UNIVERSITÉ DE CAEN BASSE-NORMANDIE & ENSICAEN — ANNÉE 2010/2011

1ere session : 17 Février 2011

Master 2 Recherche "Electronique et automatique"
Epreuve "Procédés submicroniques, fluctuations dans les composants"
Terminal, durée 1h30

Tous types de document autorisés et calculatrice autorisée

Chaque candidat doit, en début d'épreuve, porter son nom dans le coin de la copie qu'il cachera par collage après avoir été pointé. Il devra, en outre, porter son numéro de place sur chacune des copies, intercalaires ou pièces annexées.

Ce sujet contient trois exercices indépendants. Les deux premiers exercices portent sur la première partie du cours. Le troisième exercice concerne le bruit et les fluctuations dans les composants

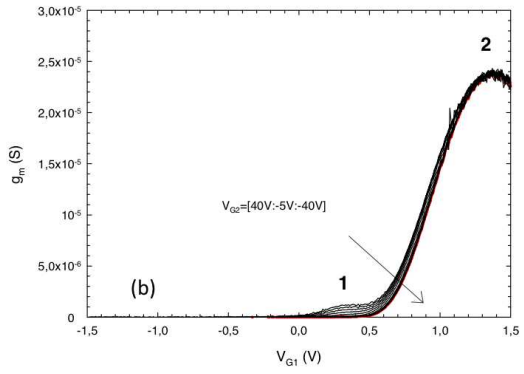
Exercice 1 : mobilité dans des composants nMOSFET GeOI

L'utilisation du germanium dans les technologie avancées a eu un regain d'intérêt avec le remplacement de la silice pour la grille par des diélectriques à forte contante diélectrique. En effet, ces diélectriques peuvent se déposer assez facilement sur du germanium : on pourrait alors profiter des perfomances plus intéressantes du germanium en particulier la mobilité des porteurs. A l'heure actuelle, des potentialités d'augmentation des performances ont été démontrés pour les nMOS mais les transistors pMOS se révèlent être décevants. Pour cette raison et du fait des coûts plus élevés engendrés par cette nouvelle technologie, le développement de ces technologies s'est arrété.

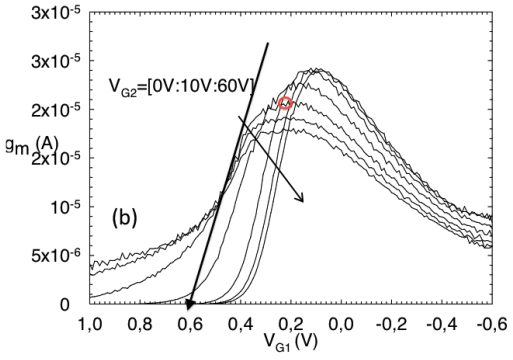
Dans cet exercice, on s'intéresse à la mobilité de transistors nMOSFET GeOI. On considère un transistor MOS sur substrat SOI, l'épaisseur de l'oxyde du substrat (BOX) est de 300 nm, l'épaisseur du film semiconducteur est de 40 nm. L'oxyde est constitué d'une couche de Si02 de 1 nm puis d'une couche de HfO2 de 4,5 nm. La largeur W des composants est de 10 μm, la longueur des composants L est de 1 μm.

Les figures ci-dessous (extraite de la thèse de J.E. Gyani "Caractérisation de dispositifis MOS SOI et GeOI avancées par l'étude et la modélisation du bruit basse fréquence", Université de Montpellier II) donnent la mesure de la transconductance en fonction de la tension V_{G1} appliquée sur la grille. Sur ces figures, on a $|V_{DS}| = 100\,mV$ et différentes valeurs de la tension appliquée sur la grille arrière (V_{G2}) sont reportées. On effectura l'analyse pour $V_{G2} = 0\ V$ (pour les pMOS, un rond et une flèche indique la caractéristique correspondant à $V_{G2} = 0\ V$). On rappelle que le maximum de transconductance est donné par l'expression : $g_{mMAX} = \frac{\mu_0 \cdot C_{OX} \cdot W \cdot V_{DS}}{L}$, que $\epsilon_0 = 8,85 \cdot 10^{-12} F/m$ et que pour la silice $\epsilon_{r_{OX}} = 3,9$.

nMOS :



pMOS :



Questions :

- 1. Calculer l'EOT ("Equivalent Oxide Thickness") de cette technologie.
- 2. Extraire des figures la valeur max de la transconductance et de la tension de seuil pour les transistors nMOS et les transistors pMOS.
- 3. En déduire la valeur de la mobilité des porteurs.
- 4. Comparer ces valeurs à celles des transistors nMOS et pMOS silicium.

