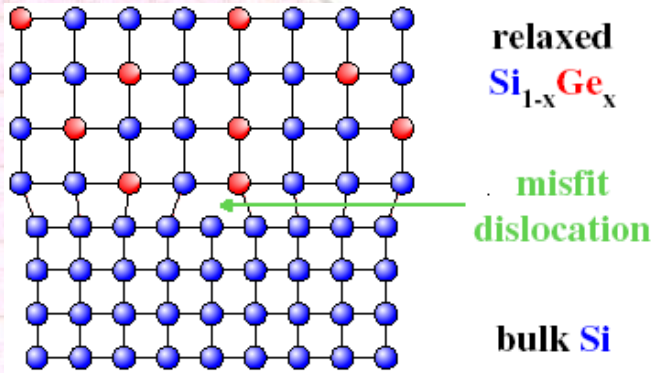
C'est une solution pour la tenue aux radiations

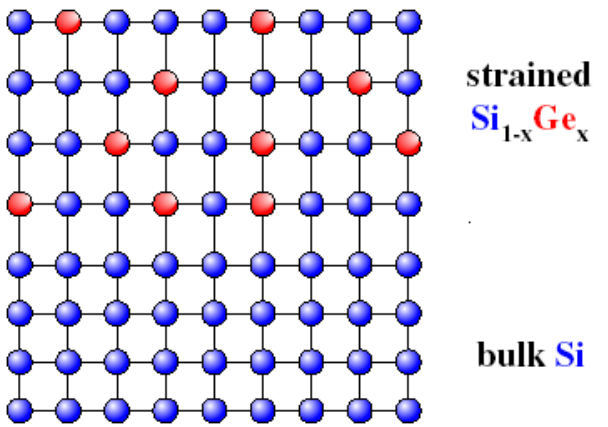
Technologie SiGe

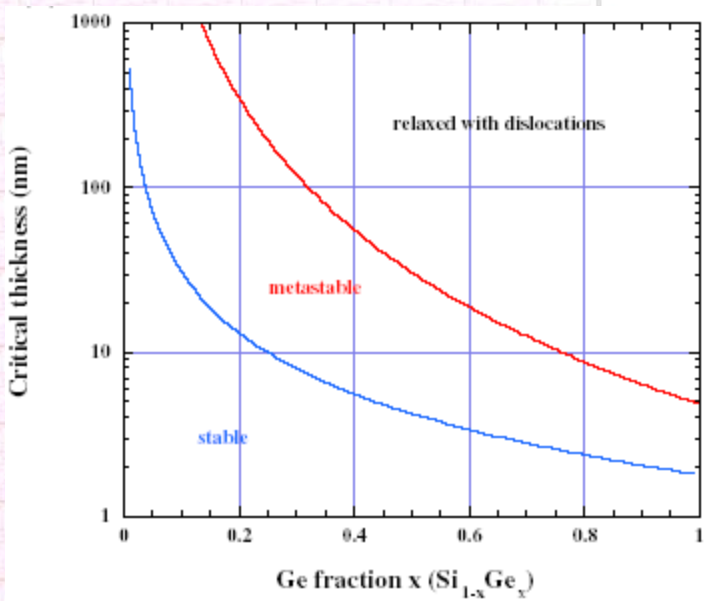
- Création de transistors bipolaires à hétérojonction (HBT) rapides en BiCMOS
- utilisation des propriétés du silicium contraint par l'épitaxie Si / SiGe (50nm)
- Technologie plus compliquée, donc plus chère et rendement plus faible

