

Recent Topics for Realizing Low-Power, High-Speed VLSI's

Takayasu Sakurai

Center for Collaborative Research and Institute of Industrial Science, University of Tokyo

E-mail: tsakurai@iis.u-tokyo.ac.jp

Abstract

Power consumption of VLSI's will limit the ever-increasing trend of system integration, if it increases at the current rate. Thus the quest of low-power yet high-speed realization of VLSI's is essentially important to think about the scalability of MOS devices. In this paper, some of the recent topics for realizing low-power yet high-speed VLSI's are described. Focus is put on the approaches based on cooperation between levels, such as device-circuit cooperation and hardware-software cooperation.

1. Introduction

The supply voltage of VLSI's is ever decreasing to assure sufficient reliability of the thin gate oxide used in deep submicron transistor. Although decreasing voltage is beneficial in decreasing power consumption, the power consumption of VLSI's is still increasing as shown in Fig.1 and 2. The threshold voltage, V_{TH} , is to be lowered to maintain high-speed characteristics of MOSFET's but the low V_{TH} is the source of leakage current (Fig. 3,4). In this sense, low-power design in the low supply environments is a battle against the ever-increasing leakage current. Moreover, other leak component, such as gate tunneling leakage and junction leakage, will be added up as shown in Fig.5. To take trade-off between power and delay, many proposals have been made and they are categorized in Fig.6. In this paper, recent development to cope with leakage control is summarized putting stress on cooperation between levels. Other recent approaches to low-power and high-speed VLSI design are also described.

2. Device-Circuit Cooperation

Earlier proposals for suppress leakage in a standby mode have scalability problems as shown in Fig.7. In order to mitigate the leakage problem in a standby mode, it is effective to insert a non-leaking power switch in series to a normal logic gate block operating at low voltage less than 1V. The non-leaking power switch can be realized by a high- V_{TH} (0.6V for

example) MOSFET that is turned on in an active mode and turned off in a standby mode. Higher voltage like 1.5V-2.0V is applied to the gate of the MOSFET to achieve higher drivability and hence higher speed, which is called boosted gate MOS scheme (BGMOS). The power switch should have higher oxide thickness to endure the higher voltage and thus can be said the cooperation between technology level and circuit level. MOSFET's tuned for the higher voltage is also helpful in SRAM, I/O and analog designs (Fig.8).

3. Hardware-Software Cooperation

As a countermeasure to the leakage problem in an active mode, dual V_{TH} is a well known technique where the higher V_{TH} is assigned to gates in non-critical paths while the lower V_{TH} is applied to critical path gates and achieves power saving of about 30%. In cases, it is more effective to use V_{TH} -hopping where V_{TH} is hopping between two V_{TH} levels in time through the use of back-gate bias. For typical real-time multimedia applications, power saving of about 80% is expected. In V_{TH} -hopping, V_{TH} is controlled to be a little higher in high-frequency mode, while in half the frequency mode V_{TH} is controlled to be a little lower. It is responsibility of software to choose a frequency between the higher and the lower frequency without degrading the performance of an application. In this sense, this scheme is based on cooperation between software level and circuit level. There is an algorithm to guarantee real-time execution of software in application level and O/S level while achieving low power (Fig.9-18, Ref.1-5).

Another software-hardware cooperative scheme for low power is bus shuffling, where bus layout is shuffled according to signal properties on the bus. The bus layout is shuffled so as to minimize the power consumed by coupling capacitance among lines. Virtually no overhead is observed for the scheme, which is different from the earlier encoding schemes for low power. Bus shuffling needs signal pattern information and is applicable to special-purpose

systems. The power saving of about 40% is observed by this simple idea (Fig. 19-21, Ref.7). Software tool is developed to estimate power distribution in the circuit (Fig.22-26, Ref.6).

3. Other low-power and high-speed approaches

Abnormal Leakage Suppression (ALS) scheme is effective in reducing abnormal standby current caused by manufacturing fluctuations in SRAM's (Fig.27-28, Ref.8). Interconnection delay is hindering the high-speed VLSI realization. If interconnect delay can be mitigated, supply voltage can be further reduced so that further low power can be achieved. Dual-rail bus (DRB) scheme and of staggered firing bus scheme are two of the schemes to cope with interconnection delay problems (Fig.29-35, Ref.9).

