

NOVEL LOCAL BIT LINE DESIGN BASED ON FORCED-KEEPER TECHNIQUE FOR ON-CHIP MEMORIES

Zezhong Yang¹, Jinhui Wang^{1*}, Lina Wang¹, Ligang Hou¹, Na Gong²

¹VLSI and System Lab, Beijing University of Technology, Beijing 100124, CHINA

²Dept. of Electrical and Computer Engineering, North Dakota State University, ND 58102, USA

*Corresponding Author's Email: wangjinhui@bjut.edu.cn

ABSTRACT

A local bit line (LBL) based on forced-keeper technique for on-chip memories is proposed in this paper. The keeper transistor is forced to be turned off adaptively to decrease the leakage current and the contention current to achieve high performance. With SMIC 65 nm technology, the simulation results show that the proposed LBL chosen by the row decoder achieve 89.9% Power-Delay-Product (PDP) reduction. And the LBL not chosen by row decoder achieve 94.1% and 93.8% power saving with the active clock. The larger memories scale are, the more total power is saved.

INTRODUCTION

As key parts in microprocessors, register files determine the overall performance of microprocessors. The register files are on the critical path and their read delay limits the achievable maximum operation frequency [1]. Therefore, dynamic circuits are used for LBL and global bit lines (GBL) to speed up the evaluation. As the bit lines takes up to more than 70% of the dynamic power consumption of the register files [2], the power has become a main challenge to design LBLs, especially for the register files of battery-powered devices.

There have been many efforts to reduce the power of LBLs, such as scaling down supply voltage [3], adaptive keeper [1], tri-state technique [4], leakage-proof technique [5-6], and sleep vector control for bit line [7]. Scaling down supply voltage is known to be the most effective way to reduce power, but it's hard to meet the performance requirements of speed and noise immunity; In [1], the adaptive keeper is used to reduce the contention current of LBL during pull-down operation, but ignore the power of LBL which is not pulled down or which is not chosen by the row decoder. If there is a conductive pull-down path, the LBL must be in chosen state (CS) by row decoder. In active mode, the LBL in CS amounts to a small percentage of the total LBLs. And in standby mode, all of the LBL is not chosen by row decoder. Register files are extremely large in modern microprocessors. For example, the level-3 cache of Intel Core i7 is as large as 8MB. So the power of LBL in not chosen state (NCS) by the row decoder cannot be ignored. In [4-7], the power is reduced effectively, but all of these techniques reduce power when memories are in sleep mode when the clock of LBL is fixed. LBLs work until it is chosen by row

decoder. And, these techniques cannot decrease the power when memories are in active mode, and meanwhile they influence the speed [5-6] of memories.

To solve the problems presented above, a novel LBL based on forced-keeper technique is proposed in this paper, which can reduce the power of LBL in CS and NCS, meanwhile the LBL has a higher speed than the conventional one.

PROPOSED NOVEL LBL

Figure 1 shows the schematic of proposed novel LBL with N pull-down paths. Three transistors M1, M2 and M3 are added to the conventional LBL. A transmission gate consists of M2 and M3 can cut off the input (P) of keeper from the output of LBL node OUT controlled by signal T and TB. And Vdd is connected to P through M1.

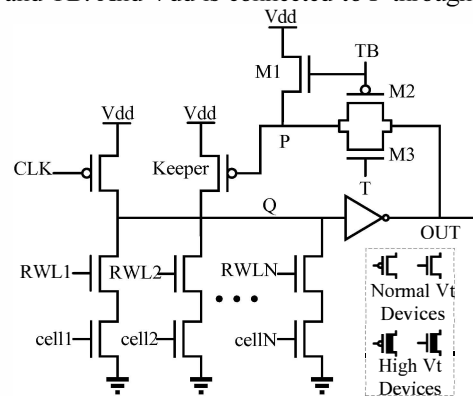


Figure 1: Proposed novel LBL

Proposed LBL in not chosen state by row decoder

In active mode, most of LBLs are in NCS. The LBL in NCS contributes to the most power consumption of a large scale register files. In this state the evaluation node Q cannot be discharged to Gnd, so the node OUT is always low. In the conventional LBL, OUT is connected to the gate of the keeper directly. As process advances to sub-65 nm, the gate oxide thickness gets scaled down to 12-16Å [6], such thin oxide leads to great gate leakage currents by various direct tunneling mechanism [5], [8]. As shown in Figure 2, in conventional LBL, the gate of the keeper is connected to Gnd, while the drain and source are connected to the Vdd, so there is leakage current from drain and source to Gnd through the gate and the NMOS of the inverter. Meanwhile, the leakage current from drain to gate get the charge of drain less, so the keeper sup-

plement the charge continuously. The current above is one dominant source of total power in NCS.

