Robust and low power register file in 65 nm technology*

Zhang Xingxing(张星星)[1], Li Yi(李毅)[1], Xiong Baoyu(熊保玉)[1], Han Jun(韩军)[1], Zhang Yuejun(张跃军)[2], Dong Fangyuan(董方圆)[1], Zhang Zhang(张章)[3], Yu Zhiyi(虞志益)[1,†], Han Jun(韩军)[1], Cheng Xu(程旭)[1], and Zeng Xiaoyang(曾晓洋)[1]

[1]State Key Laboratory of ASIC and System, Fudan University, Shanghai 201203, China
[2]Institute of Circuits and Systems, Ningbo University, Ningbo 315211, China
[3]School of Electronic Science and Applied Physics, Hefei University of Technology, Hefei 230009, China

Abstract: A register file (RF) with 32 × 32 capacity and 4-read 2-write (4R2W) ports is presented and analyzed in detail. A new output structure using a MUX and a latch is proposed. It eliminates any dynamic or analog circuit in the read path, and thus it can improve robustness and reduce power at the same time. We also simplify the timing sequence due to the output scheme. The simplified timing circuit not only cuts down the power but also improves the robustness. In addition, less power is achieved when successive read of "0" or "1" is performed. The RF has been fabricated in TSMC 65 nm technology, and the chip test demonstrates that it can operate at 0.8 GHz, consuming 7.2 mW at 1.2 V.

Key words: register file; 65 nm; robust; low power; multi-port
DOI: 10.1088/1674-4926/33/3/035010 **EEACC:** 2570

1. Introduction

Parallel computing technologies such as out-of-order and multi-threading often require a multi-port RF to feed multiple execution units at the same time. In Motorola's M.CORE processor [1], the power consumed by the RF is 16% of total power, and 42% of the data path power. A study[2] shows that the RF is responsible for about 25% of total processor power consumption. So it is essential to design a multi-port and low power RF.

Traditional RFs, which normally use a sense amplifier in the output structure, always require high symmetry, and they are susceptible to noise and process-voltage-temperature (PVT) variation. The output structure using a pseudo static bit line technique proposed in Ref. [3] can lessen these problems. However, it still needs dynamic circuits: the pre-charge and the keeper circuit. Noise tolerance of wide dynamic gates degrades rapidly with technology scaling as transistor subthreshold leakage increases exponentially[4]. The keeper circuit also causes a short-cut current and increases the read delay, because of the competition between the read bit line and the keeper. Another dynamic circuit problem is that it must recover its state after each read operation, thus extra power will be consumed.

In this paper, we propose a new output structure that can eliminate the aforementioned problems. It is made up of a MUX and a latch, which are full-static circuits. Because dynamic or analog circuits are avoided in the structure, robustness and both low power are achieved. The simplified timing control module also improves the robustness and cuts down the power. With the co-operation of the cells, the RF will consume less power when successive read of "0" or "1" is performed.

2. Design details

The cell array is divided into four banks to obtain the low access time on the read path shown in Fig. 1. Signals entering the decoder and input module are latched to avoid unpredictable change when operations are performed.

2.1. Cell array

The traditional 6T cell structure accessed by two nMOSs is not suitable for multi-port read. When a single-read operation is performed, the differential voltage on the read local bit line (RLBL) may disturb the data in the cell. This impact may be aggravated when multi-port read is performed in one cell. To eliminate this problem, a robust cell[5], shown in Fig. 2, is described. Isolating inverters are added to isolate the coupled inverters and the RLBL. To measure the stability of our cell, the static noise margin (SNM)[6] is simulated. The comparison results are shown in Table 1 at the supply voltage of 1.2 V. It manifests that the traditional cell is almost not accepted, while our cell manifest is much more robust.

In traditional design, the RLBL is pulled up to VDD after each read operation is completed. However, because of the use of transmission gates in this cell, the RLBL can keep the corresponding state. Thus it will consume less power when successive read of "0" or "1" is performed[7]. The new output architecture proposed in this paper will further explore the non-recovery scheme, and reduce power in the best way.