References

[1] S. Lee and T.Sakurai, Run-time Power Control Scheme Using Software Feedback Loop for Low-Power Real-time Applications, " ASPDAC'00, A5.2, Jan. 2000.
 [2] S. Lee, and T.Sakurai, "Run-Time Voltage Hopping for Low-Power Real-Time Systems," Proceedings of Design Automation Conference, pp. 806-809, June 2000.

[3] Y.Shin, K. Choi, and T.Sakurai, "Power Optimization of Real-Time Embedded Systems on Variable Speed Processors," ICCAD'00, Nov. 2000.
 [4] H.Kawaguchi, Y.Shin, and T.Sakurai, "Experimental Evaluation of Cooperative Voltage Scaling (CVS): A Case Study, " IEEE Workshop on Power Management for Real-Time and Embedded Systems", Taiwan, May 2001.
 [5] Y.Shin, H.Kawaguchi and T.Sakurai, "Cooperative Voltage Scaling (CVS) between OS and Applications for Low-Power Real-Time Systems," CICC'01, pp.553-556, May 2001.
 [6] Y.Shin and T.Sakurai, "Estimation of Power Distribution in VLSI Interconnects," International Symposium on Low-Power Electronics and Design, June 2001.
 [7] Y. Shin and T. Sakurai, " Coupling-driven bus design for low-power application-specific systems, " Proc. Design Automation Conf. (DAC), pp.750-753, June 2001.
 [8] K.Kanda, N.D.Minh, H.Kawaguchi and T.Sakurai, "Abnormal Leakage Suppression (ALS) scheme for low standby current SRAM's," Digest of Papers, ISSCC'01, MP11.4, Feb. 2001.
 [9] K.Nose and T.Sakurai, "Two Schemes to Reduce Interconnect Delay in Bi-directional and Uni-directional Buses," Symp. on VLSI Circuits, pp.193-194, June 2001.

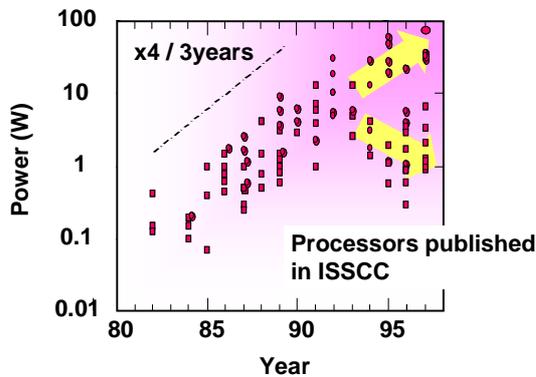


Fig.1 Ever Increasing VLSI Power

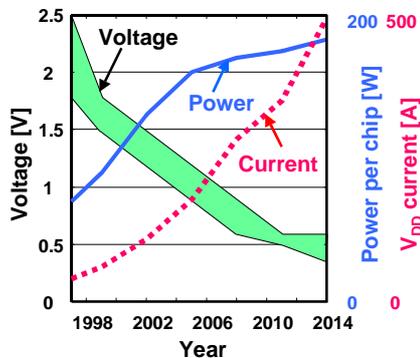


Fig.2 Trend in voltage and power (ITRS)

$$Power = p_t \cdot f_{CLK} \cdot C_L \cdot V_{DD} + I_0 \cdot 10^{\frac{V_{TH}}{s}} \cdot V_{DD} + Gate / Drain leak$$

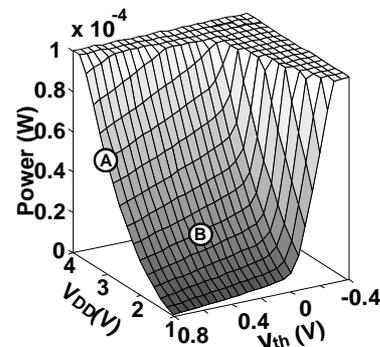


Fig.3 Power Dependence on VDD & VTH

$$Delay = \frac{k \cdot Q}{I} = \frac{k \cdot C_L \cdot V_{DD}}{(V_{DD} - V_{TH})^\alpha}$$

