# Design of High Performance SRAM Based on Single-Port Sense Amplifier

Shunrui Li[1, a *], Zuocheng Xing[2,b], Jianjun Chen[3,c] Zhentao Li[4,d]

[1]Collage of computer science of Nationl University of Defense Technology, Hunan Changsha 410073,China

[2]Collage of computer science of Nationl University of Defense Technology, Hunan Changsha 410073,China

[3]Collage of computer science of Nationl University of Defense Technology, Hunan Changsha 410073,China

[4]Collage of computer science of Nationl University of Defense Technology, Hunan Changsha 410073,China

[a]email: lisr824467100@163.com, [b]email: zcxing@nudt.edu.cn,

[c]email: cjj192000@163.com, [d]email: vlsi_report@yeah.net

**Keywords:** SRAM circuit design; low power consumption; high performance;8T-SRAM[1]

**Abstract.** Through the rapid development of integrated circuits,high performance and low power consumption always as the same goal for the IC designer, Especially in SRAM design.This paper described a full-custom design of high performance SRAM,the full-custom design using 40nm process, based on a new single-port sense amplifier to the traditional full-custom on the read out circuit,the timing and the performance has been greatly improved,the power consumption has been greatly reduced.So,compared with the traditional full-custom,the new design have a great advantage of the high performance and low power consumption.

## Introduction

For the DSP chip design in nowaday, the cache architecture uses two stage, one is the first-level cache (L1), and the other is the second-level cache (L2). L2cache chip design, as the maincenter of the storage path in chip processor core, is the critical path of nuclear CPU corepac's throughput data, and also considered to be the data sharing interface between corepac and other nuclear. So L2cache's performance is criticial influence to the overall performance of the chip. As the result, in a certain degree, improve the performance of L2cache can greatly improve the performance of computer systems. The key to improving L2cache's performance is to enhance the performance of a single internal L2 SRAM bank.

For low-power high-performance SRAM design, academic and business-related research is currently focused on the 90nm to 28nm technology. According to the annual Technology Trends 2014 International Solid State Circuits Conference (ISSCC conference), pointed out three design challenges: 1) increasing leakage power and dynamic power consumption brings low-power design challenge; 2) process down, especially after scale down to 30nm technology,increasing low process voltage disturbance cause reliability design challenges; 3) Under new process node ，According to Moore's Law that the speed of 0.5 times reducing chip area, it brings the challenges of circuit layout cnstomization design. This fully shows that in the 40nm process, the deeply study of the method of high-performance low-power SRAM is very important. For dynamic power consumption, the use of low voltage is an effective way to reduce power consumption, and also is the most directly and effective way for the static power. So low-power SRAM must ensure that it can running reliably even at the lowest VDD case. But with the shrinking process nodes, low voltage has been unable to go further, of which the memory cell circuit, read-and-write circuit, SA circuit and SAE generating circuit that will be extremely sensitive to PVT disturbance at a low voltage, high temperature and

extreme process corner, and low power SRAM design is particularly complicate. The key technology of low-power SRAM design is concentrated around the core circuit to start over.

Through the study of traditional memory, this paper starts from two angles-performance and power consumption, and focuses on the core module of SRAM circuit --- sense amplifier SA, analyze and design a sensitive amplifier circuit which is suitable to use in the SRAM of L2cache. The capacity of this SRAM bank is 512X10, and 8T memory cell is used to achieve the desired performance and low power requirements. Through this design, gives out a summary of some important methods about SRAM design. Known these methods will be great helpful for the future of SRAM designers.

## The Circuit Structure

The memory of the Full-custom SRAM512X10 is made of the storage array which formed by 8T cells, and based on this storage array, then build the peripheral circuit of the memory which includes the latch module, the decoding module, the I / O module and the clock-gating module. Based on the reliability, the design of the SRAM also pursues optimum design, such as minimum delay and reduces the power consumption to least, and so on.

