李志坚文集

（下）

科学出版社
www.sciencep.com

李志坚文集

（下）

科学出版社
北京

图书在版编目(CIP)数据

李志坚文集.下/李志坚著.—北京:科学出版社,2010
ISBN 978-7-03-027090-0

Ⅰ.李… Ⅱ.李… Ⅲ.电子技术-文集 Ⅳ.TN-53

中国版本图书馆 CIP 数据核字(2010)第 051105 号

责任编辑:刘红梅 杨 凯/责任制作:董立颖 魏 谨
责任印制:赵德静

北京东方科龙图文有限公司 制作
http://www.okbook.com.cn

科学出版社 出版
北京东黄城根北街 16 号
邮政编码:100717
http://www.sciencep.com

北京天时彩色印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2010 年 5 月第 一 版 开本:B5(720×1000)
2010 年 5 月第一次印刷 印张:110 3/4 插页:8
印数:1—1 100 字数:1 736 000

定 价:360.00 元

(如有印装质量问题,我社负责调换)

目 录

战略展望

器件物理

集成电路技术

系统集成技术

战略展望

A Discussion on the Extension of Moore's Law[1)]

Abstract Both theory and ITRS predict that the scaling of conventional silicon CMOS is going to approach its ultimate. Several emerging devices based on different new mechanism are proposed as possible successors. But still, as classical 2-level switches, they have common limits of quantum uncertainty and thermodynamic fluctuation. In order to further enhance and improve the ability and performance of ULSI systems, either more efficient devices or processing systems with new model and architecture are required. In this article, these topics are briefly discussed in unified view of the product of device delay time and its power dissipation.

Ⅰ. Introduction

Almost three decades ago Gordon Moore of Intel Corporation observed that the number of transistors per chip had been doubling annually for years[1]. This astute observation has become known as "Moore's Law". Today the Moore's Law is still valid. But this trend certainly has its ultimate. When decreasing the channel length and supply voltage, some secondary effects will become more and more significant, and the increasing leakage current I_{leak} will increase the stand-by power dissipation. When this increasing stand-by power dissipation exceeds some percents (10% for example) of maximum power dissipation and the transconductance is no longer larger than the output conductance, our scaling down strategy meets its ultimate.

This article will have a review and discussion on the extension of Moore's Law from the energy point of view of switching transition. In section Ⅱ a discussion of ultimate CMOS, base on theory. ITRS and experiments, is given. In section Ⅲ data

---

1) In: International Conference on Solid-State and Integrated Circuits Technology Proceedings, ICSICT, v1, 2004 7th International Conference on Solid-State and Integrated Circuits Technology Proceedings, ICSICT 2004, 2004: 251—254 (Coauthor: Tiefu Li).

of some recently developed emerging devices are involved. In section Ⅳ the limit of classical 2-level switching devices will be shown. Finally in section Ⅴ, a discussion on how to overcome the Heisenberg limit is given.

Ⅱ. Ultimate Si CMOS

Si CMOS is the foundation of today's IC technology. The enhancement of IC performance is mainly by appropriately scaling down of device dimension and voltage. But this scaling is not boundless[2]. When scaling down the channel length and voltage, some secondary effects, such as SCE, DIBL, Punch-Through and Hot-Carrier Effects, will tend to be more and more significant, which make the leakage current I_{leak} and stand-by power dissipation increasing. When the stand-by power dissipation in total power dissipation reaches a definite proportion (different for different applications) and device output conductance gets over the transconductance. MOSFET's scaling down meets its ultimate[3]. For example, ITRS has set that the maximum power dissipation (P_{max}) of unit chip area is 100W, and for HP(High Performance) CMOS devices the stand-by power dissipation. P_{st}, should be less than 10% of P_{max}. With this stand-by power dissipation and supply voltage (V_{dd}), we can get the maximum leakage current (I_{leak}). And then the effective channel length. L_{eff}, thickness of oxide layer, t_{ox}, and finally the ultimate E_b and performance can be decided. For Si CMOS there are also limits on signal delay, which is mainly because of the finite velocity of electrons. With the saturated velocity of electrons (v_{sat}) we can deduce the minimum delay, $\tau_{min} = L/v_{sat}$, and the maximum delay with the minimum supply voltage, $\tau_{max} = L^2/\mu V$. Thus for every possible E_b, CMOS can be designed on a segment of $\log\tau$-$\log P$ plot with τ_{min} and τ_{max} as terminals.