2.2. Output module

Traditionally, a pre-charge circuit and keeper are needed to pull up and hold the RLBL (Fig. 3). When the read operation

* Project supported by the National Significant Science and Technology Projects (No. 01-Special-2010ZX01030-001-001-03).
† Corresponding author. Email: zhiyiyu@fudan.edu.cn
Received 8 September 2011, revised manuscript received 8 November 2011

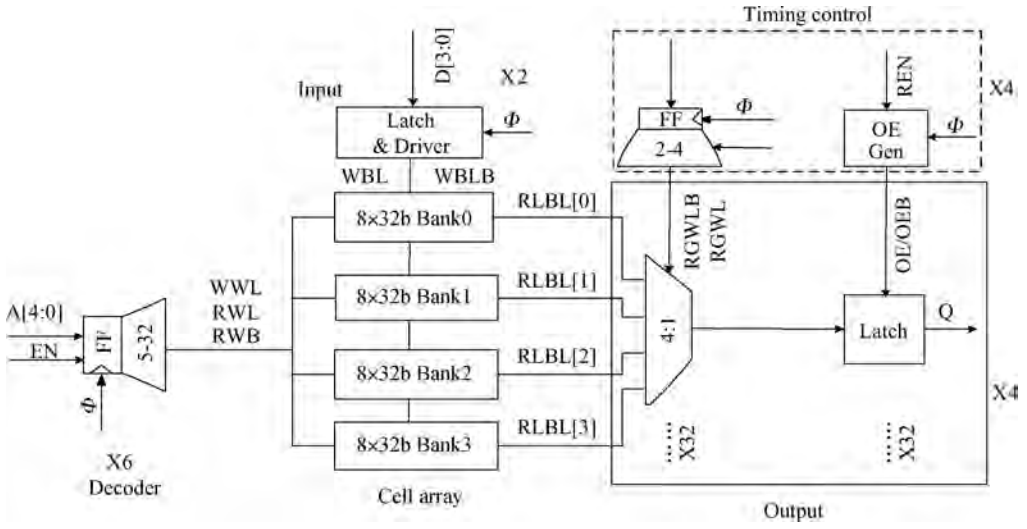

Fig. 1. The overall structure of RF.

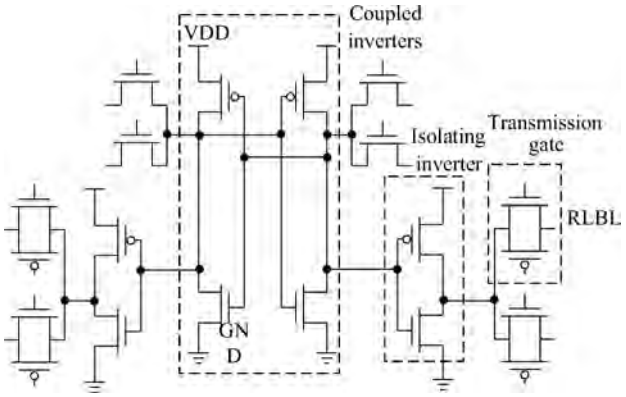

Fig. 2. Cell schematic.

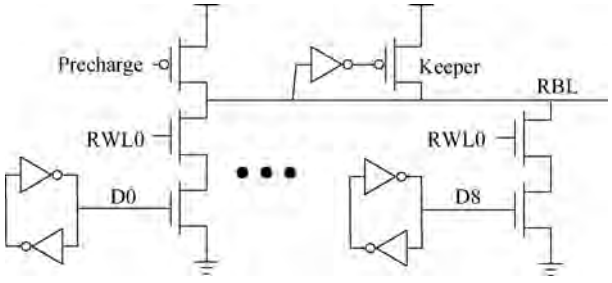

Fig. 3. Bank structure with pre-charge and keeper circuit.

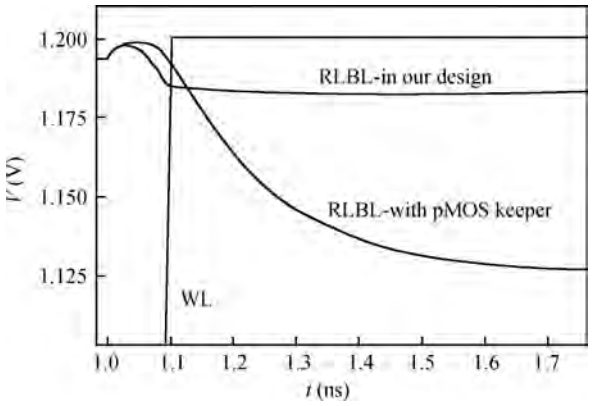

Fig. 4. RLBL DC droop in 65-nm at 1.2 V 125 ℃.