## The Circuit construction

The circuit structure of the SRAM is Composed of the address latch, the data latch, reading and writing decoder, the storage array, the I / O circuit and the clock-gating circuit. The structure diagram is showing in Fig 1:
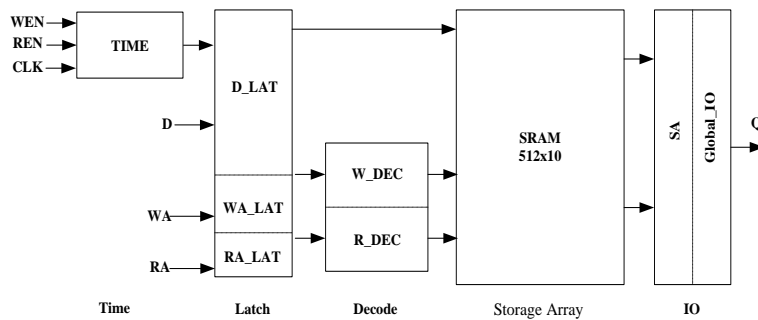


Fig 1    The diagram of SRAM

Under the control of the chip enable signal and write/read enable signal, SRAM generating the corresponding signals of the signals, including address latch clock, the data latch clock, read clock, the read decode clock, the write clock, the write decode clock ,and the control clock of sensitive amplifier. During the time of valid write enable signal, the write address is latched by latch clock, and at the same time, the input data is also latched. Then decoder begin to decode the write address, the pass transistor of the corresponding memory cell will turn on by the word line that decoded from write address, and at this time, the data has been latched by latch module, and delivered to the bit line, waiting to be written. So, when the transfer tube is opened, the data could be written to the memory cell directly to complete the write operation. When read enable valid, in the same way, the read address is latched, the decoder opens the corresponding word line so that make the bit line discharge, and let the stored memory cell's value quickly appears at the output port under the influence of the sensitive amplifier.

The key point in the SRAM circuit design is the design of the read circuit. The performance of the read circuit determines the SRAM's speed. As shown in the figure 2, the read circuit becomes the critical path of the SRAM, and reducing the delay of the critical path can effectively increases the storage's performance.
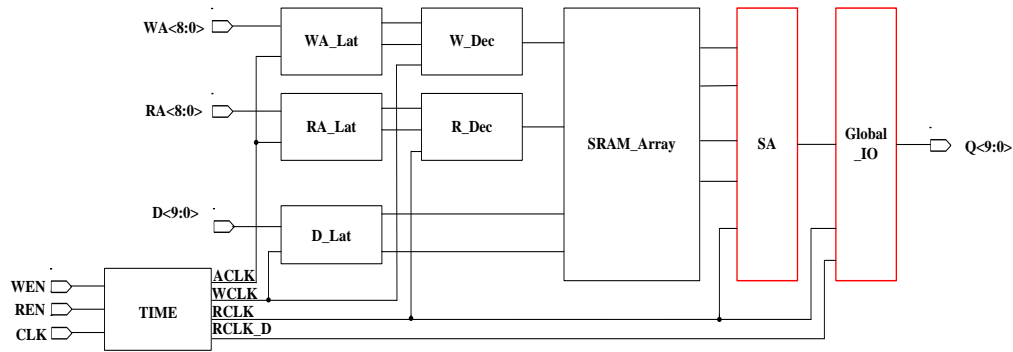
Fig 2 The structure of SRAM512X10