Technologie SiGe : principe



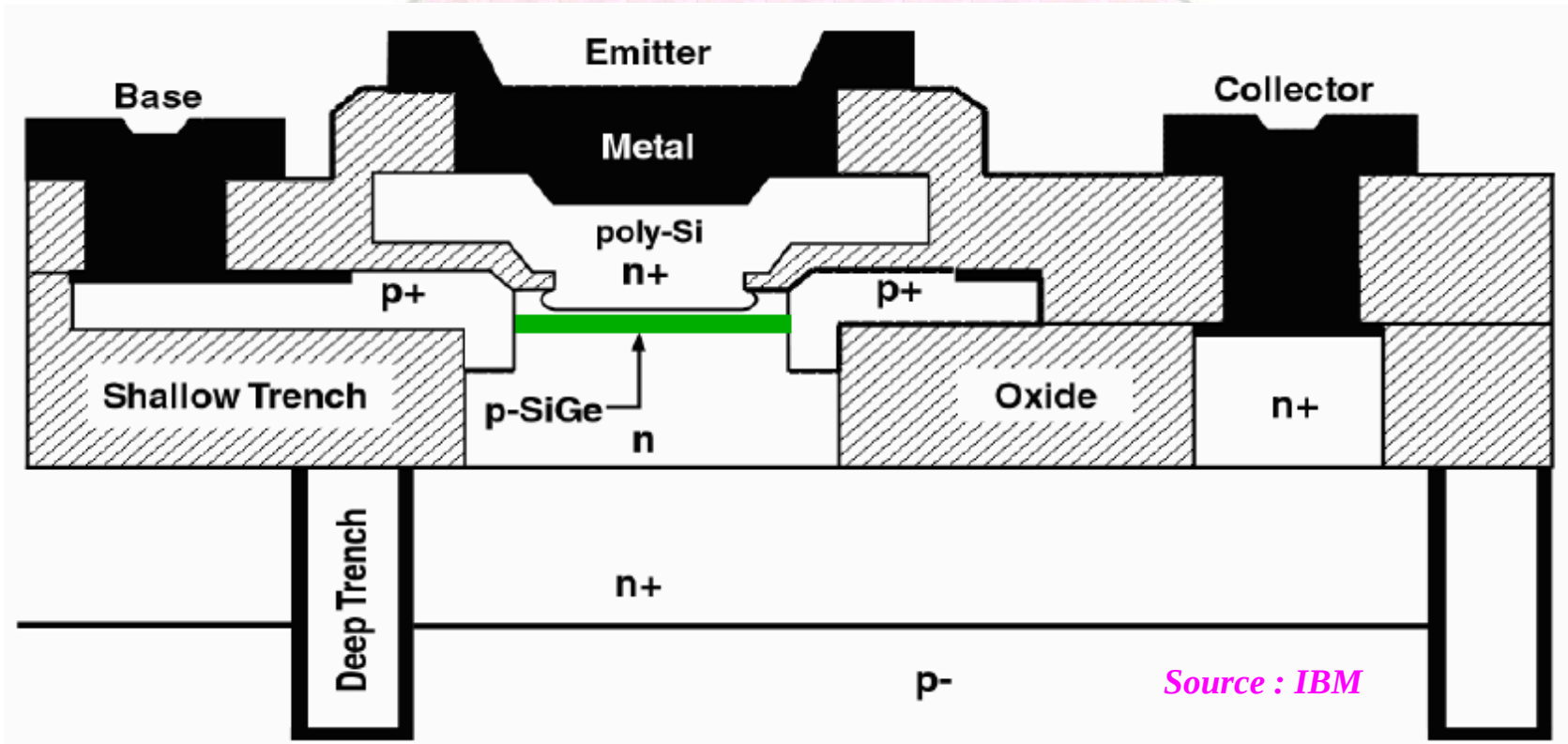






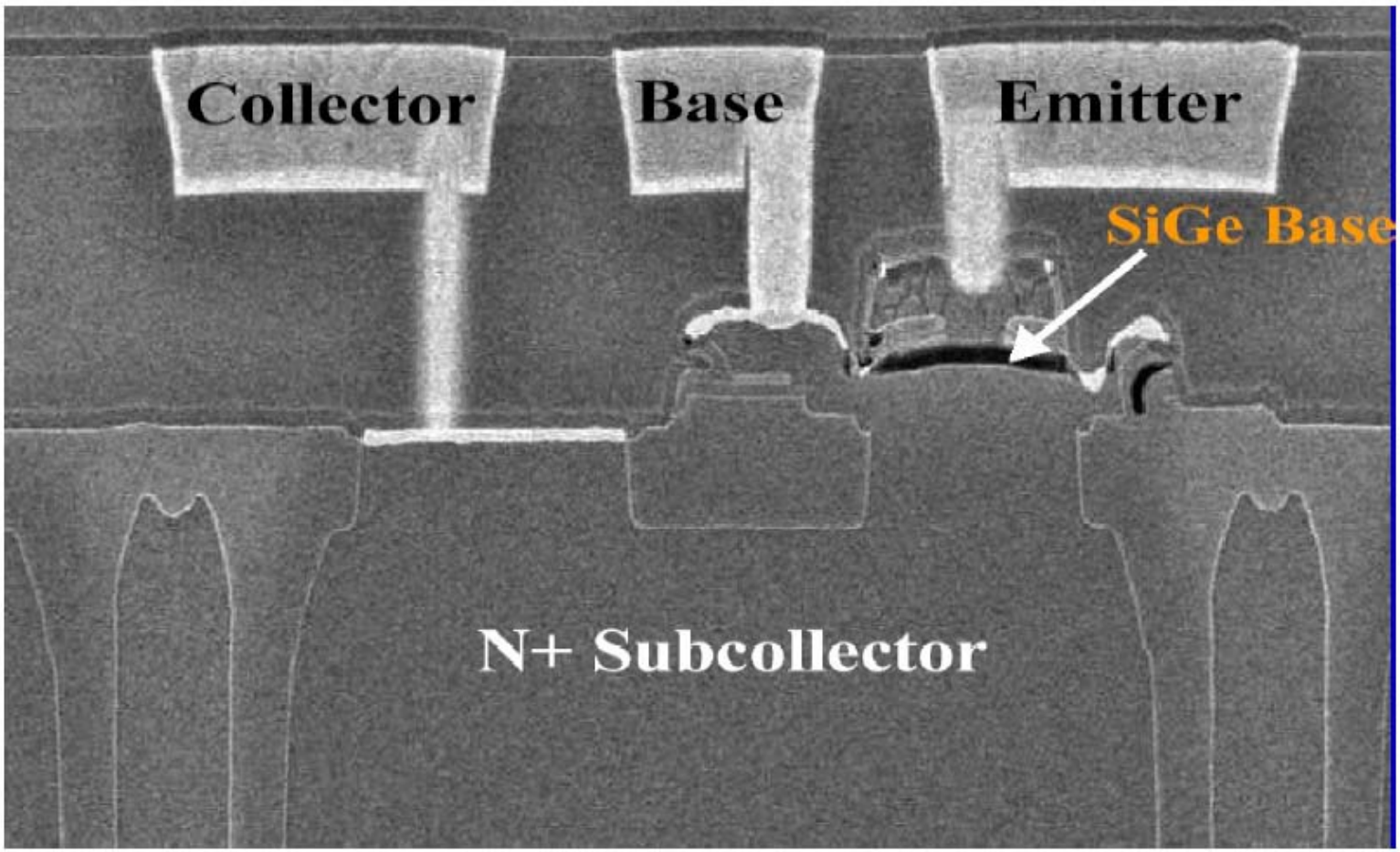
Obtention d'une couche mince de SiGe par épitaxie sur un substrat Si