Exercice 2 : analyse d'un article scientifique

1) La réduction des dimensions du transistor MOS entraine l'apparition de différentes sources de courant de fuite. Quelles sont les deux sources physiques identifiées par les auteurs de l'article ?

2) En figure 1 sont montrés des simulations des variations de courant I_{on} en fonction de la longueur de grille L_G (en imposant un courant I_{off} constant) pour différents nœuds technologiques. Comment sont expliquées ces variations anormales de I_{on} ?

3) Quels sont les principaux effets indésirables mentionnés dans l'article liés à la réduction d'échelle des composants?
Expliquer de façon succincte l'impact sur le fonctionnement d'un transistor MOS de l'effet d'abaissement de la barrière de potentiel par la polarisation du drain (« DIBL » en anglais)

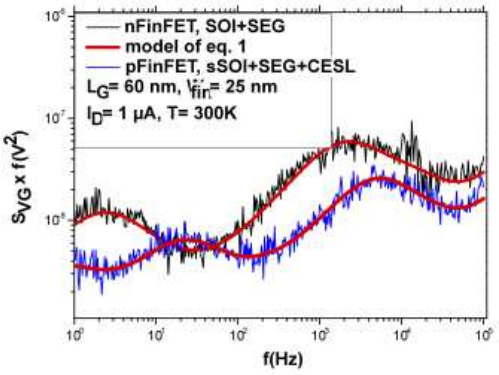
4) D'après les prévisions d'ITRS'2003 (illustré en figure 7 de l'article) quel doit être le niveau du courant de fuite I_{off} pour les dispositifs « low stand-by power » en technologie 32 nm ? Dans ce cas, quelles sont les solutions envisageables mentionnées dans l'article de manière à augmenter les performances des composants (d'augmenter le niveau de courant I_{on})?

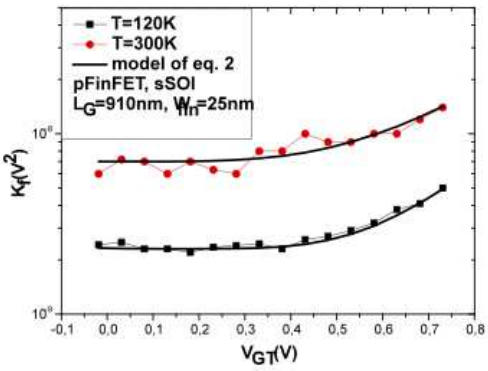
5) Quelle est la longueur physique de la grille d'un composant CMOS en technologie 22 nm ?

6) D'après les auteurs de l'article, quelle sera la plus petite longueur physique de grille qui pourra être réalisée technologiquement sans dépasser les limites critiques de point de vue de la physique de composants et de l'état de l'art actuel ?

Exercice 3 : Bruit basse fréquence dans des transistors MOS n et p FINFET

Les figures ci-dessous sont extraites d'un article proposé par R. Talmat et al. qui sera présenté à la prochaine conférence ICNF 2011 à Toronto ("International Conference on Noise and Fluctuations 2011"). La figure de gauche donne des densités spectrales de bruit de transistors n et p FINFETS. La densité spectrale de bruit a été reportée à l'entrée du composant et **multipliée par la fréquence** pour identifier plus facilement les différentes contributions apparaissant dans les spectres de bruit. La figure de droite donne la valeur du bruit en 1/f à 1 Hz extraite des spectres de bruit reportées à l'entrée, mesurées pour différentes températures et pour différentes valeurs de la tension $V_{GT}(= V_{GS} - V_T)$ appliquée sur la grille de transistors p FINFET .





Questions :

- 1. Rappeller les 3 types de bruit qui peuvent être distinguer sur un spectre de bruit.
- 2. Sur le spectre de bruit donné sur la figure, quels sont les types de bruit qui sont clairement identifiés ?
- 3. Rappeller les différentes sources de bruit basse-fréquence à prendre à compte dans un transistor MOS.
- 4. à l'aide de la figure de droite, indiquer les sources de bruit relevées par l'étude effectuée à 300 K.
- 5. Pour le bruit du canal, peut on conclure sur l'origine de volume ou de surface ?