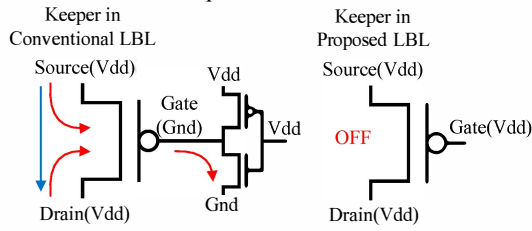


Figure 2: Keepers of different LBL in NCS

When the proposed LBL is in NCS, T is low, TB is high, respectively. M1 is turned on, and M2 and M3 are turned off. So the transmission gate is turned off and node P is charged to Vdd. As shown in Figure 2 the keeper is forced to be turned off, which has no influence on the pre-charge stage and the charge of Q, so the proposed LBL can wait to be chosen at any time in NCS of active mode.

Proposed LBL in chosen state by row decoder

For the conventional LBL, node P and Q are connected directly. At the beginning of evaluation stage Q is discharged to Gnd, the speed is decreased and the power is increased due to the large contention current of the keeper [1].

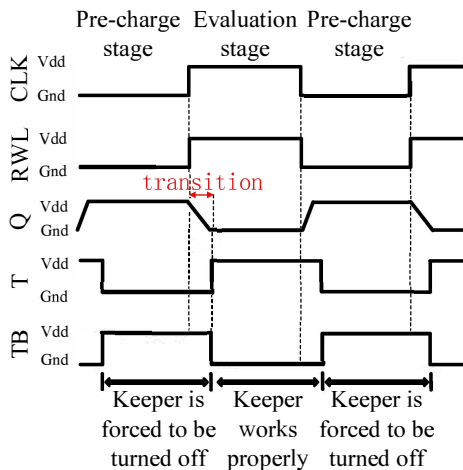


Figure 3: Timing constraint of CLK, RWL, Q, T, and TB

In this state, it's hypothesized that the LBL is always chosen and there is a conductive pull-down path in evaluation stage. The principle of the proposed technique is shown in Figure 3. In this proposed LBL, at the beginning of evaluation stage, the transmission gate is turned off, and node P is charged to Vdd, so the keeper is forced to be turned off. Therefore, at the process of discharging Q to Gnd, the speed is higher and the power is smaller than the conventional one, because the contention current is almost eliminated. After this short transition process, T and TB are set to Vdd and Gnd, alternatively, and the keeper works properly to increase the robustness to noise as the node P is connected to Q. In the pre-charge stage, the power is also reduced as the same reason with the LBL in

NCS. Notes that if a LBL is in CS the timing constraint of CLK, T, and TB is not changed, whether there is a conductive pull-down path or not.

External circuits of LBL in SRAM

To implement the proposed technique, a method is proposed to generate T and TB. RQ is the output of the row decoder, and RQB is the reverse of RQ. The row decoder is designed that the output is the output of AND gate, so the RQB can be extracted from the output of the NAND gate which consist in the AND gate. As shown in Figure 4, the delay array is used for delaying RQ and RQB to T and TB to meet the timing constraint in Figure 3. The delay time from RQ and RQB to T and TB, respectively, can be adjusted by changing the voltage of Vadj0 and Vadj1. The architecture of a SRAM based on proposed LBL is shown in Figure 4. The function of row decoder is to judge which LBL is chosen, and column decoder is to judge which RWL in the LBL is high. Thus, some stored data in cell can be output through the GBL through the control of the decoder circuit.

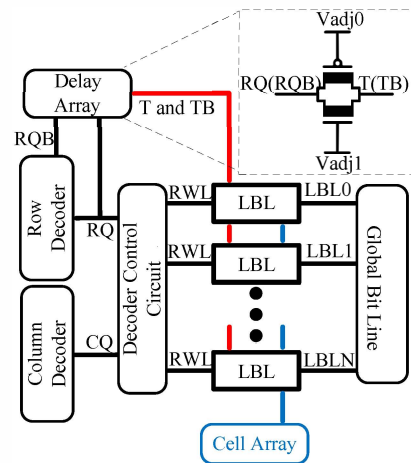


Figure 4: Block diagram of the SRAM

SIMULATION RESULTS

With SMIC 65 nm [9] technology, 4 kinds of LBL are simulated. They are, respectively, 2 proposed LBLs with 8 and 16 pull-down paths, and their corresponding conventional ones in CS and NCS. All of the 4 LBLs are turned to operate at 1GHz clock frequency and the power supply is 1.2 V. In NCS, the CLK is active and there are no conductive pull-down paths. The probability is 50% that there is a conductive pull-down path In CS.