Technologie SiGe : bipolaire hétérojonction

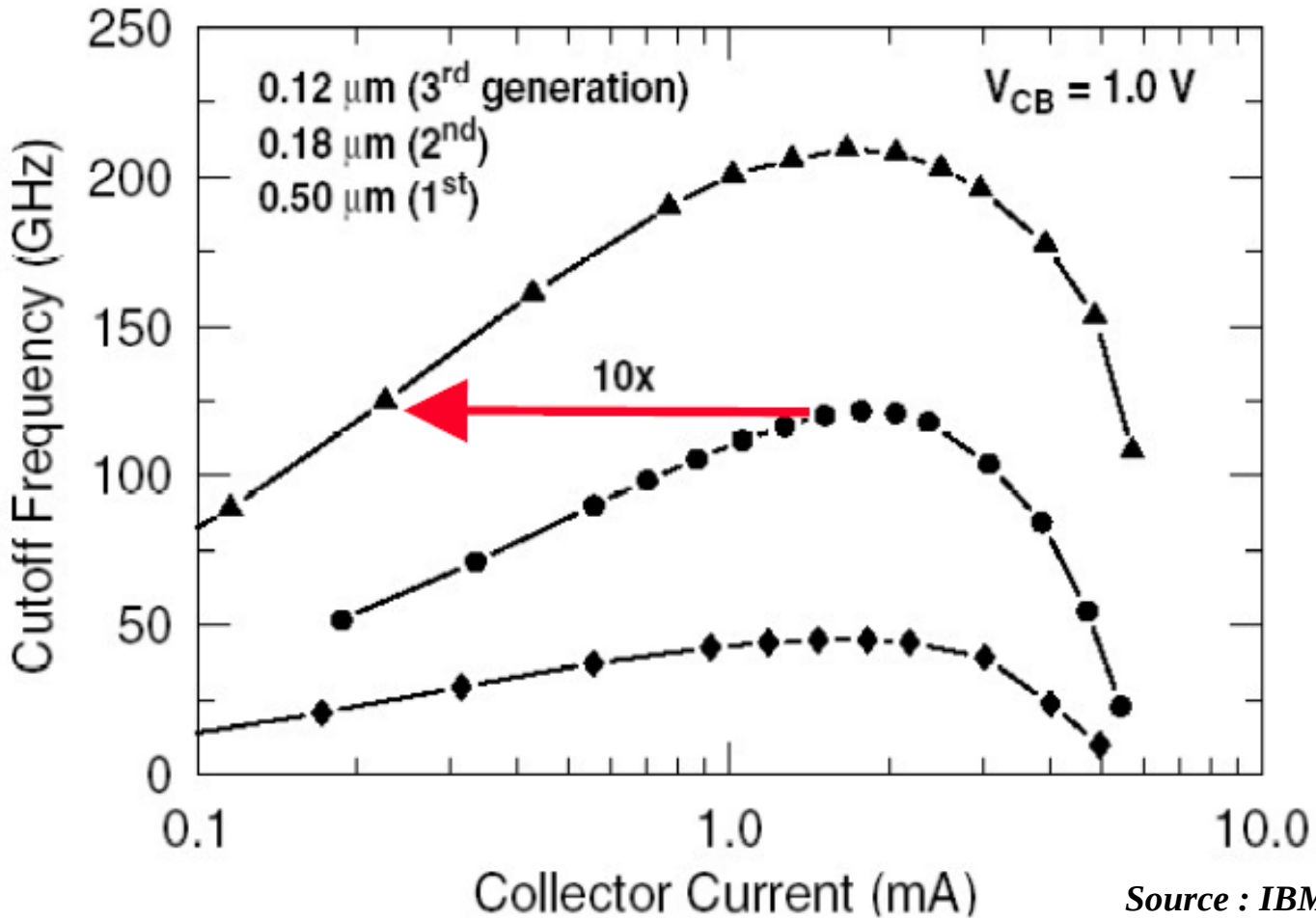


Transistor bipolaire hétérojonction : base SiGe fortement dopée épitaxiée sur un collecteur Si, émetteur en Si-poly ⇒ $f_{MAX} \approx 120GHz$
technologie BiCMOS