(α=1.3)

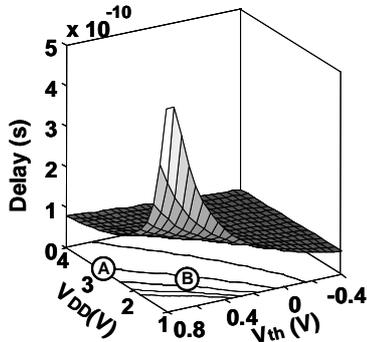


Fig.4 Delay Dependence on VDD & VTH

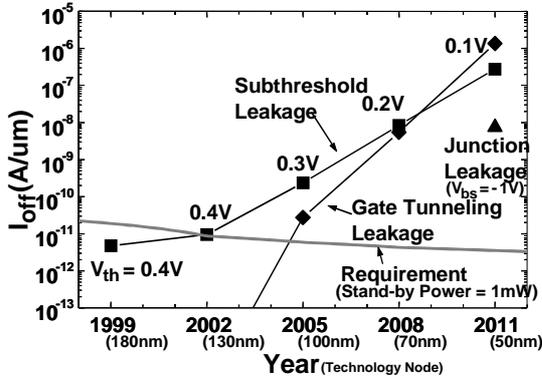


Fig.5 Transistors go leaky

Low power → Low V_{DD} → Low speed → Low V_{TH} → High leakage → V_{DD}-V_{TH} control

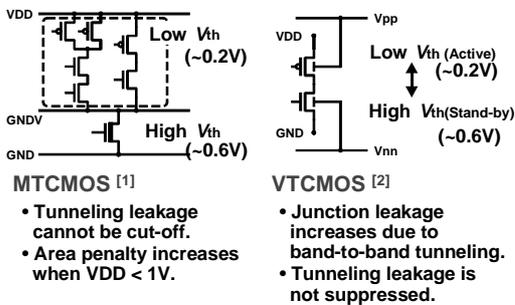
	Active	Stand-by
Multiple V _{TH}	Dual-V _{TH}	MTCMOS
Variable V _{TH}	V _{TH} hopping	VTCMOS
Multiple V _{DD}	Dual-V _{DD}	Boosted gate MOS
Variable V _{DD}	V _{DD} hopping	

Software-hardware cooperation (points to V_{TH} hopping and V_{DD} hopping)

Technology-circuit cooperation (points to Boosted gate MOS)

- *) MTCMOS: Multi-Threshold CMOS
- *) VTCMOS: Variable Threshold CMOS
- Multiple : spatial assignment
- Variable : temporal assignment

Fig.6 Controlling VDD and VTH for low power



[1] S.Mutoh et al. IEEE, JSSC, 1995. [2] T.Kuroda et al. IEEE, JSSC, 1996.

Fig.7 Scalability problems of early proposals

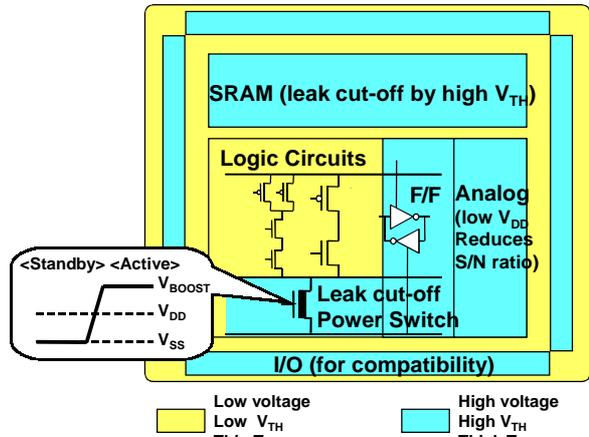


Fig.8 Multi Tox scheme: device-circuit cooperative approach for future low-power GSI