Table 1. Comparison of SNM.

MODE		SNM (mV)
Traditional cell	1 port read	210
	2 ports read	130
	3 ports read	88
	4 ports read	59
Our cell	1 port read	448
	2 ports read	448
	3 ports read	448
	4 ports read	448

occurs, the pre-charge circuit will turn off, and the RLBL will be driven down by the cell or kept high by the keeper. However, the RLBL is susceptible to noise due to the high active leakage during evaluation, when it should stay high. It is particularly more sensitive when the node charge is smaller and the dynamic structure is wider[8].

In an LVT process, when the RLBL is required to stay high, a 75 mV DC droop (6.2% of VDD) is observed (Fig. 4) in the RLBL in the worst case. A straightforward solution is upsizing the pMOS keeper, which can compensate for the leakage current. However, the competition between the RLBL and the keeper will increase the power and delay. Figure 5 shows the delay and power increase as the keeper is upsized from 10% to 100% of the effective nMOS pull-down strength. So a trade-off must be made between the robustness and other performance factors.

The new output architecture shown in Fig. 6 solves the problems. When the RWL is effective, the keeper circuit (coupled inverters in MUX) will be cut off. Thus, the short-cut current caused by the keeper is reduced. Compared with the dynamic structure (pMOS keeper is 10% of the effective pull-down strength), the power and delay reduce 8% and 39% respectively. Furthermore, the dc droop is only 15 mV (1.2% of VDD) as shown in Fig. 4. The output structure

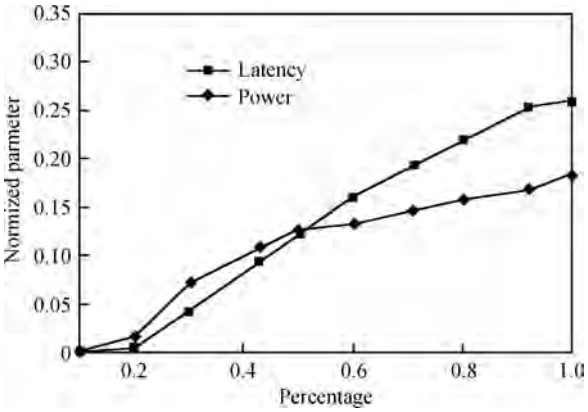

Fig. 5. Keeper upsizing versus delay and short-cut power in 65-nm at 1.2 V, 125 ℃.

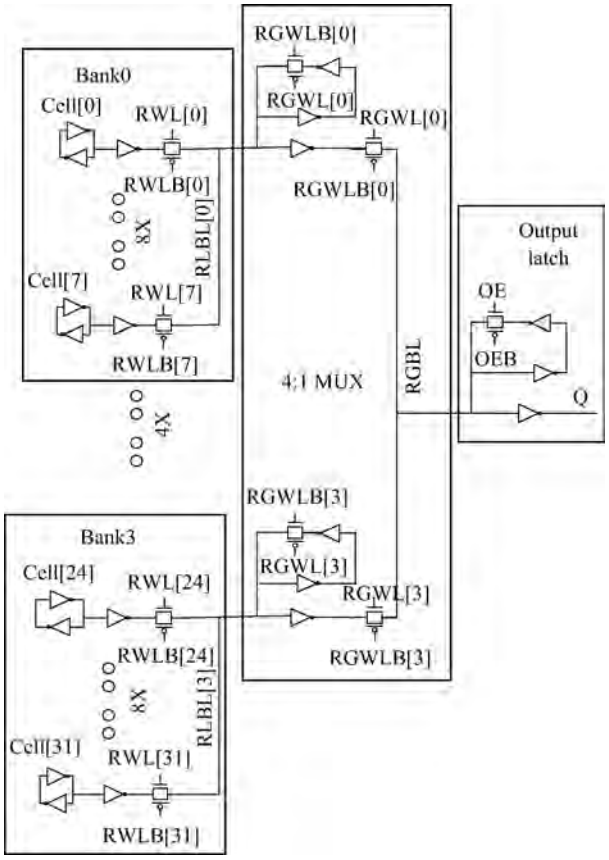

Fig. 6. Output module in our design.