Bipolaire SiGe technologie IBM : F_T = 120GHz

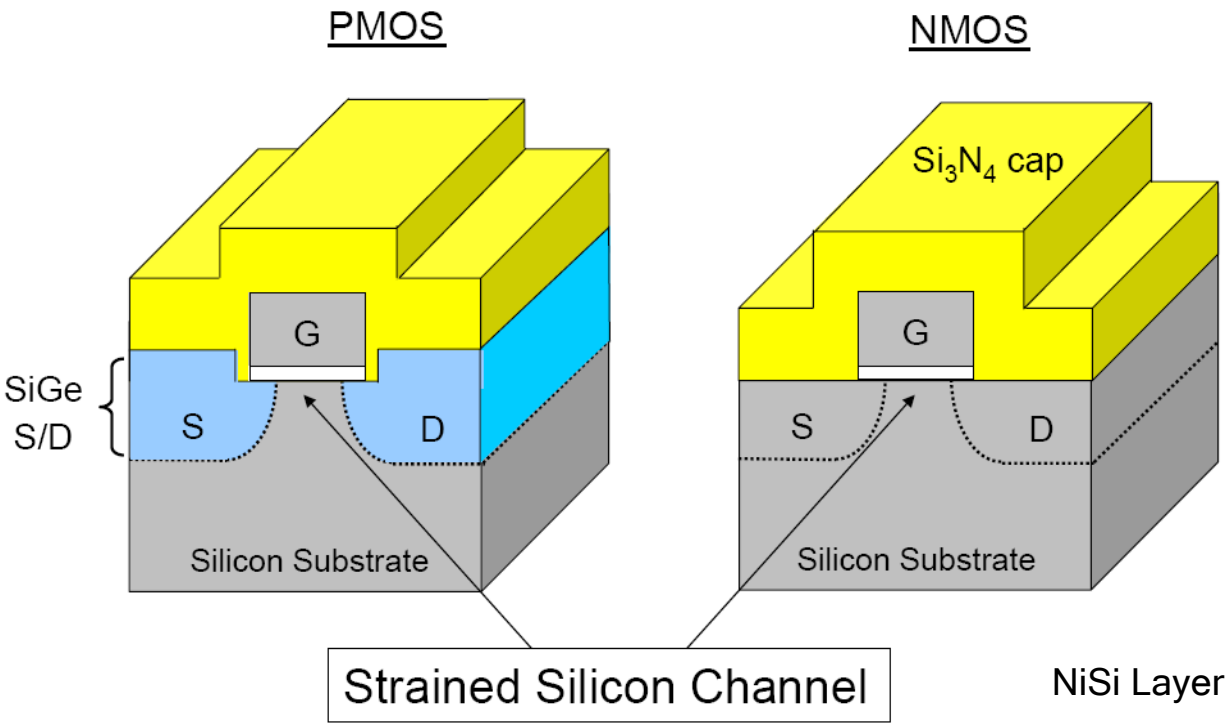


Technologies bipolaires SiGe : évolutions

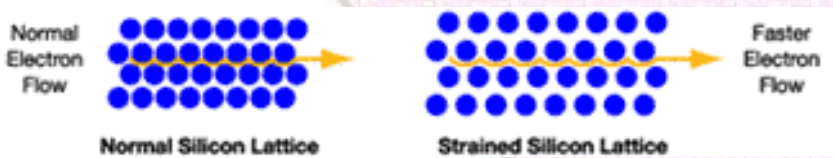


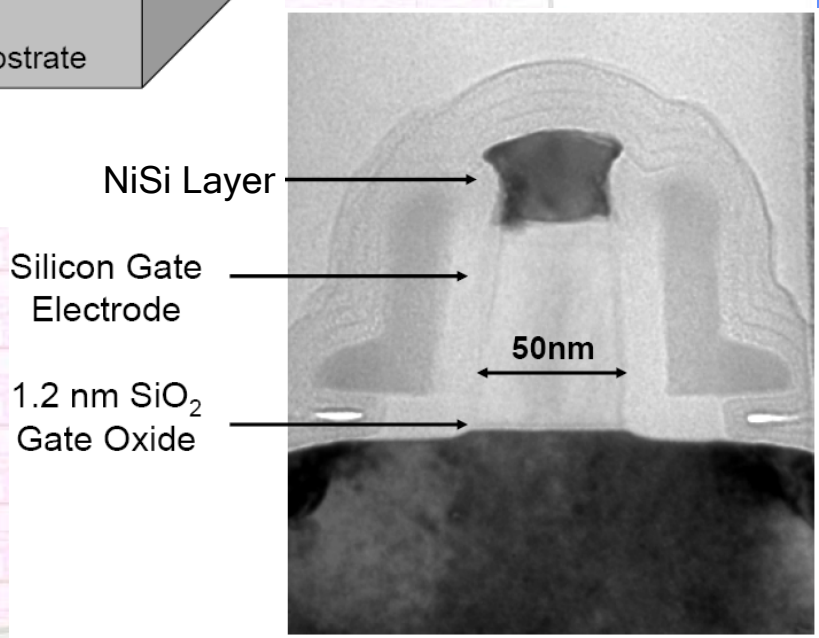
Source : IBM

Transistors MOS en SiGe/Si contraint