## The structure of the Traditional single-port read circuit

The read circuit of the traditional single-port is designed based on the 8T SRAM. Compared with 6T SRAM, because of having the better stability, it is higher in reliability. As joins the column selection control into the 8T storage cell, the storage cell includes 10T actually. Although the number of tube increased, the length of the bit line can be greatly reduced, which can highly improve the speed of charging and discharging , and enhances the performance. The 8T memory cell structure used in the design shows in Figure 3:
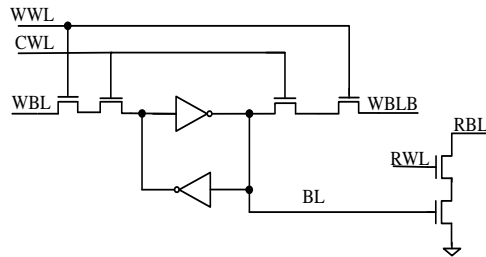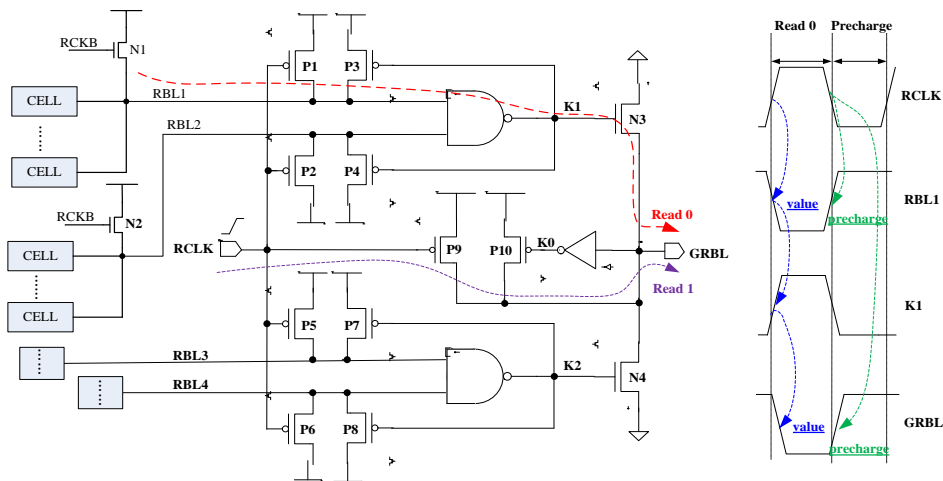


Fig 3 The circuit of 8T cell



Fig 4 The conventional read circuit of 8T SRAM

The conventional read circuit of 8T single-ended is shown in Fig 4. It is realized by the way of the dynamic circuit, such as precharge and evaluation, to select and output the data of bit line. As we can see from the output waveform of the 0-read and 1-read, in which RCLK is read gated-clock, when RCLK is low, the read circuit is in a pre-charge stage, the five bit lines, RBL1, RBL2, RBL3, RBL4 and GRBL global line, will be precharged to VDD. when RCLK is high, the read circuit is in

189

the evaluation phase, and by this moment, precharge transistor is off, and there is only one of these four bit lines is valid .

The data path of 0-read and 1-read is shown in figure4. When reading 1, all of the four bit lines are 1 so that both of the K1 and K2's value are 0, N1 and N2 are closed, GRBL is 1 as the pre-recharge value, so the value of the RBL reached to the GRBL. When reading 0, if RBL1 = 0, the RCLK is high, P1, P2, P5 and P6 are closed and the RWL is effective, RBL1 discharged to a low level, for RBL2=1,BL3=1,RBL3=1,so K1 = 1, K2 = 0. Then, at this same time, N1 turns on, the GRBL is pulled down to a low level, so that RBL1 = 0 value reaches to the GRBL, GRBL=0.

It can be seen from the waveform of the traditional reading circuit that the bit line voltage swing is from 0 to VDD during the pre-charge and evaluation phases. With the depth of the SRAM increasing, the time of both precharge and discharge time for the bit line will greatly increase. It's not good for seeking high-performance. Additionally, it will generate a great dynamic power，which will not be benefit for reducing power consumption.