For every 2-level switch we have $E_b = P\tau$ and plot $\log\tau$ vs $\log P$ gives a straight line with constant E_b. For CMOS $E_b = \frac{1}{2}CV^2$ where C and V are capacitance and applied voltage of the switch. Fig. 1 shows the scaling down trend of Si CMOS on this plot, data is from ITRS(2003). The points represent the delay, and power dissipation of Si CMOS devices with different channel length and corresponding V_{dd} and V_T. Two lines at the left lower corner represent the limits set by Heisenberg Uncertainty Principle and Thermodynamics[4]. First of all, from this plot we can see that with devices scaling down, points move towards the lower left corner of the plot, that is the smaller delay and power dissipation, better performance. Second, as mentioned before, the scaling down strategy has its ultimate. The theoretical predicted value for ultimate CMOS E_b is approximately the same as that proposed by ITRS[5]. For HP devices, using given P_{max} and P_{st}, it has been found that ultimate

$L_g \approx 10$nm, corresponding ultimate $E_b \approx 10^{-18}$ J for 300K. So the corresponding constant E_b line in the log$P$-logτ plot is the ultimate E_b of Si CMOS for HP use. Third, for one E_b, the devices also have limit on signal delay time as mentioned before. The delay time can only be between the τ_{min} and τ_{max}. As approximately shown in the plot, the area surrounded by three lines is the zone for Si CMOS. All Si CMOS devices can not be out of this region.

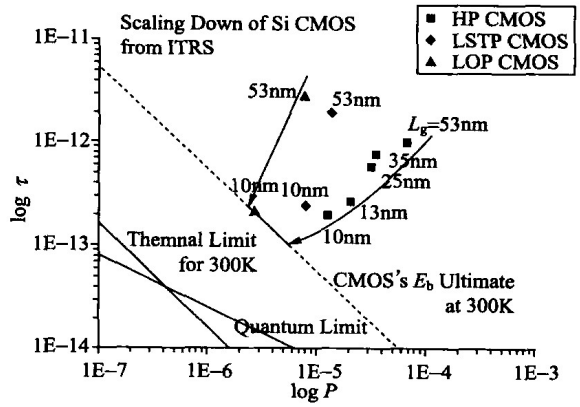

Fig. 1 Scaling down trend of Si CMOS from ITRS

Recently many research centers and companies, such as IME of CAS[6]. Stanford, IBM and Intel, have developed MOSFET with nano-scale channel length(from 6nm to 70nm). Fig. 2 shows their properties (data from [6] and IEDM(2001～2003)) comparing with the trend from ITRS. We can see from this plot clearly that these experimental points are all in the discussed region of Si CMOS, especially they are all above the CMOS's E_b limit. They can have different power dissipation or delay, but no one of them is located beyond the limit of Si CMOS devices.

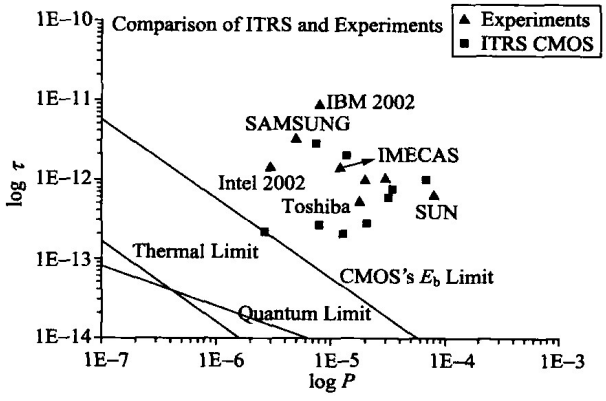

Fig. 2 Scaling down trend from experiments and ITRS

Ⅲ. Emerging Devices

In order to overcome the Si CMOS limit recently a series devices based on new mechanism are proposed. Some perspective examples are discussed in this section.