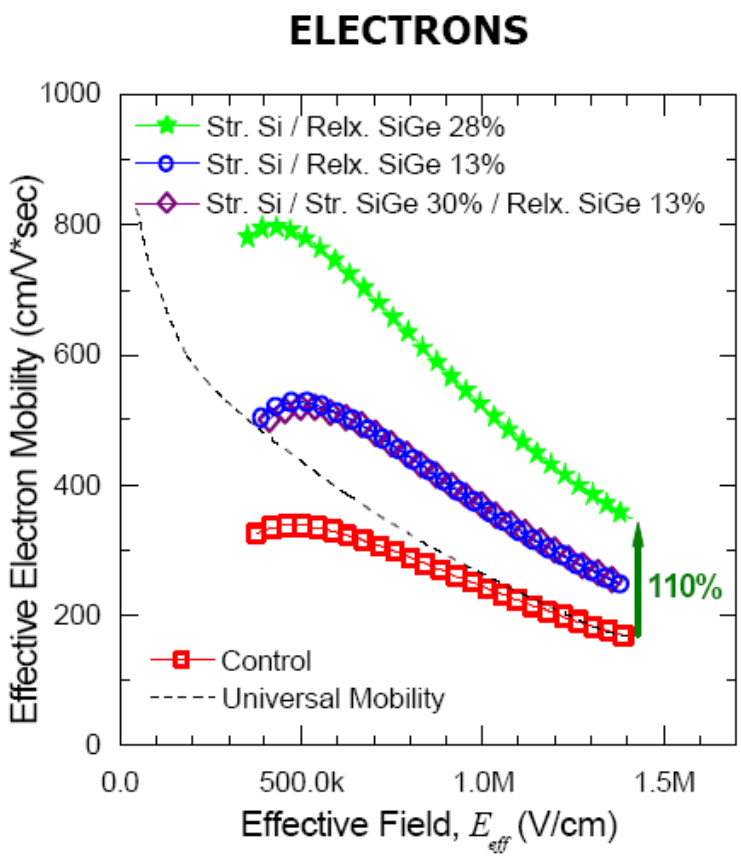


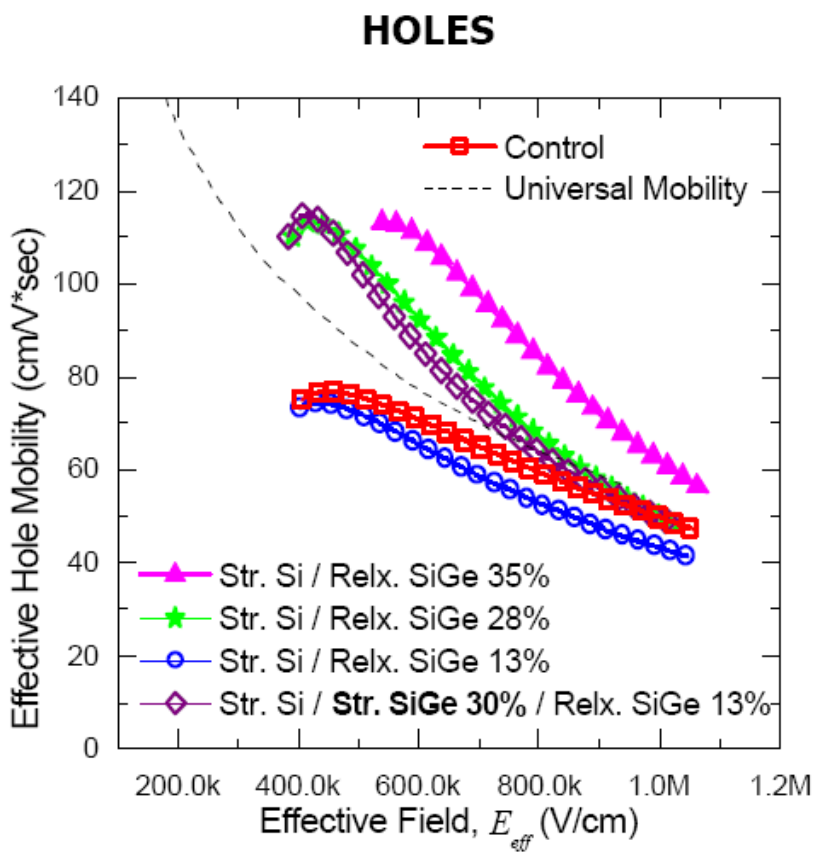
Source : Intel





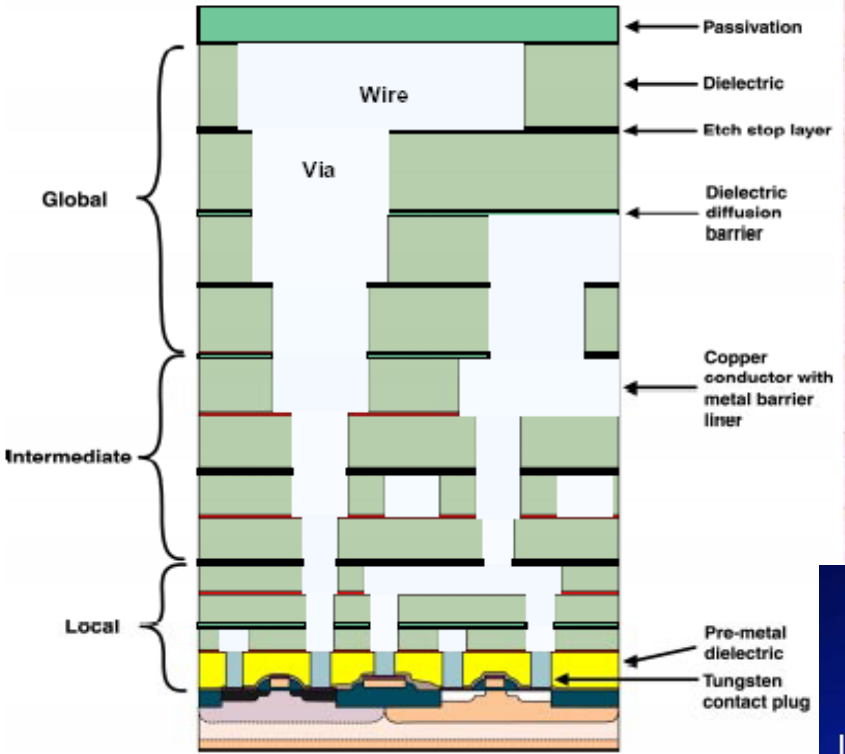
Si contraint sur SiGe : augmentation de la mobilité





K. Rim, S. Koester, M. Hargrove, et al.,. 2001 Symp. VLSI Technol., Digest p.12.