ELSEVIER

Materials Science and Engineering B 124–125 (2005) 3–7

MATERIALS SCIENCE & ENGINEERING B

www.elsevier.com/locate/mseb

New materials and device architectures for the end-of-roadmap CMOS nodes

Thomas Skotnicki [a,*], Frederic Boeuf [a], Robin Cerutti [a], Stephane Monfray [a], Claire Fenouillet-Beranger [b], Markus Muller [c], Arnaud Pouydebasque [c]

[a] *STMicroelectronics, Crolles, France*
[b] *CEA-LETI, Grenoble, France*
[c] *Philips, Crolles, France*

Abstract

Conventional bulk CMOS scaling starts to fail. In order to prolong the life of Moore's laws, at least one technological booster (innovation) per node has to be introduced starting from the node 32 nm on. This presents a big technological challenge for the semiconductor industry. On the other hand, accumulation of the boosters permits to retrieve healthy scaling down to sub-10 nm gate lengths. This strategy even if technologically very challenging, is prospected to prolong the CMOS competitiveness till at least 2020.

Keywords: Semiconductor; CMOS competitiveness; Devices

1. Introduction

Operational MOS transistors shorter than 10 nm have been demonstrated at recent CMOS conferences. Yes, but their electrical characteristics remain by far behind the specifications (ITRS roadmap). Does it mean that scaling is dead? Does it mean that the future of the Moore's laws is menaced? We will analyse these questions in detail and identify the physical causes behind the prospective CMOS performance deficiency.

In the second part of the paper, we will show how non-classical device structures: ultra thin single- and double-gate devices such as SOI, FinFET, Silicon On Nothing (SON) and new materials (HK-dielectrics, metallic gates, strained-Si) can help with retrieving HEALTHY scaling. We will also deliberate on how they can help to overcome or at least attenuate the problems with short-channel effect, with drain-induced barrier lowering effects, with high-field effects, with mobility degradation, etc. Finally, we will show how these new device structures and materials can prolong the Moore's laws up to the end of the roadmap, and even beyond, into the nano-world.

2. Where does the trouble come from?

The problem with CMOS scaling generally originates from different sort of leakages that arise when shrinking the transistor dimensions. Of course this is almost a phenomenological answer, and as researchers we would like to understand why it is so? The first physical source resides in the non-scalability of certain transistor physical parameters, such as the subthreshold slope, build-in voltage, carrier injection velocity, dark-space in the inversion layer, effective field, etc. To illustrate the above statement, let us take example of the subthreshold slope. Its minimum value in majority of electronic devices is 60 mV/dec (at room temperature). Therefore, the reduction of the threshold voltage, resulting from scaling, inevitably leads to an increase in the source–drain leakage current I_{off}.

The second source of troubles lies in the fact that scaling down certain parameters we awake some new physical phenomena that before were negligible but now come to play. A good example here may be the gate tunnelling current. It practically did not exist when gate oxide was thicker than 5 nm, but becomes a dominant source of leakage when the oxide gets close to 1 nm.

All these new phenomena cause the scaling to break down. A good example here can be with gate length scaling. As well known from any text book on transistor physics, the Source–drain current (when the transistor is on) should increase as $1/L$ when shortening the gate. The entire scaling theory [1] is

* Corresponding author.
E-mail address: thomas.skotnicki@st.com (T. Skotnicki).

0921-5107/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.mseb.2005.08.122

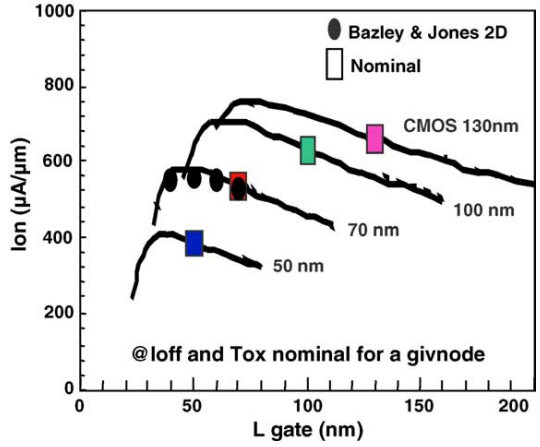


Fig. 1. Anomalous scaling of transistor gate length at constant I_{off}.