1. CNT

CNT(Carbon Nano Tube), as a kind of natural nanowires, has been shown to have larger mobility than that of bulk silicon and due to one dimensional transportation the ballistic transport is expected in it. Further, since CNT has a very large breakdown electric field (up to 10^6 V/cm), the carrier drift velocity can greatly exceed that in silicon inversion layer.

We can see from Fig. 3 and Fig. 4(data extracted from ITRS(2003)) that by scaling Si CMOS, the electron mobility decreases, while the electron drift velocity nearly keeps constant. So we can not reduce delay as we want. This is velocity saturation. But it is not this case for CNT. By ballistic transport in CNT, nearly constant mobility with increasing of electric field and thus velocity over-shoot can be excepted.

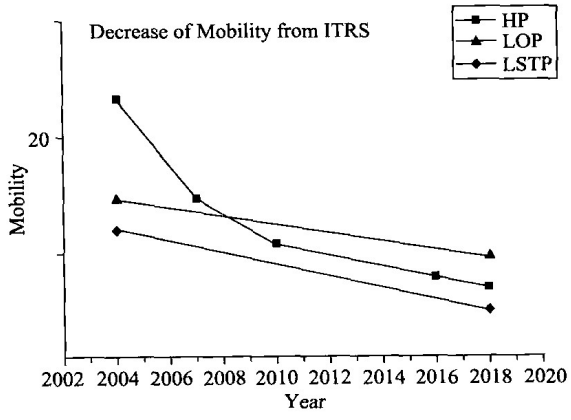

Fig. 3 The change of mobility of CMOS

We have three sets of data about CNT shown in Fig. 5. The EXP is an experimental data from IEDM, L_g = 260nm V_{dd} = 2.5V. Set Ⅰ and Ⅱ are deduced from EXP, with L_g = 25nm V_{dd} = 2.5V and L_g = 25nm, V_{dd} = 0.5V, respectively, with assumption of ballistic transport. We can see from Fig. 5 that a CNT of 260nm long has nearly the same delay as a CMOS transistor of 25nm long but a larger power dissipation. When we reduce CNT's length and keep voltage invariant as in case Ⅰ, it has a very small delay that CMOS can not reach but still a larger power dissipation. But when we reduce both the voltage and length, CNT has not only small delay but

also small power dissipation. We should notice in case Ⅱ the CNT has already had a E_b smaller than that of ultimate CMOS. And moreover in case Ⅱ CNT also overcomes the delay time barrier of CMOS.

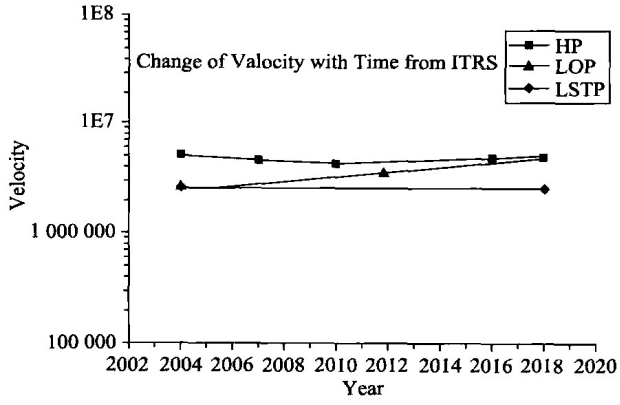

Fig. 4 The change of velocity of CMOS

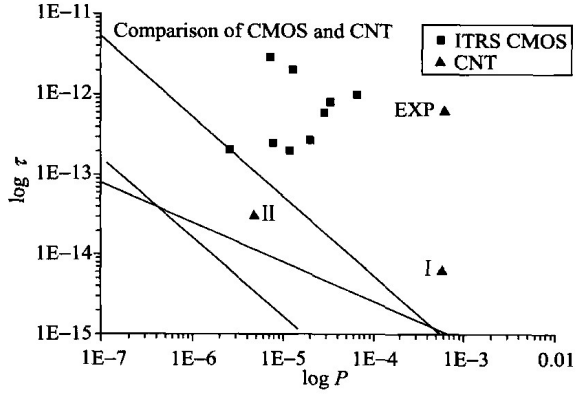

Fig. 5 Comparison of ITRS CMOS and CNT

2. Other Emerging Devices

Other emerging devices (generally called nano-devices) include RSFQ (Rapid Single Flux Quantum), Optical Switch. NEMS (NanoElectroMechanical Systems) Switch, Molecular Switch, etc. We won't discuss them in detail, they are still in stage of laboratory investigation. In Fig. 6 the corresponding data presented in ITRS (2003) for these devices are quoted, indicated and compared with Si CMOS and CNT.