interconnections

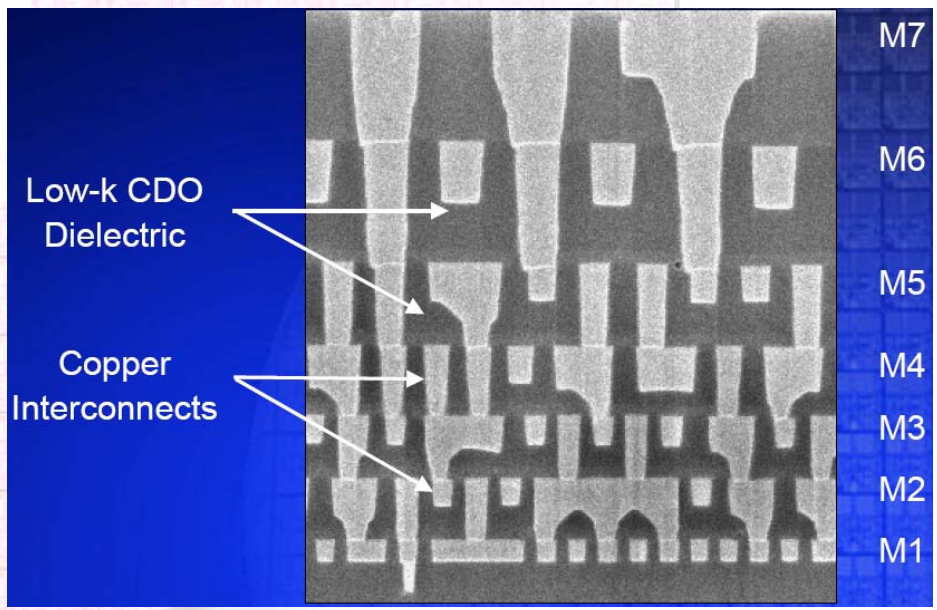


Année	2001	2005	2010	2016
Nombre de niveaux	8	10	10	11
Longueur totale des interconnections (m/cm^2)	4000	9000	16000	33000
Pas des métallisations intermédiaires (nm)	450	240	100	50