performs as follows (we take Bank0 for example): (1) the RGWL[0]/RGWLB[0] turn off/on the corresponding transmission gates in the MUX; (2) the RLBL[0] is driven to the corresponding state (VDD or GND) by the cell; (3) RLBL[0] is chosen out of the 4:1 MUX, and the MUX result is read global bit line (RGBL); (4) when the RGBL is stable, the OE/OEB disable the transmission gate in the output latch and output the result. When the read operation is finished, the corresponding coupled inverters are connected again. Thus the RLBL is statically held again by the coupled inverters. If the next read result is the same (successive read of "0" or "1"), the RGBL does not need to change its state. So less power will be consumed. In sit-

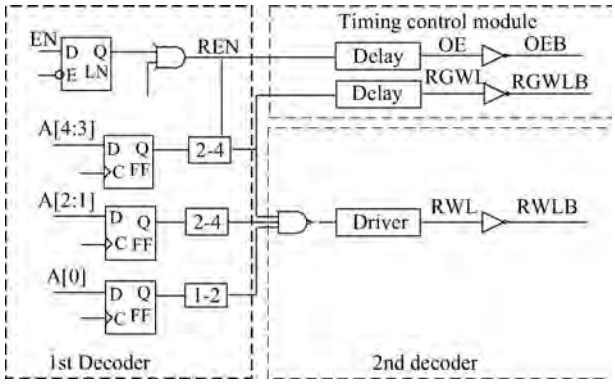

Fig. 7. Decoder and timing control Module.

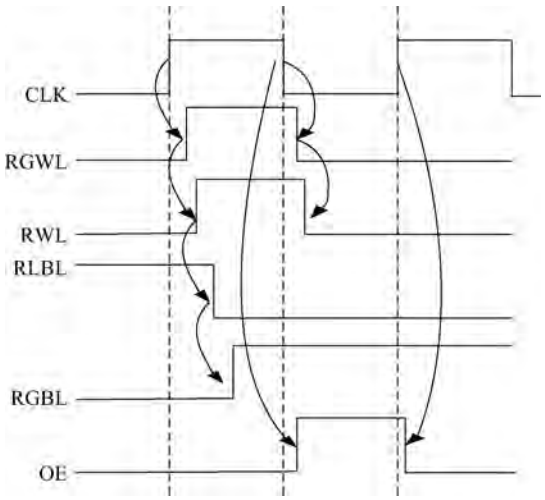

Fig. 8. Timing sequence of the internal signal.

uations where there is long successive data, the RF power can be reduced the most. We will analyze the impact of different kinds of data later. The area occupied by the output module is about 20% of the RF. In the condition of every bit switchover, the output module consumes about 32% of total power. So it is necessary to cut down the power consumed by the output module.

2.3. Decoder and timing control module

The timing sequence is simplified due to the new output architecture. The two-stage static decoder and timing control module shown in Fig. 7 is used. The read enable (REN) is generated through the NAND gates controlled by the CLK and the EN. The first stage decoder and the timing control module share the REN signal on order to cut down on power. The simulation shows that the timing control module consumes about 13% of total power. In the traditional design, the word line pulse is required to be appropriate. A short word line pulse will reduce the discharge time on the RLBL, resulting in uncertainty of the result. In the design using a sense amplifier, the long word line pulse will discharge the RLBL more, inducing more power to be consumed. In addition, the time left to the precharge circuit is shortened, which means that the RLBL may not be charged to VDD before the next read operation. However, in our design, it is no longer a consideration. As shown in

Table 2. Comparison of the RFs with previously published designs.

Paper	Capacity $N_{words} \times N_{bits}$	Port	Process (nm)	Power (mW)	Frequency (GHZ)	Supply (V)	P_{norm} (mW/GHz)
ISSCC'06[9]	16 × 64	1R1W	65	198	8.8	1.2	0.1758
ESSCC'07[10]	48 × 32	1R1W	65	47	6.3	1.2	0.1156
ISSCC'10[11]	64 × 32	1R1W	32	72	7.5	1.0	0.15
ISSCC'11[12]	144 × 78	4R2W	45	59	2.3	0.9	0.0547
Ours	32 × 32	4R2W	65	7.2	0.8	1.2	0.0469