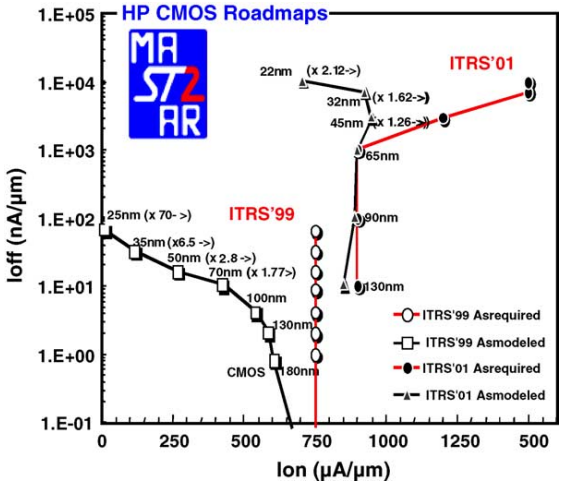


Fig. 2. Impact of scaling deficiency on the CMOS roadmap.

built on this kind of physics. In now-a-day technologies, however, this physics fails, see Fig. 1 (calculations made with the use of the MASTAR model, [2]). The explanation is very simple [3]: as the gate shortening leads to a drop in the threshold voltage value V_{th} (due to SCE and DIBL) and thus to an increase in I_{off}, we need to accompany the gate shortening with stronger channel doping to keep the I_{off} constant. Doping increase leads, however, to reduced carrier mobility that diminishes the current more efficiently than the channel length reduction increases it.

The impact of this scaling deficiency on the CMOS roadmap is dramatic. As shown in Fig. 2, the actual roadmap 1999[1] (open squares) diverges rapidly from the targets (open circles). In the 2001 ITRS roadmap edition, we have improved the situation by better (optimal) choice of the nominal gate lengths pour each node. Actually the gate nominal lengths had been shortened so to correspond to the maxima in the curves shown in Fig. 1. This simple operation is very efficient and renders the 65 nm node feasible with Bulk. Nevertheless, the divergence between reality and the targets persists—it is just diminished and belayed. Therefore, in order to really get in line with the Moore's laws, we need to add new more efficient technologies, structures and materials to the game.

3. How non-classical technologies, structures and materials can help?

As we have already stated DIBL, SCE and subthreshold slope S are among the main causes of the trouble with scaling. All of them can be improved by shallowing the junction. Junction design used to be governed by one of the so called "good technology design rules" requiring the junction depth to be 1/3 of the gate length. The problem is that this, as well as all the other good design rules, are violated already today. To understand the consequences of such the violation we need some physical insight into the "good technology design rules". They appeared as an empirical observation guarantying that if the ratios:

$$\frac{T_{ox}}{L_g} = \frac{1}{30},$$
$$\frac{X_j}{L_g} = \frac{1}{3} \text{ and}$$
$$\frac{T_{dep}}{L_g} = \frac{1}{3}$$

are conserved when designing a new technology, its subthreshold behaviour would be good. Only recently a model has appeared [2] that derives these rules from the transistor physics. It has turned out that SCE and DIBL are explicit functions of the ratios:

$$\text{SCE} = 0.64\frac{\varepsilon_{Si}}{\varepsilon_{ox}}\left(1+\frac{X_j^2}{L_{el}^2}\right)\frac{T_{ox\_el}}{L_{el}}\frac{T_{dep}}{L_{el}}\phi_d$$

$$\text{DIBL} = 0.80\frac{\varepsilon_{Si}}{\varepsilon_{ox}}\left(1+\frac{X_j^2}{L_{el}^2}\right)\frac{T_{ox\_el}}{L_{el}}\frac{T_{dep}}{L_{el}}V_{DS}$$

It is easy to calculate with these expressions that if the good technology design rules were respected and if no difference were supposed between L_g and L_{el} as well as between T_{ox} and T_{ox_el}, DIBL would be kept constant equal to 30 m (V/V) across all CMOS nodes.