**The new single—port Sense Amplifier read circuit structure**

The new reading circuit includes two parts, which are sense amplifier[2] and feedback circuit[3].
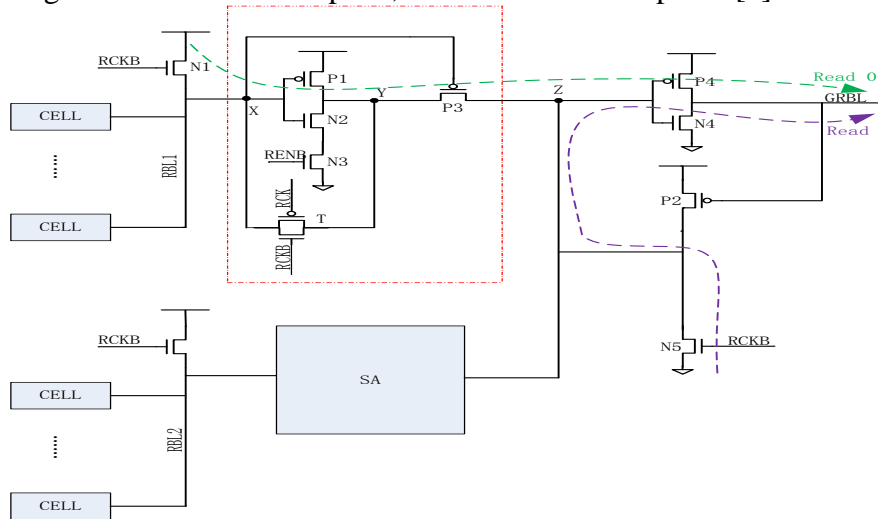


Fig 5 The circuit of sense amplifier

Firstly introduce about the structure of the sensitive amplifier circuit. The structure of the sensitive amplifier circuit is shown in Figure 5. This Sensitive amplifier circuit is a single port sense amplifier, which in the sense amplifier module is made up of a static complementary inverter INV, a balance PMOS transfer P3, a transfer gate T, the NMOS transistor N3 which is under the control of the enable signal, and a precharge NMOS N1. Among those components, the inverter, in which the pull-down network of it is control by the enable signal RENB. This can effectively reduce the static leakage current of the sensitive amplifier. RENB is an inverted signal of the read enable signal, due to the effective read enable signal is low, which means REN = 0, that can ensure RENB = 1 during the enable read period, which means the read enable signal of the sensitive amplifier is effective. At the rest of the time, REN = 1 RENB = 0, the sense amplifier doesn't work, N3 and N2 can series to form the DIBL effect which can reduces the static leakage current of the sense amplifier.

The working process of the sensitive amplifier is also divided into two stages, the pre-charge stage (RCKB = 1) and the evaluation phase (RCKB = 0). The clock signal RCK and RCKB are complementary signals. In the pre-charge stage, RCK = 0, RCKB = 1. RBL1 and RBL2 are precharged by transistor N1    and the transfer tube T is keeping open in the pre-charge stage, then X and Y are connected. So the voltage of X and Y are equal in the pre-charge stage. It should be noted that the voltage of RBL1 and RBL2 is lower than VDD after the precharge have done，    the voltage is clamped at the trip-point of the inverter, which is due to a voltage by the inverter can be

clamped around the INV deflection points. During in the pre-charging NMOS transistor N5 also open to pre-charge to Z points, but pre-charge point Z is to pull down the voltage of Z point to GND. In the evaluation phase, RCK = 1, RCKB = 0, the precharge transistor N1 and N5 is closed, and the transfer T is off too. When reading 0, supposed that the value 0 will appeared on the RBL1, then the bit line RBL1 will discharging, due to the bit line voltage value is at the trip-point of the inverter, so the voltage gain is very big at this point. When the bit line voltage drops, the output voltage of y point is quickly pulled high, then P3 MOSFET opens, Z point is pulled high, output GRBL = 0 by inverter, and P2 is turned on at the same time, making the voltage of Z pulled up faster to the steady high level state. P2 tube formed a positive feedback circuit, which can speed up the sense amplifier's speed. When reading 1,it will be auto-zero[4], all of the bit line voltage will not drop, the voltage of the X point will not up or down, so the valtage of the Y point will not drop, P3 will not open, the voltage of Z point will remain low, outputs GRBL = 1, the feedback circuit is closed. The waveform of the Sensitive Amplifier in precharge and evaluation stage is shown in Fig 6.
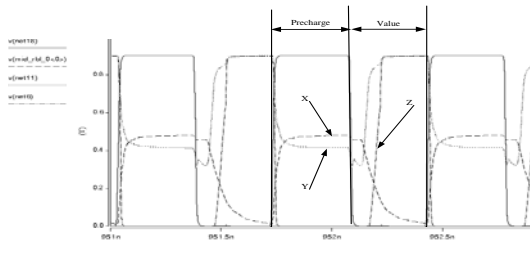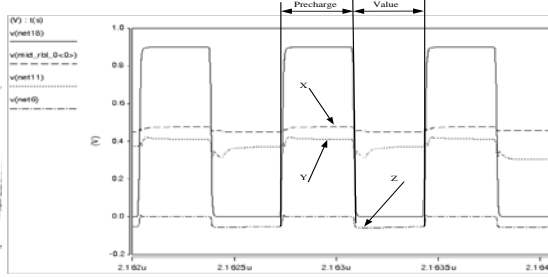