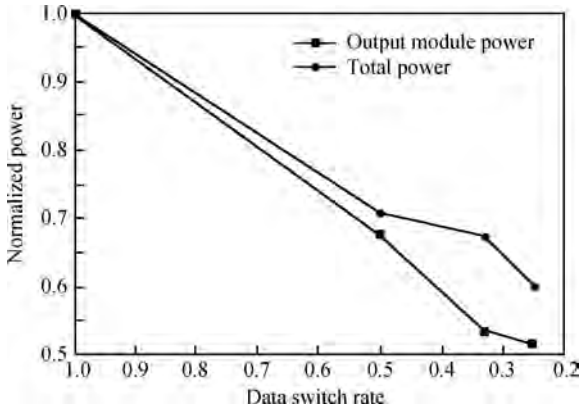

Fig. 9. Output and total power versus data switch rate.

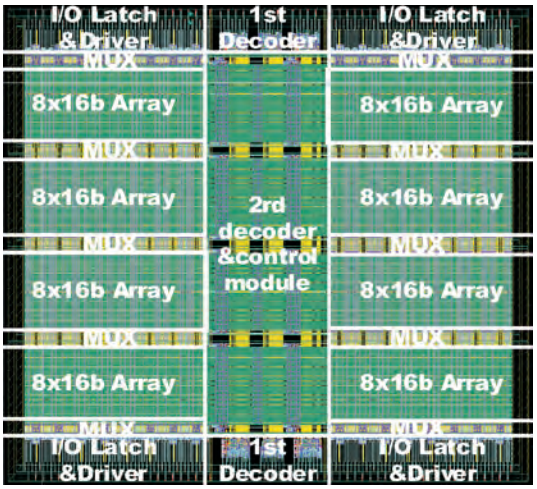

Fig. 10. Layout of the RF.

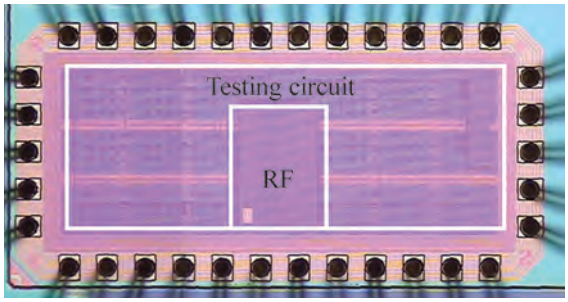

Fig. 11. Chip micro photo of the RF.

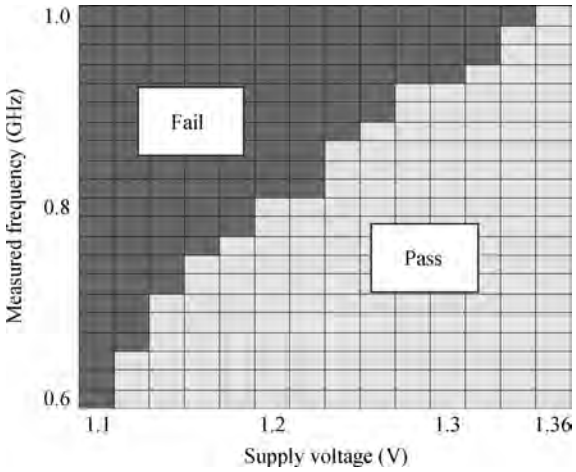

Fig. 12. Frequency characteristics of RF versus supply voltage.

Fig. 8, the width of the word line pulse is decided by the CLK pulse. So long as the clock pulse is not too short, the word line pulse will be long enough to ensure the function of the RF. The long word line pulse will have no influence on the output, since RLBL and RGBL do not need to recover to VDD.

3. Analysis and testing result

To obtain the impact of successive data on the RF, we change the data's switch rate. When the switch rate varies from 1 to 0.25 per cycle, the power decreased dramatically, as shown in Fig. 9. It also manifests that the power consumed by the output module decreases more rapidly than the total power, which means that the output module is energy efficient in the RF. The RF is especially suited to the audio or video equipment, because of the frequent successive data type.