In the past, the differences between L_g and L_{el}, and between T_{ox} and T_{ox_el} could be neglected as small in comparison with their bases. Today they are of the same order of magnitude and have to be taken into account. Therefore, if we admit that:

$L_{el} = L_g - \gamma X_j$ (due to subdiffusion γ was around 1.2 in old technologies, and is close to 0.8 in modern technologies)

[1] The I_{off}–I_{on} curves labelled as the roadmaps (open and filled circles) are calculated in the following way: I_{on} is adjusted in order to suite the traditional 17%/year improvement rate (Moore's law) in intrinsic frequency ($1/(CV/I)$), then V_{th} is adjusted to suite these I_{on} values, and finally the I_{off} results from the required V_{th}.

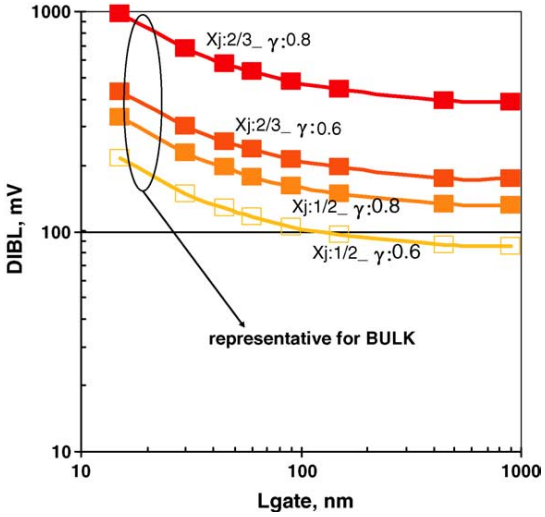


Fig. 3. DIBL as function of L_{gate} for X_j and γ representative for bulk—calculated with MASTAR.

and

$T_{ox\_el} = T_{ox} + 8A$ (a halve of the $8A$ term is due to poly depletion and a halve due to dark-space),

DIBL will grow rapidly when shortening the gate length, even if the good technology design rules are respected. But in fact they are not.

Is it, thus possible, to survive with BULK? If we admit that X_j used to be closer to $2/3L_g$ rather than 1/3, the DIBL values sky rocket, see Fig. 3. What happens then that the real measured DIBL values even if exceeding the desired limit of 100 mV are not as large as these predictions? The answer comes again from the analysis of the model. By limiting diffusion thanks to advanced annealing techniques we can better limit the junction depth. In today technologies, X_j can be closer to $1/2L_g$ rather than to $2/3L_g$. In addition, by countering the extension doping with the use of hallo/pocket doping, we can strongly limit the subdiffusion. Therefore, the γ factor can possibly be reduced from 0.8 to slightly lower values. Also the use offset spacers helps to artificially reduce the γ value, although for other reasons they are less and less frequently used in advanced technologies.

As shown in Fig. 3, if we succeed to cumulate both $X_j = 1/2L_g$ and $\gamma = 0.6$, we can keep DIBL below 200 mV down to 15 nm gate length even with Bulk. For certain applications such value may however be too large. Are there means to reduce DIBL further?

In the past, the largely used leverage was to over-dope the channel, thus, reducing the T_{dep}/L_g ratio below 1/3, and therefore, to compensate for too large values of X_j/L_g and T_{ox}/L_g. The problem is that this method can no longer be applied since the channel doping has already been brought to such high values ($>10^{18}$ cm^{-3}) that junction breakdown become a strong concern. In addition, this method cannot be applied to ultra thin body (UTB) structures that are supposed to work with undoped fully depleted channel.

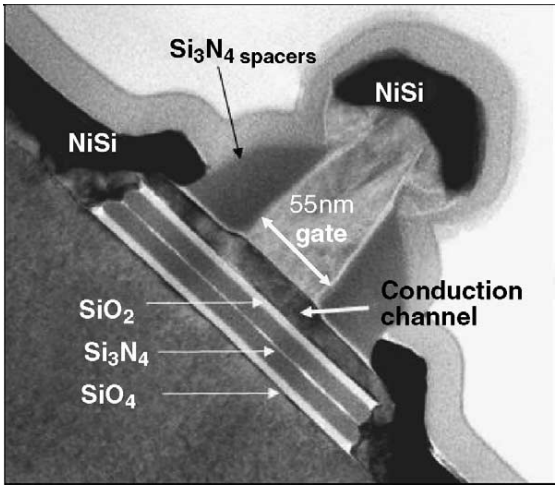


Fig. 4. Silicon-On-Nothing (SON) transistor structure, [4].