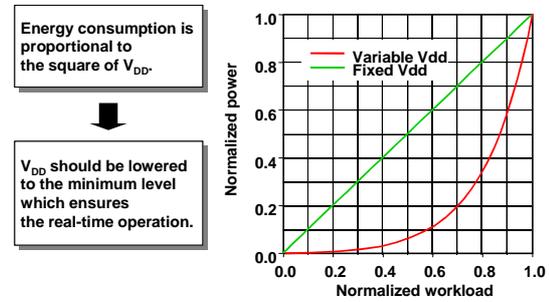


Fig.9 V_{DD}-hopping basics. If you don't need to hussle, V_{DD} should be as low as possible

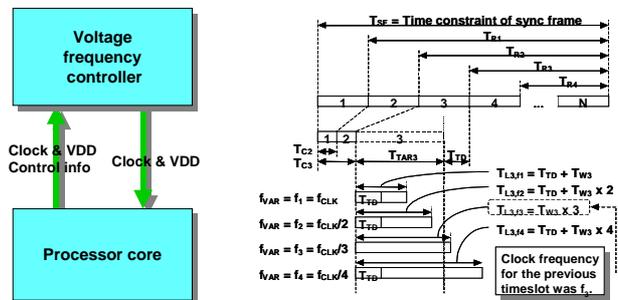


Fig.10 Application slicing and software feedback loop in Voltage Hopping

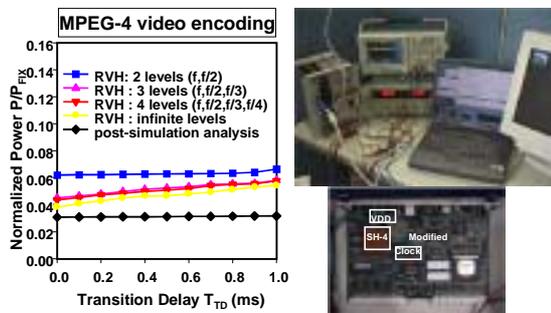
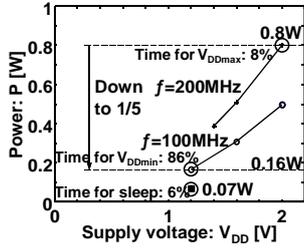


Fig.11 Run-time Voltage Hopping reduces power to less than 1/10

Total power = $0.8 \times 0.08 + 0.16 \times 0.86 + 0.07 \times 0.06 = 0.2W$



VDD hopping can cut down power consumption to 1/4

Fig.12 Measured power characteristics

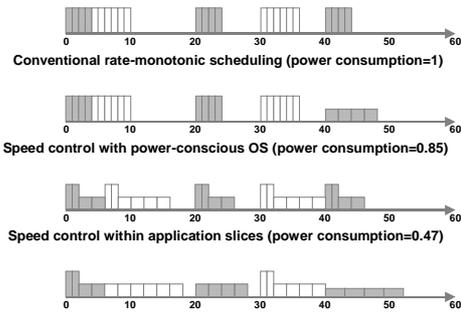


Fig.13 Power Conscious OS & Application Slicing. Cooperative Voltage Scaling (CVS) between OS and Applications for Low-Power Real-Time Systems

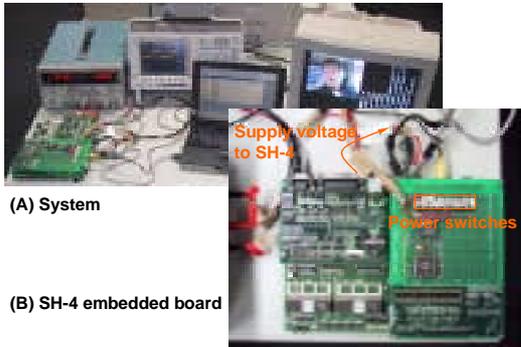


Fig.14 Hardware for Cooperative Voltage Scaling (CVS)