Source : SIA

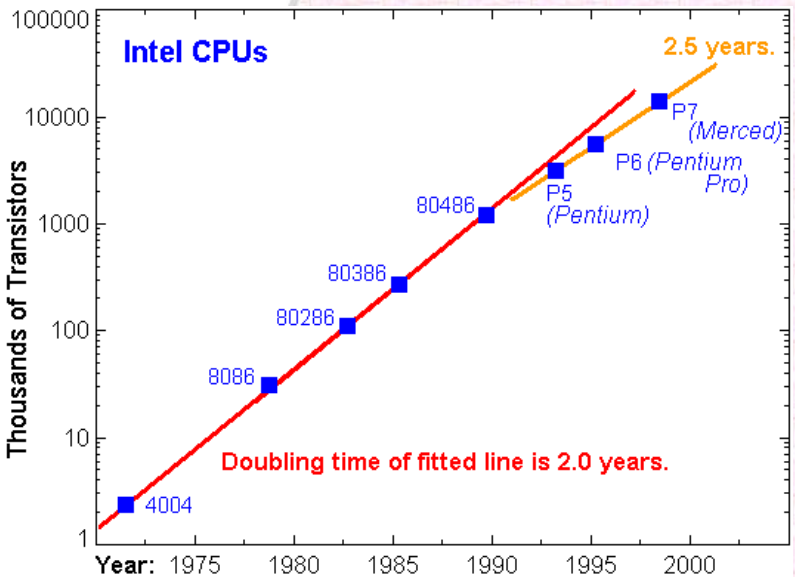
Solutions techniques à l'étude

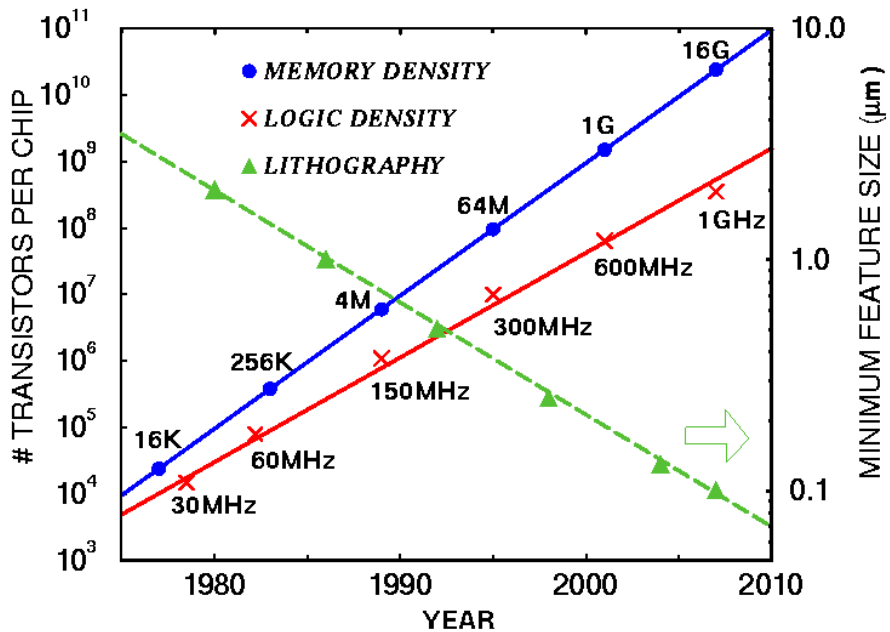
Pas de solution connue



- **Métaux utilisés :**
  - **W, Ti pour le niveau local (dépôts sélectifs)**
  - **Cu pour le niveau intermédiaire**
  - **Cu pour le niveau global avec répéteurs pour les liaisons très longues**
- **Futur :**
  - **liaisons RF, supraconducteurs, …**

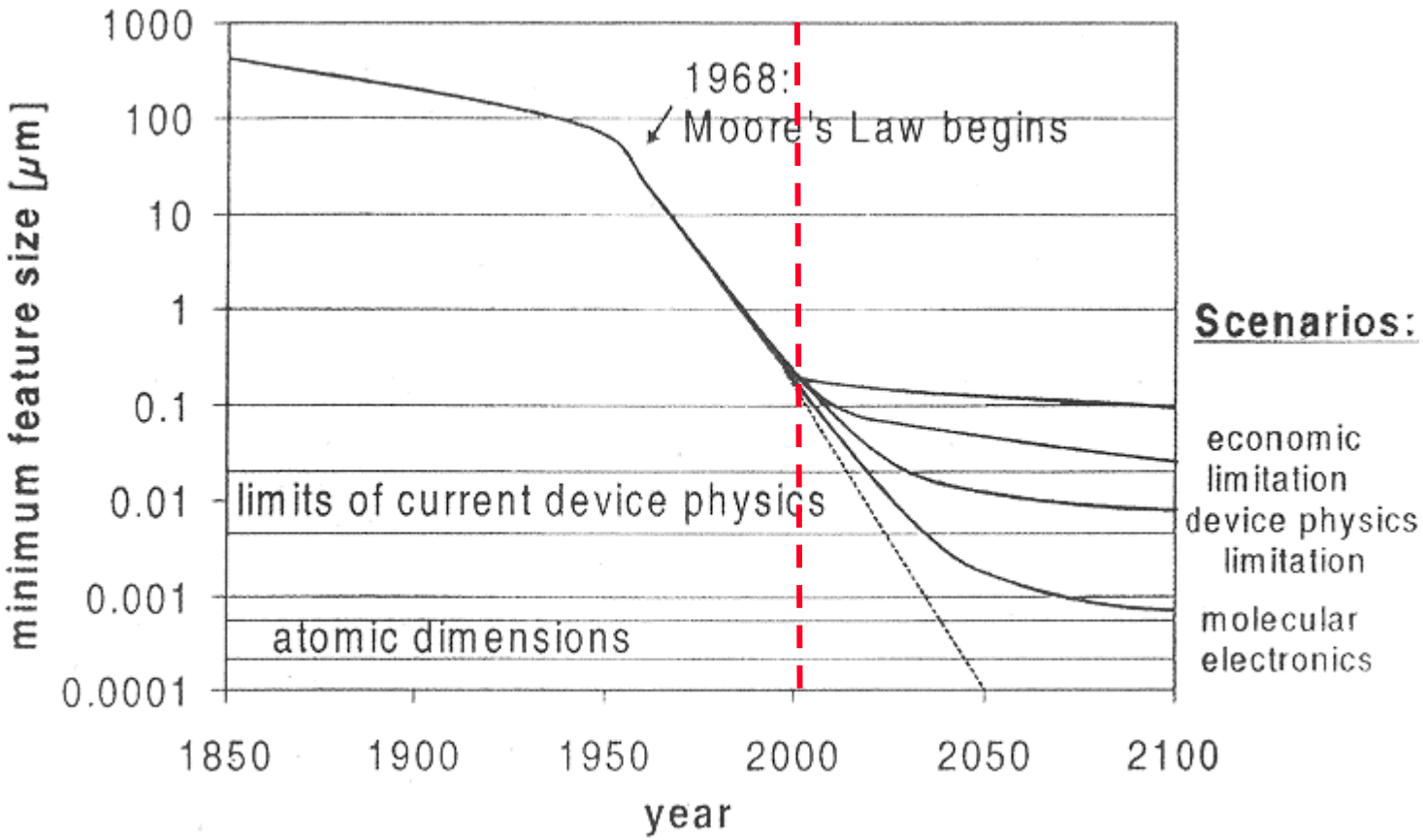
Loi de Moore : densité doublée tous les 2 ans