The simulation results are shown in TABLE I. As the contention current and leakage current is reduced, the proposed LBL with 8 and 16 pull-down paths in chosen state (8P-CS, 16P-CS) reduces delay by 85.2% and 93.7%, power by 31.7% and 38.1% (the power of delay array is counted), and PDP by 89.9% and 89.9% than conventional ones. And the proposed LBL with 8 and 16 pull-down paths in not chosen state (8P-NCS, 16P-NCS)

reduce the power by 94.1% and 93.8%, respectively, as the leakage current is reduced.

A parameter R is defined in (1), which can reflect the scale of a SRAM.

$$R = N_{ALL} / N_{NCS} \quad (1)$$

Where N_{NCS} is the number of LBLs which are chosen in some clock cycle of active mode, N_{ALL} is the number of the total LBLs. As shown in Fig 4, the larger memories scale are, the more total power is saved. And the maximum of the total power saving in active mode is 94.1% and 93.8%, respectively.

In standby mode, all of the LBLs are in NCS, so the power saving of standby mode equals the power saving in NCS, and the maximum of the total power saving in active mode as well.

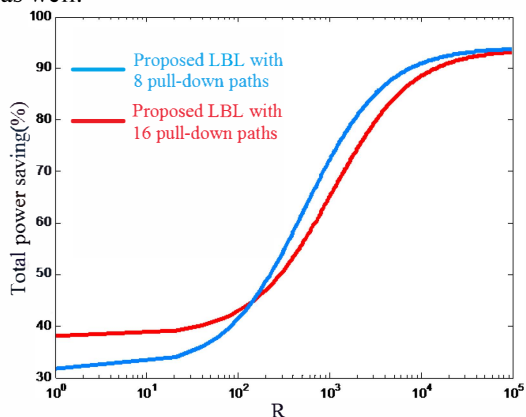


Figure 4: Total power saving of the proposed LBL in active mode

SUMMARY

A novel LBL based on forced-keeper technique for on-chip memories is proposed in this paper. The proposed LBL reduces the leakage current and the contention current in the evaluation stage to reduce delay time and power. With SMIC 65 nm technology, the simulation results show that 8P-CS, 16P-CS reduce delay by 85.2%

and 93.7%, and power by 31.7% and 38.1% than conventional ones, respectively. 8P-NCS, 16P-NCS reduce the power by 94.1% and 93.8%, respectively, with active clock. It is shown that the larger memories scale are, the more total power is saved.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No.61204040, 60976028), Beijing Municipal Natural Science Foundation (No.4123092), Ph.D. Programs Foundation of Ministry of Education of China (No.20121103120018), and Plan Program of Beijing Education Science and Technology Committee (No. JC002999201301).

REFERENCES

- [1] N. Gong, et al. *Proceedings of IEEE International Systems on Chip (SOC) Conference*, 2011, pp. 30-35.
- [2] N. Gong, et al. *Electronics letters*, vol. 48, 2012, pp. 1104-1105.
- [3] K. Takeuchi, et al. *Proceedings of IEEE International Solid-State Circuits Conference*, 2011, pp. 514-515.
- [4] N. Gong, et al. *Proceedings of the IEEE International Systems on Chip (SOC) Conference*, 2011, pp. 323-328.
- [5] G. Yang, et al. *Proceedings of 17th International Conference on VLSI Design*, 2004, pp. 222-227.
- [6] J. Wang, et al. *Proceedings of 8th International Conference on Solid-State and Integrated Circuit Technology*, 2006, pp. 1864-1866.
- [7] N. Gong, et al. *IEEE Transactions on Circuits and Systems I*, vol. 0, 2014, pp. 1-14.
- [8] H. Momose, et al. *IEEE Transactions on Electron Devices*, vol. 43, 1996, pp. 1233-1242.
- [9] SMIC, <http://www.smics.com/eng/index.php>

TABLE I. SIMULATION RESULT OF PROPOSED AND CONVENTIONAL LBL

| Different Circuits | Simulation Result Of Proposed And Conventional LBL | | | | |
|--------------------|--|-----------|----------|-----------|------------------|
| | Delay(ns) | Reduction | Power(w) | Reduction | Reduction of PDP |
| 8C-CS | 0.115 | - | 5.18E-06 | - | - |
| 8P-CS | 0.017 | 85.20% | 3.54E-06 | 31.70% | 89.9% |
| 8C-NCS | - | - | 9.68E-09 | - | - |
| 8P-NCS | - | - | 5.73E-10 | 94.10% | - |
| 16C-CS | 0.153 | - | 6.90E-06 | - | - |
| 16P-CS | 0.025 | 93.70% | 4.27E-06 | 38.10% | 89.9% |
| 16C-NCS | - | - | 6.51E-09 | - | - |
| 16P-NCS | - | - | 4.05E-10 | 93.80% | - |