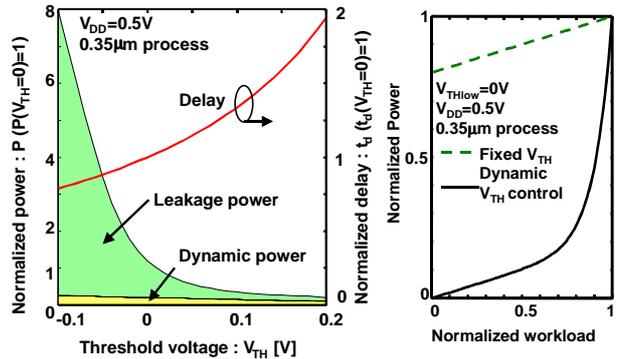


Fig.15 VTH-hopping basics

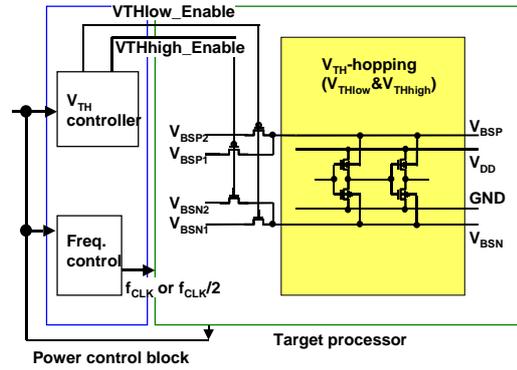


Fig.16 Schematic of VTH-hopping

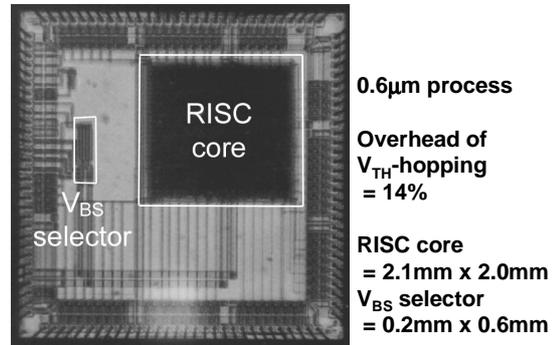


Fig.17 Microphotograph of RISC processor

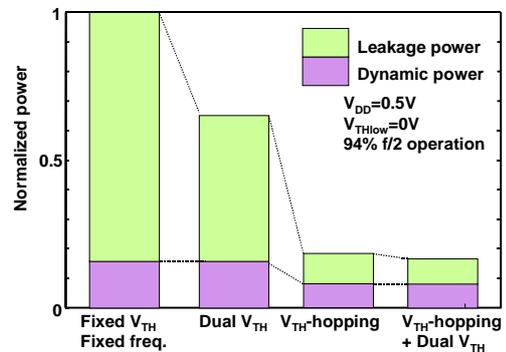


Fig.18 Power comparison

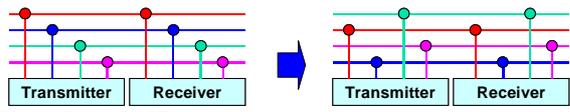


Fig.19 Bus shuffling scheme

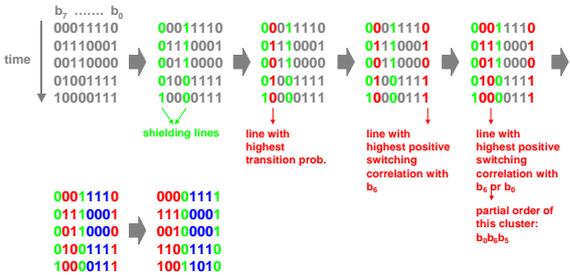


Fig.20 Proposed heuristic bus shuffling algorithm

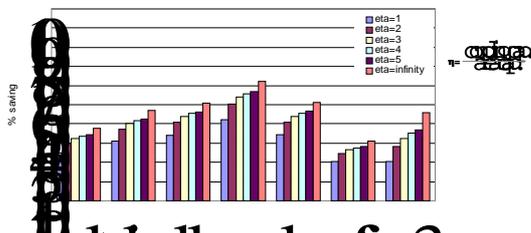


Fig.21 Measured result for dual-rail bus

- Reliability problem
 - Current density in metal lines increases
 - Temperature of interconnect increases
 - MTF (Mean Time to Failure) decreases
- Problem of power distribution estimation