Nevertheless, with the use of UTB device structures it becomes pretty straightforward to bring X_j to 1/3 of L_g. As X_j does no longer result from diffusion but rather from the silicon film thickness T_{Si}, we can realise the $X_j = 1/3L_g$ relation with these kind of structures. In the example shown in Fig. 4, the SON transistor structure has a conventional deep SD junction (to facilitate silicidation) and very shallow extensions determined by the silicon film thickness (to suppress the SCE and DIBL). In contrast, it is rather difficult to imagine γ being further improved beyond 0.6 because in SON/SOI structures, tendency is not to dope the channel. Therefore, we can no longer counter the lateral expansion of the junction (subdiffusion) by pocket/hallo implantations. Nevertheless, even with $X_j = 1/3L_g$ and $\gamma = 0.6$ and metallic gate (MG) we can conduct the SON/SOI structures to 15 nm gate length with DIBL remaining below 100 mV, see Fig. 5.

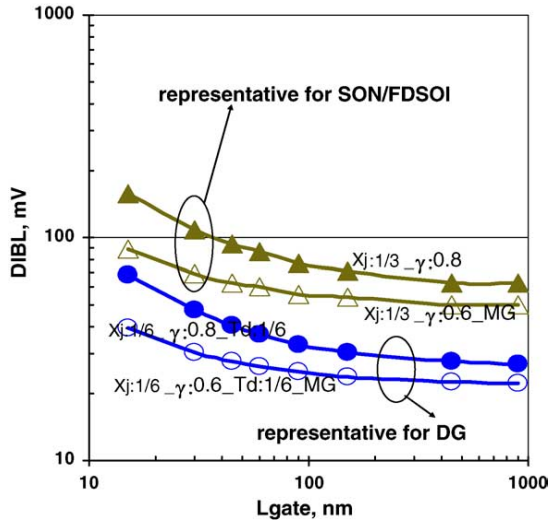


Fig. 5. DIBL as function of L_{gate} for X_j and γ representative for SON/SOI and for DG—calculated with MASTAR.

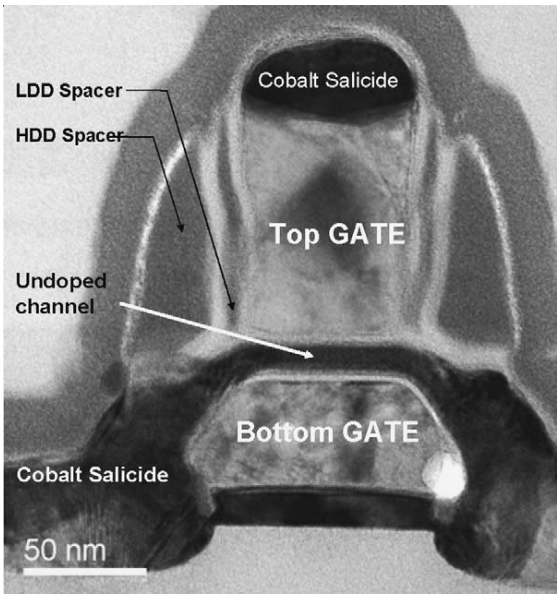


Fig. 6. Planar DG transistor structure obtained with the DG SON process.

Still further improvement is possible with the use of DG structures. Their strength with DIBL suppression results from the vertical symmetry of the potential distribution in such structures. This causes that only half of the film thickness contributes to the junction depth as well as to the depletion depth. In other terms, DG structure (Fig. 6) can be seen as a superposition of two planar transistors having an axis of symmetry in the middle of the Si film. Therefore, if the film thickness follows the same scaling rule as junction in the case of Bulk (i.e. $T_{Si} = 1/3L_g$), the junction depth and the depletion depth will both scale as $1/6L_g$. As shown in Fig. 5, such a scaling (only possible with DG structures) allows DIBL to be kept smaller than 40 mV down to 15 nm gate lengths.

4. Ushering CMOS into nano-world

It is interesting to see how far all these innovative structures, technologies and materials can bring CMOS. The answer is of course different depending on the product category. This is because of the I_{off} difference between these products, that implies that the advantage of a better subthreshold slope (advantage of all UTB devices) shows up or not. Suppose LSTP products where I_{off} is very small, let say 10 pA/μm—we need the V_{th} adjusted at the level of 300 mV, if the subthreshold slope is 60 mV/dec, but as much as 450 mV if the subthreshold slope is 90 mV/dec. The threshold is supposed to correspond to 1 μA/μm in this simple example meaning five decades difference between the I_{off} and the threshold level. To assess the impact of such a difference in the threshold voltages, assume that I_{on} is linear with the $(V_{dd}–V_{th})$. Then 150 mV difference in V_{th} corresponds to 27% difference in I_{on} current at 1 V supply. In HP products, however, the I_{off} is very large and close to the V_{th} level. Therefore, only scant advantage in I_{on} can be produced due to better subthreshold slope.