Fig 6 The waveform of sense 0                    Fig 7 The waveform of sense 1

In order to speed up the SA,it is necessary to make sure that the voltage of the X point as near as possible to the trip-point of the inverter at the pre-charge stage. The voltage of the X point is decided by the pre-charge transistor N1, the transfer T and the pull-down net of the inverters, and by changing the size of each MOSFET can make sure the voltage of the X-point keep steady in the trip-point, and enhance the sensibility of the SA. Using the NMOS to pre-charge can reduce the pre-charge time. Make the precharging voltage slightly higher than the inverter trip-point to make the SA insensitive to the leakage of the LBL, and avoid SA reading wrong result because of the static leakage, so precharging in NMOS can enhance the noise immunity of the SA.

As the pre-charge voltage is lower, the charging time is shorter, the discharge speed is faster, the speed of SA is higher, the voltage swing of bit line is smaller, the dynamic power is lower. In order to making the pre-charge voltage lower, we decided to use low-skewed INV[5] in SA. The theory is shown in Fig 7.
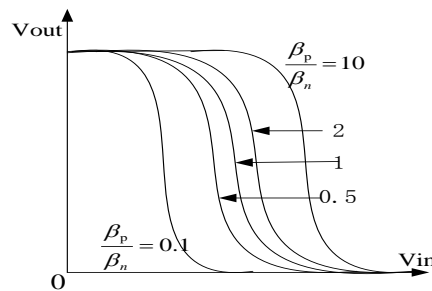


Fig 7    The VTC curve of the skewed INV

The high skewed inverter has a strong PMOS and a weak NMOS, so it has higher threshold voltage. The low skewed inverter has a weak PMOS and a strong NMOS, so this inverter has lower threshold voltage. According to the analysis to the influence of the parameter $\beta$ ( $\beta = W_{PMOS}/W_{NMOS}$) to the inverter VTC curve, we can see that the VTC is moving to the right with the increasing $\beta$, and the threshold voltage of the inverter is higher; when reducing the $\beta$ ratio, the VTC is moving to

the left with the reducing of β, and the threshold is lower. Although changing the value of β will making the threshold changing, the voltage transition is still steep.

In this design,the inverter of Sensitive Amplifier hope to get faster pull-down speed when read 0. According to adjust the the MOSFETs' size of the inverter, it can satisfy the demand of pull-down speed, which then becomes a low-skewed inverter. When read 1, the voltage of the X point is slightly higher than the trip-point of the INV, that makes SA output pull low, so that the output voltage of Y is 0. P3 is closed at this time, Z point is still in low voltage as in pre-charge. In fact, the sensitive amplifier speed is faster when reading 1 than when reading 0, because P3 is off when SA reading 0, So the speed of reading 1 is faster.

**Performance Comparison**

This paper presents a new single-port Sense amplifier SRAM, which compares with the traditional custom SRAM and the businessly memory compiler SRAM in performance and power consumption. Contrasting performance and power consumption in SS, TT, FF corner at the 40nm process with XA, the results is in the Table1 and Table2.