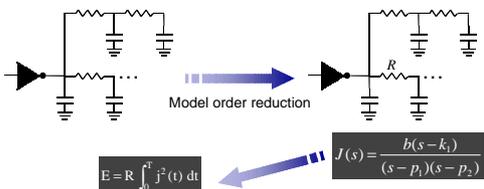


Fig.22 Power Distribution Estimation

- Theorem
 - If the Laplace transform of a time-domain signal $j(t)$, denoted by $J(s)$, has q simple poles in the left half of the s -plane,

$$\int_0^{\infty} j^2(t) dt = \sum_{i=1}^q r_i J(-p_i)$$

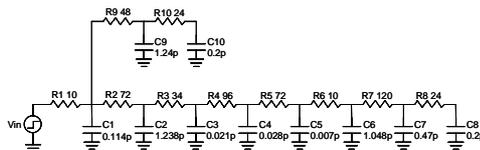
r_i : residue of $J(s)$ at the pole p_i of $J(s)$

Fig.23 Fundamental theory in Power Distribution Estimation

- Prototype tool
 - SPICE-in and power-out
 - Moment matching-based model order reduction
- Estimation accuracy
 - Source of error: area under the square of $j(t)$
 - Comparison with SPICE

$$E = R \int_0^{\infty} j^2(t) dt$$

Fig.24 More than 1000 times faster heat generation simulator than SPICE



Resistor	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	Avg. error	Max. error
SPICE	5.12	8.42	0.88	2.42	1.76	0.24	0.43	0.01	5.54	0.05		
1-pole	3.12	7.18	0.89	2.43	1.76	0.24	0.41	0.01	4.69	0.04	9.4%	39.1%
2-poles	4.81	8.39	0.88	2.42	1.76	0.24	0.44	0.01	5.53	0.05	1.2%	5.9%
3-poles	4.96	8.38	0.88	2.42	1.76	0.24	0.43	0.01	5.50	0.05	0.5%	3.2%

Fig.25 Numerical example of the power consumption estimation (or heat generation)

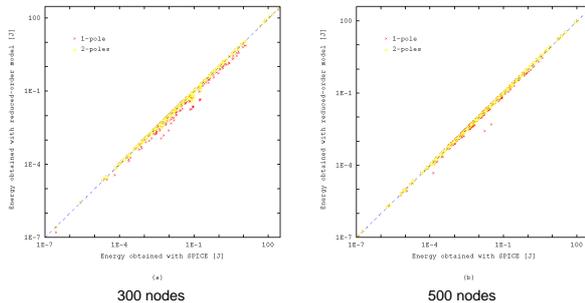


Fig.26 Experimental Results for Randomly-generated circuits

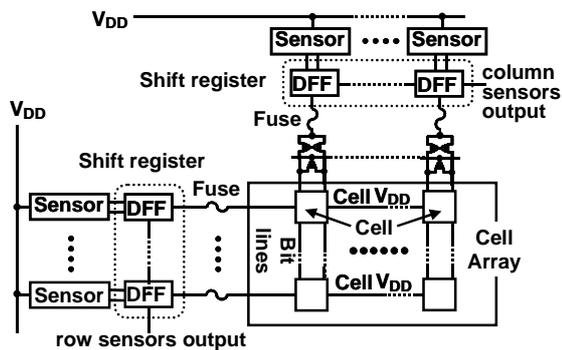


Fig.27 Whole structure of Abnormal Leakage Suppression (ALS) SRAM

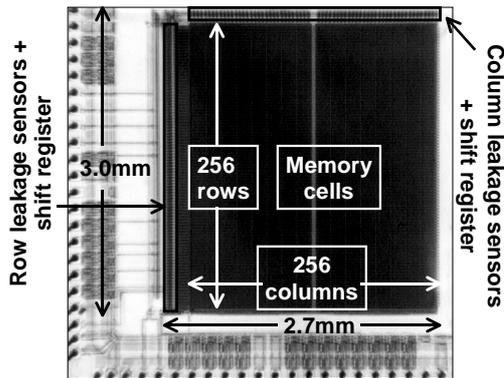
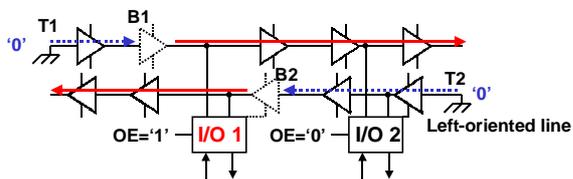


Fig.28 Test chip fabricated by 0.6mm design rule. ALS detects 1μA order leakage current and area overhead is about 1% in 4Mb SRAM.