The layout of the RF is shown in Fig. 10; each module is scaled out in detail. Decoders are placed in the center to balance the word line latency. The cell array occupies almost 50% of the total area. MUXes are designed to be near the cell bank to short the wire length of the RLBL. The I/O latches and decoder addresses are distributed equally in the top and bottom sides to weaken congestion. The RF has been fabricated in the TSMC 65 nm low power (LP) process, the chip photo of the RF is shown in Fig. 11. The testing circuit is specially designed to measure the performance of RF. The RF is supplied independently to get its accurate power. Figure 12 shows the measured frequency versus supply voltage. The frequency increases linearly with the growth of voltage. At the standard voltage of 1.2 V, it can work well at 0.8 GHz. The chip test manifests that it consumes 7.2 mW at 1.2 V in the situation of each bit

Table 3. Summary of the measured result.

Parameter	Value
Process	TSMC 65 nm CMOS LP 1p9m
Organization	32×32 bit
Single cell size	$5.4\ \mu\text{m} \times 3.8\ \mu\text{m} = 20.52\ \mu\text{m}^2$
Total area of RF	$0.19\ \text{mm} \times 0.25\ \text{mm} = 0.046\ \text{mm}^2$
Supply voltage	1.2 V
Frequency	800 MHz
Power	7.2 mW
Leakage power	18 μW (simulated in TT corner, 1.2 V, 27 ℃)

changes every other cycle.

$$P_{\text{norm}} = \frac{P}{f(N_{\text{R}} + N_{\text{W}})N_{\text{bits}}}. \quad (1)$$

The power of the RF is mainly determined by the number of ports and the word width. To compare with other previously published RFs, Equation (1) is used to calculate the normalized power. The comparison results (Table 2) manifests that our RF is the most energy efficient. Due to the precision of the power meter, the leakage power is obtained through the back-annotated simulation. Detailed results are shown in Table 3.

4. Conclusion

In this paper, we describe and give a detail analysis of the 4R2W RF. The new output architecture supports robust read access. The successive read of "0" or "1" will consume less power. Furthermore, the simple timing control scheme cuts down power in the best way.

The chip test demonstrates that it can operate up to 800 MHz at 1.2 V supply voltage with 7.2 mW total power. Because of its low power characteristic in the successive data, the RF is suitable to be used in audio or video systems. Also it can be used in military or aerospace domains due to its robustness.

References

[1] Scott J. Designing the lower power M.core architecture. IEEE Power Driven Microarchitecture Workshop, 1998: 145

[2] Zyuban V, Kogge P. The energy complexity of register files. ISLPED, 1998

[3] Sriram V, Mark A A, Nitin B, et al. 5-GHz 32-bit integer execution core in 130-nm dual-V_t CMOS. IEEE J Solid-State Circuits, 2002, 37(11): 1421

[4] Anders M, Krishnamurthy R, Spotten R, et al. Robustness of sub-70 nm dynamic circuits: analytical techniques and scaling trends. Symposium on VLSI Circuits, 2001: 23

[5] Xiong Baoyu, Zhang Xingxing, Han Jun, et al. Design of a single-ended cell based 65nm 32×32b 4R2W register file. International Conference on ASIC, 2011: 339

[6] List F J. The static noise margin of SRAM cells. European Solid-State Circuits Conference, 1986: 16

[7] Hiroki H, Okumura S, Iguchi Y, et al. Which is the best dual-port SRAM in 45-nm process technology? –8T, 10T single end, and 10T differential. Integrated Circuit Design and Technology and Tutorial, 2008: 55

[8] Ran K K, Atila A, Ganesh B, et al. A 130-nm 6-GHz 256×32 bit leakage-tolerant register file. IEEE J Solid-State Circuits, 2002, 37(5): 624

[9] Hsu S, Agarwal A, Anders M, et al. An 8.8 GHz 198 mW 16 $\times$ 64b 1R/1W variation tolerant register file in 65 nm CMOS. International Solid-State Circuit Conference, 2006: 1785

[10] Agarwal A, Banerjee N, Hsu S K, et al. A 200 mV to 1.2 V, 4.4 MHz to 6.3 GHz, 482b 1R/1W programmable register file in 65 nm CMOS. European Solid State Circuits Conference, 2007: 316

[11] Agarwal A, Mathew S K, Hsu S K, et al. A 320 mV-to-1.2 V on-die fine-grained reconfigurable fabric for DSP/media accelerators in 32 nm CMOS. International Solid-State Circuit Conference, 2010: 328

[12] Ditlow G S, Montoye R K, Storino S N, et al. A 4R2W register file for a 2.3 GHz wire-speed POWER processor with double-pumped write operation. International Solid State Circuits Conference, 2011: 256