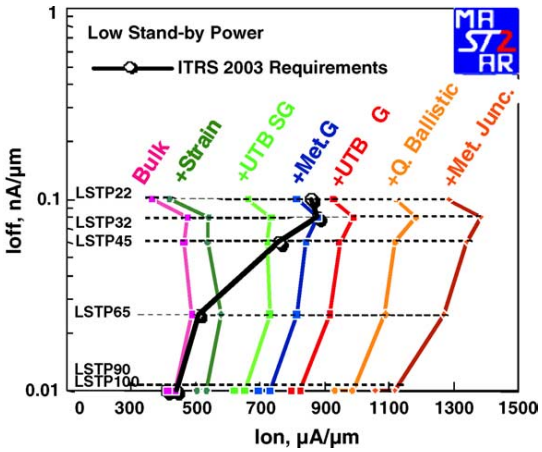


Fig. 7. CMOS 2003 roadmap for LSTP products. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

In Fig. 7, we show the CMOS roadmap for LSTP products (black line) and a sequence of roughly parallel colour lines representing performance of CMOS technologies equiped with more and more boosters. The pink line corresponds to conventional Bulk and is convergent with the roadmap only till the 65 nm node. To catch the 45 nm node we need to add strained channel, replace Bulk with SON/SOI and switch to metallic gate. The good news is that this collection of boosters keeps valid till the end of the roadmap.

As shown in Fig. 8, this is different for the HP products. To catch the last node 22 nm, we need to develop DG device structures with metallic gate and strained channel and produce a semi-ballistic transport in the channel. The good news is that in both cases there remains a booster (metallic junction) that gives a margin to extend the CMOS roadmap beyond the 22 nm node.

Taking into account that the physical gate length for HP 22 nm node reads at 9 nm, we can estimate that CMOS has real potential to devices in the range of 5 nm.

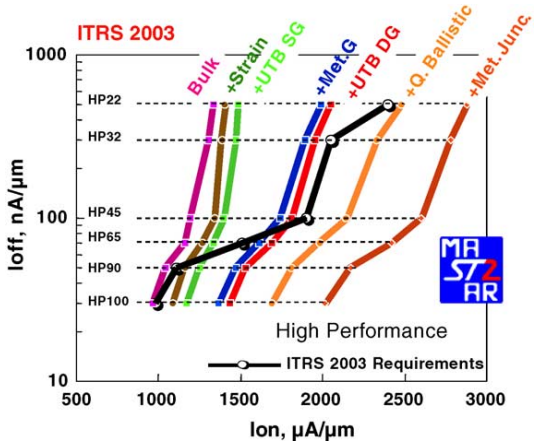


Fig. 8. CMOS 2003 roadmap for HP products.

5. Conclusions

An extensive innovation reach R&D is in progress to remain on the path of Moore's laws. According to physical analysis and simulations, CMOS has potential to remain on this path up to the 16 nm node. Economical and market related issues are of course the major consideration, so are also integration issues such as dispersions, power dissipation, etc. Nevertheless, it is good to know that at least regarding the physics, no critical show-stopper is seen to impede operational transistors in the scale of 5 nm.

References

[1] R.H. Denard, et al., IEEE J. SSC (September) (1974) 256–268.
[2] MASTAR Code and Instructions—available on request to the authors.
[3] T. Skotnicki, F. Boeuf, in: H.R. Huff, L. Fabry, S. Kishino (Eds.), Proceedings of the Ninth International Symposium on Silicon Materials Science and Technology, ECS Volume 2, 2002, pp. 720–734.
[4] S. Monfray, D. Chanemougame, S. Borel, A. Talbot, F. Leverd, N. Planes, D. Delille, D. Dutartre, R. Palla, Y. Morand, S. Descombes, M.-P. Samson, N. Vulliet, T. Sparks, A. Vandooren, T. Skotnicki, IEDM Tech. Dig. (December) (2004) 635–638.