- Two buffered interconnects per bit
- All nodes before B1 and B2 are kept '0'.

Fig.29 Dual-rail bus (DRB) scheme

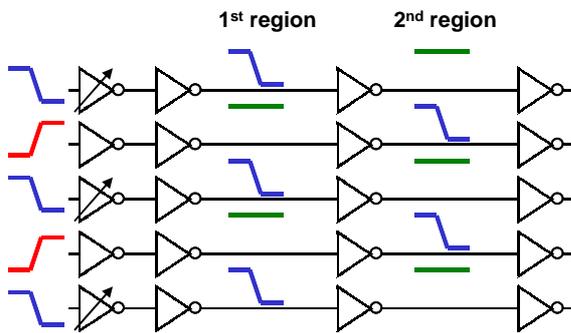
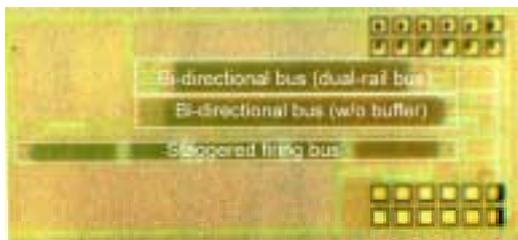


Fig.30 Operation of staggered firing bus scheme



- 0.13μm CMOS process
- 10mm bus length
- 5 buffers (dual-rail bus), 11 buffers (staggered firing bus)

Fig.31 Microphotograph (0.13mm process)

0.6μm process ($L_{INT}=60mm$)

conv. (w/o buffer)	17.7 ns	31% improved	Dual-rail bus	12.2 ns
--------------------	---------	--------------	---------------	---------

0.13μm process ($L_{INT}=10mm$)

conv. (w/o buffer)	2.25 ns	44% improved	Dual-rail bus	1.27 ns
--------------------	---------	--------------	---------------	---------

- To make the comparison fair, line width and spacing are doubled for the conventional scheme (w/o buffer).

Fig.32 Measured result for dual-rail bus

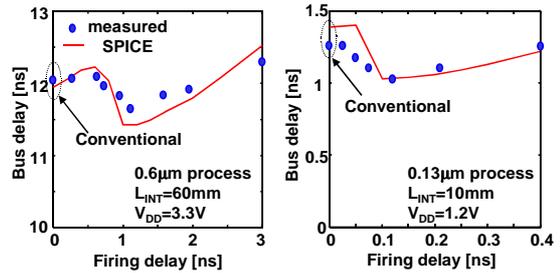


Fig.33 Measured result of staggered firing bus

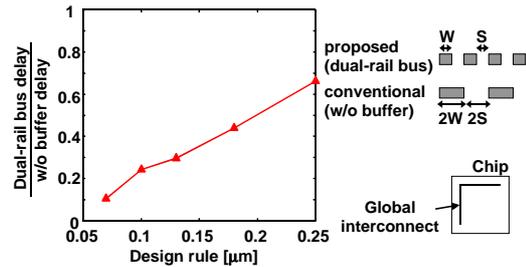


Fig.34 SPICE estimation of benefit of dual-rail bus scheme for the future

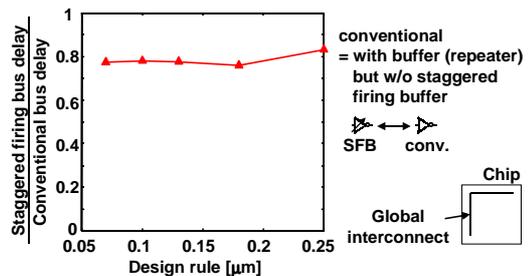


Fig.35 SPICE estimation of benefit of staggered firing bus scheme for the future