Encore pour combien de temps ?

Loi de Moore : perspectives



Evolution des technologies : projections

Year	2001	2003	2005	2007	2010	2013	2016
DRAM ½ pitch (nm)	130	100	80	65	45	32	22
DRAM generation	512M	1G	2G	4G	8G	32G	64G
MPU transistors/chip	97M	153M	243M	386M	773M	1.55G	3.09G
Local clock (GHz)	1.7	3.1	5.2	6.7	11.5	19.3	28.8
Number wiring levels	8	8	10	10	10	11	11
Total wire length (km/cm^2)	4.1	5.8	9.1	11.2	16.1	22.7	33.5
Interlayer eff. permittivity	3.3	3.3	2.8	2.5	2.1	1.9	1.8
High-perf. logic physical gate length (nm)	65	45	32	25	18	13	9
High-perf. logic EOT (nm)	1.3-1.6	1.1-1.6	0.8-1.3	0.6-1.1	0.5-0.8	0.4-0.6	0.4-0.5
High-perf. VDD (V)	1.2	1.0	0.9	0.7	0.6	0.5	0.4
High-perf. Power (W)	130	150	170	190	218	251	288
Low-power logic physical gate length (nm)	90	65	45	32	22	16	11
Low-power logic EOT (nm)	2.0-2.4	1.6-2.0	1.2-1.6	1.0-1.4	0.8-1.2	0.7-1.1	0.6-1.0
Low-power VDD (V)	1.2	1.1	1.0	0.9	0.8	0.7	0.6
Low-power Power (W)	2.4	2.8	3.2	3.5	3.0	3.0	3.0

Source : SIA ITRS

Conclusion

- En 2004, on approche les limites des technologies silicium conventionnelles
- Pour repousser ces limites, il faut imaginer des techniques entièrement nouvelles
  - composants quantiques
  - composants moléculaires
  - supraconducteurs
  - …
- On peut faire progresser les architectures des circuits et augmenter leur taille
- Pour l'analogique, if faut adapter les architectures aux très basses tensions d'alimentation

Perspectives en physique des hautes énergies

- **Pour les ASICs numériques :**
  - **Les logiciels de conception et de synthèse vont probablement être rapidement hors de nos moyens financiers.**
  - **La solution FPGA/DSP répond actuellement à nos besoins**
- **Pour les circuits analogiques :**
  - **Il faut adapter les architectures aux très basses tensions d'alimentation et trouver des technologies permettant de réaliser des résistances et des capacités (SiGe, AMS BiCMOS 0.35µ) : *voir R&D VLSI 0.35***
  - **Le marché de la physique des hautes énergies est bien trop réduit pour justifier à lui seul le maintien d'une technologie chez un fondeur.**
  - **Il faut trouver des partenaires qui ont besoin d'utiliser des technologies similaires. Les technologies SOI développées pour l'industrie automobile et les télécommunications peuvent nous intéresser**

Références

- **Physics of semiconductor devices** **S. M. SZE** ***Wiley***
- **IBM research journal** ***www.research.ibm.com***
- **INTEL technology journal** ***www.intel.com/labs***
- **International Technology Roadmap for Semiconductors (ITRS)** ***public.itrs.net***
- **University of Cambridge, Cavendish Laboratory, semiconductor physics group,** **D. J. Paul (SiGe)**