A 110MHz / 1Mbit Synchronous Tag RAM

Yasuhiro Ueda, Tsugue Kobayashi, Tsukasa Shiratori, Yukihiro Fujimoto, Takayoshi Shimazawa, Kazuhiro Nogami, Takehiko Nakao, Kazuhiro Sawada, Masataka Matsui, Takayasu Sakurai, Man Kit Tang**, and Bill Huffman**

1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan
Silicon Opciones Inc.**
110 North Shoreline Blvd. Mountain View, CA, 94039, USA

1. INTRODUCTION

Most of the recent micro-processors include the first level cache on chip. However, a chip size constraint limits the on-chip cache capacity to 16KBytes. Therefore a large off-chip secondary cache is indispensable for high-performance computer systems. The synchronous Tag RAM reported in this paper addresses holds addresses and status bits of cached data and can be used to build a secondary cache system of up to 16MBytes with external commodity synchronous SRAMS. In order to handle the large secondary cache the between Tag RAM contains 1.89Mbit of 4T SRAM cells, the largest capacity ever reported for a Tag RAM.

Short cycle time and small clock to DOUT (data output) delay of the Tag RAM is crucial for a high-performance cache system. 9ns cycle operation and clock to DOUT of 4.7ns in typical condition are achieved by a use of circuit techniques such as a pipelined decoding scheme, a single PMOS load BICMOS main decoder, a BICMOS sense-amplifying comparator, a highly linear Voltage-Controlled Oscillator (VCO) for a Phase Locked Loop (PLL) and doubly placed self-timed write circuits. Since pure CMOS implementation can not achieve the required speed, the device is manufactured with 0.7um double-polySilicon and double-metal BICMOS technology.

2. FEATURES

Figure 1 shows a block diagram of the present Tag RAM. It contains 8K entries x 4-ways x 20bits Tag memory, 8K entries x 4-ways x 12 bits memory for S/A bits and 8K entries x 4-way x 4 bits of Dirty bits. 8 redundancy rows are included. It also contains comparators to compare read-out Tag address with Higher Physical Address. Double WL structure is adopted for reducing memory cell power consumption and WL delay. In the State bits, Virtual Synonym and a Dirty memory WL structure and in consequence the P1 can be designed to have high drivability to realize high-speed pull-up. The present circuit reduces the address decoding time by 0.5ns compared with the conventional full CMOS decoder + BICMOS buffer scheme.

3. SENSE-AMPLIFYING BICMOS COMPARATOR

Conventionally, a comparator for a cache is built with a MOS comparator inserted between a bit line (BL) and a BICMOS sense amplifier (SAC), whose circuit diagram is shown in Fig.3, replaces the conventional MOS comparator with a bipolar comparator (Q3 & Q6) merged into a bipolar sense amplifier (Q1, Q2, Q4, Q5). This configuration eliminates the MOS comparator delay and gains 0.1ns.

Because the Tag data read-out is also required a fast operation in the present Tag RAM, a Tag data sense amplifier (S/A, Q6 - Q10) is placed in parallel to the SAC. All circuits from the BL through the HRT signal generator take ECL-based configuration to reduce critical path delay.

3.4 ON-CHIP PLL

A PLL is integrated on chip which cancels internal clock delay. The linearity of the VCO is a key to obtain a large lock frequency range. The proposed VCO is shown in Fig.4, together with a linearity comparison with the conventional VCO. Due to the high linearity, the PLL is measured to lock frequencies from 50MHz to 150MHz stably with a 0.4ns jitters. The inclusion of the PLL on a chip can reduce the cycle time by 1ns.

3.5 DOUBLY PLACED SELF-TIMED WRITE CIRCUITS

In order to minimize the clock to DOUT delay, sense amplifiers should be placed near to the pads. This rules out the possibility of BL precharge. In the presented Tag RAM, the WL delay is reduced by 1ns and the write operation delay by 1ns. This reduces the cycle time by 2ns. In the case where the write operation determines the cycle time, the write operation becomes the slowest path.

4. RESULTS

Figure 6 shows a total chip layout whose size is 14.8mm x 14.8mm. The minimum clock cycle time is 9ns in typical condition which corresponds to 110MHz clock frequency. If the RAM is designed with a pure CMOS technology without the circuit ideas mentioned in Section 3, the clock cycle is estimated to be 16ns. If the RAM is designed with a BICMOS technology without the above-mentioned circuit ideas, the clock cycle is estimated to be 13ns. The 4ns improvement of the present design over the conventional BICMOS design comes from the additive speed gains of the circuit ideas described in chapter 3.1 through 3.4.

Figure 9 shows a simulation of the delay distribution of a Tag look-up cycle operated at 9ns cycle time, respectively.

REFERENCES

Figure 1 Memory core architecture for Tag look-up operation

Table 1 Features

Figure 2 BICMOS Main Decoder

Figure 3 Sense Amplifying Comparator

Figure 4 Voltage Controlled Oscillator and measured property

Figure 5 Doubly placed precharge and write control

Figure 6 Chip layout

Figure 7 Simulated waveforms of a Tag look-up cycle

Figure 8 Distribution of delay time in Tag look-up cycle