Table1 CLK-Q_delay

| CLK-Q_delay | Traditional_full_custom | compiler | SA |
|---|---|---|---|
| SS    corner | 0.609ns | 0.46ns | 0.421ns |
| TT    corner | 0.373ns | 0.283ns | 0.242ns |
| FF    corner | 0.246ns | 0.187ns | 0.152ns |

Table2 Dynamic_Power

| Dynamic_Power | Traditional_full_custom | compiler | SA |
|---|---|---|---|
| SS    corner | 3.8e-03 | 3.25e-03 | 1.64e-03 |
| TT    corner | 4.98e-03 | 3.90e-03 | 2.65e-03 |
| FF    corner | 6.55e-03 | 4.86e-03 | 4.13e-03 |

Comparative and Analysis: As can be seen from the performance and power consumption comparison graph, the structure of the new single SA presented in this paper shows that both performance and power consumption of the amplifier circuit are better than the other. This is due to the traditional full-custom read circuit is full swing, which makes the voltage of the bit line pull from 0 to VDD in the pre-charge stage when read 0; at the evaluation phase, discharged from VDD to 0, resulting spend much more time on the bit line in the precharge stage as the discharge stage, and the bit line will produce a great dynamic power, so the traditional full-custom SRAM has a large output delay, and also has a lot of dynamic power. Compared to the memory compiler SRAM, which uses the same internal sense amplifier, but it uses a 6T bit cell[6], and the reading circuit uses a latch-type sensitive amplifier, it is one of dual-port sensitive amplifier, the enable signal which control sense amplifier in the reading circuit, need to wait the virtual storage unit discharged off, and the bit line voltage produce a voltage difference. This structure needs to generate some additional delay, but the design raised in this paper is relatively simple to control sensitive amplifier. No voltage difference generation time, so it will be faster in timing, and in the pre-charge stage, the pre-charge voltage is much lower, which is about VDD / 2, so the charge and discharge will be faster than others, the bit line voltage swing is reduced, and the dynamic power consumption is greatly reduced, so the advantages of reducing power consumption are obvious.   In addition, 8T cell has higher reliability than 6T.

## Summary

In this thesis, a storage capacity of 512X10 SRAM circuit is designed. This circuit's design targets are higher reliability, higher speed and lower power consumption. Through the design and verification of this circuit, and separately compared with the conventional full-custom circuit and memory compiler SRAM in performance, the conclusion is that this new single SA structure is better than them in speed and power consumption, indicating that this custom circuit design has achieved satisfactory results, and this full-custom design is very successful. In addition, there are some good design methods mentioned in this design process，it is worth learning for other designers in the custom SRAM design.

## Acknowledgments

## References

[1] Jan-M.Rabaey, Anantha-Chandrakasan, "Digital Integrated Circuits: A Design  Perspective, second Edition,2004.10.

[2] Hanwool Jeong,Taewon Kim,Taejoong Song,Gyuhong Kim,and Seong-Ook jung，"Trip-Point Bit-Line Precharge Sensing Scheme for Single-Ended SRAM，" in IEEE TRANSACTIONS ON VLSI SYSTEMS

[3] N. Verma and A. P. Chandrakasan, "A high-density 45 nm SRAM using small-signal non-strobed regenerative sensing," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2008, pp. 380–621

[4] B. Giridhar, N. Pinckney, D. Sylvester, and D. Blaauw, "13.7 A reconfigurable sense amplifier with auto-zero calibration and pre-amplification in 28 nm CMOS," in IEEE Int. Solid-State Circuits Conf. Dig. Tech.Papers, Feb. 2014, pp. 242–243.

[5] Liang Wen. "The design and realize on High Performance SRAM in 65nm CMOS,"Master Thesis, Graduate School of NUDT, 2011

[6] S. Nalam, V. Chandra, C. Pietrzyk, R. C. Aitken, and B. H. Calhoun, "Asymmetric 6T SRAM with two-phase write and split bitline differential sensing for low voltage operation," in Proc. 11th Int. Symp. Qual. Electron. Des. (ISQED), Mar. 2010, pp. 139–146.