As can be seen in the plot there is a much wider span of those kinds of devices. At this moment we ignore the two points. Bio and Quan, which will be discussed in section Ⅴ. We can see NEMS has the smallest power dissipation and largest delay, while Optic devices have the smallest delay but quite a large power dissipation. The others, Molecule Switch and RSFQ all have their own characters. Some of these data

exceed the Si CMOS limit thus able to improve the IC technology.

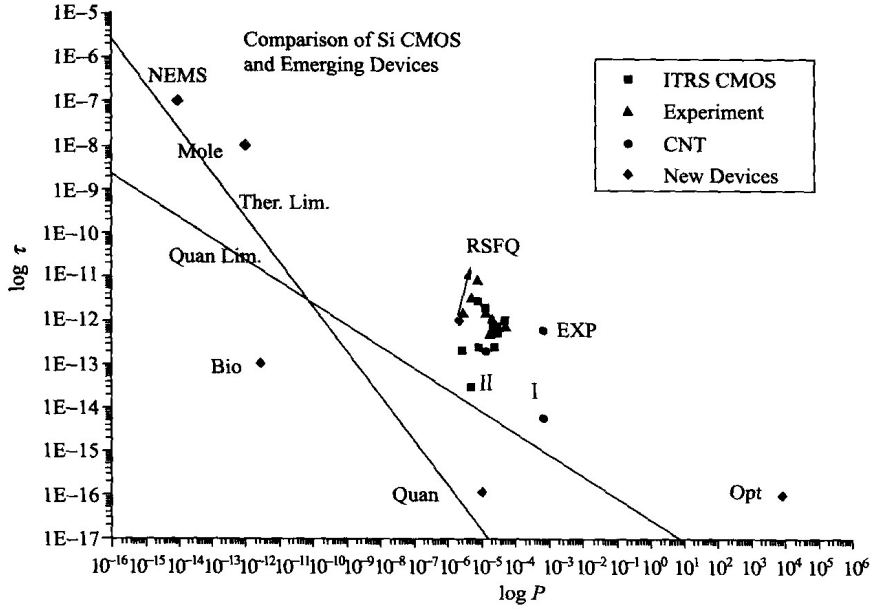

Fig. 6 Comparison of CMOS and emerging devices

Ⅳ. Limits of Classical 2-level Switching Devices

As shown in Fig. 6, although many emerging nanodevices can have performance (either speed-power product or speed and power separately) better than that of ultimate CMOS, they are still limited by two common lines corresponding to two fundamental physical principles of thermodynamics and quantum mechanics. Due to thermal noise the energy separation of two binary signal digits should be $E_b \geqslant \gamma kT$ ($\gamma = \ln 2$), as discussed by Shannon, where k is Boltzmann constant and T is absolute temperature. On the other hand by quantum uncertainty principle $E_b \tau \geqslant \hbar$, that is $P\tau^2 \geqslant \hbar$. Thus these two straight lines on the logτ vs log$P$ plot are the lowest boundary for all devices based on classical 2-level switching. We can see from Fig. 6 that any 2-level switching devices, including classical Si CMOS and emerging devices, does not exceed these limits.

Fig. 7 shows an ideal 2-level switching device. We assume when the particle on the left of the potential, it represents state "0" and when on the right it is state "1". The particle's passing through of the potential represents a switching transition from "0" to "1" or "1" to "0". We propose that the device is working in the critical region of quantum limit, so that $\Delta x \Delta p = \hbar$ and $\tau \Delta E = \hbar$, i. e. $\Delta x = \hbar / \sqrt{2m^* E_b}$ and $\tau = \hbar / E_b$. It indicates that by choosing appropriated parameter Δx (that is a in Fig. 7)