INTEGRATION, the VLSI journal 55 (2016) 12–21

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

journal homepage: www.elsevier.com/locate/vlsi

A compact low-power 4-port register file with grounded write bitlines and single-ended read operations

Khawar Sarfraz *, Mansun Chan

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 7 October 2015
Received in revised form
29 December 2015
Accepted 21 February 2016
Available online 2 March 2016

Keywords:
Multi-port
Mobile application processor
Clock management
Data retention
Noise margins

ABSTRACT

An area-efficient 4-port register file with low power consumption is presented for mobile application processors. Area efficiency at array level is achieved with a novel compact bitcell that supports single-ended one-sided read operations using the direct read access mechanism and single-ended write operations. A write-assist technique ensures robust operation down to 0.75 V. Single-ended one-sided read operations help maintain sufficient bitcell stability at 0.75 V. Factors that contribute toward low power consumption include grounded write bitlines, bitcells with low leakage currents, individual read clock generators for top and bottom halves of the array, and smaller wordline buffers and capacitance due to a smaller bitcell. For a 2-Kib array implemented in TSMC 65 nm low power (LP) dual-V_t CMOS process, the proposed design achieves 17.8% reduction in silicon area, 19.6% lower active power, and 12.8% lower standby power when compared to the conventional 4-port dual-V_t register file. These benefits are obtained by trading off operating frequency at voltages below the nominal, read and write bitcell noise margins, and data retention voltage.

1. Introduction

Multi-ported register file (RF) memories are a cornerstone of modern superscalar microprocessors, and typically operate at the same speed as the processor core. The area of a multi-ported RF grows more than quadratically with issue width [1] and can exceed that of level-1 cache [2]. RFs consume up to 37.1% of the active power of embedded platforms [3] and up to 15% of the total system-on-chip leakage power [4]. Low power RFs with a small silicon footprint are therefore highly desirable for 32-bit and 64-bit mobile application processors that are tailored for low-cost, high performance for low power, high reliability, and high flexibility. Application processors are extensively employed in smart phones, gaming consoles, navigation devices, automotive infotainment systems, tablets, and e-readers. Processors and SoCs that target these applications have been reported with operating frequencies within a wide band from 500-MHz to about 2-GHz. Dynamic voltage scaling (DVS) is demonstrated by these processors down to 0.7 V [5–11] in order to achieve energy efficiency.

The conventional eight-transistor (8T) static random access memory (SRAM) bitcell [12] is typically port-extended to realize a multi-port RF bitcell, where port extension implies increasing the number of differential write ports and the single-ended read ports to achieve dedicated multi-port functionality. Compared to the conventional 6T SRAM bitcell that is shown in Fig. 1a, the 8T bitcell that is shown in Fig. 1b demonstrates enhanced data stability characteristics with low voltage operation capability primarily due to the isolated read port and the fact that read and write port transistors can be sized independently to satisfy a given performance constraint. An 8T SRAM is demonstrated to operate down to 0.41 V [13]. The 8T SRAM bitcell is therefore over-dimensioned for mobile application processors that operate over a moderate voltage range, typically down to 0.7 V only. Since bitcell area is known to increase quadratically with the number of ports [14], the use of a port-extended version of the 8T SRAM bitcell in multi-ported RFs results in large area figures and high active power consumption [14–18].

Several techniques have been previously reported to reduce active power and area in multi-port memories. A self reverse-bias bitline scheme [19] uses the conventional compact 6T SRAM bitcell with the two access transistors driven by two separate wordlines in order to achieve dual-port functionality. Even with wordline splitting, the bitcell fails to support 2 reads and 1 write in a single clock cycle due to the shared bitlines. Furthermore, the benefit that would have resulted from a compact bitcell layout is overshadowed by a larger bitcell beta ratio that is required to support the pre-discharged bitline architecture. The pre-discharged bitlines, which are held at 220 mV in standby, reduce the bitline leakage currents but result in excessive static power dissipation in the inverters that are placed at the output of dynamic bitlines. Finally, the design does

* Corresponding author.
E-mail addresses: ksarfraz@ust.hk (K. Sarfraz), mchan@ust.hk (M. Chan).

http://dx.doi.org/10.1016/j.vlsi.2016.02.008
0167-9260/

not support DVS, which is a crucial requirement in modern processors in order to achieve energy efficiency.

The dual-port memory in [20], which relies on the bitcell in [19], shares the port constraints associated with that bitcell. In addition, the design in [20] uses the dual supply voltage (dual-V_{DD}) architecture to achieve reduction in bitline switching power. The inclusion of an additional power rail adds significantly to the cost of the chip regardless of whether it is in terms of pin count or on-chip voltage generation. Furthermore, the differential read topology limits the minimum operating voltage of the memory due to read static noise margin (RSNM) constraints.

Double-pumped writes with replicated read ports is proposed in [14]. The RF design is based on the use of the conventional 8T SRAM bitcell. The basic idea is to split a highly-ported bitcell into a bitcell with lesser number of ports. The work in [14] claims that the design of two separate sub-arrays with the smaller bitcell would result in area and power efficiency. This approach however pushes the leakage power figure of the design beyond the active power of the memory (31 mW as compared to 28 mW [14]). Moreover, the 2 write operations (one to each sub-array) are performed on the rising and falling edge of the clock signal within the same clock period. This tight timing constraint significantly increases the complexity of timing circuits considering the requirement for wide voltage range operation.

A bitcell copying technique that yields area-efficient multi-port bitcells for RFs in superscalar microprocessors is proposed in [21]. The bitcell area is reduced to 63% of the original design. This reduction in bitcell area is offset by the inclusion of multiple control signal generators that need to be designed on a per-word basis. The simplicity achieved within the array is offset by a timing-critical and large control signal generator. Secondly, the bitcell is designed with symmetrical inverters that are written to using single-ended write operations conducted via transmission gates. This architecture is prone to write '1' failures in the presence of process parameter variations. Finally, the design does not support DVS thus preventing energy efficiency to be achieved in the processor core.

Low-power peripheral circuits used in conjunction with the conventional 8T SRAM bitcells [13] are presented in [15]. The design uses gated read wordlines (RWLs) and tri-stated floating read bitlines. The floating bitline architecture consumes less power but, together with the gated wordline architecture, results in a larger area penalty compared to conventional implementation.

In this paper, an area-efficient 2-read, 2-write ported dual-V_t RF with low power consumption is presented for mobile application processors that operate from 1.2 V down to 0.75 V. A 2-read, 2-write ported architecture meets the industrial standard for RF intellectual-property cores that are applicable to a wide audience in the networking and communications domain [22,23]. The proposed RF is designed with a single supply voltage using a novel bitcell topology that employs the minimum number of transistors required to facilitate the simultaneous and independent execution of 2 read and 2 write operations over the specified operating voltage range. The bitcell is designed to support single-ended one-sided reads using the direct read access mechanism in order to achieve adequate RSNM at low supply voltages. A previously-published write assist scheme [26] is employed to ensure successful completion of write '1' operations, particularly at low supply voltages. In this work, write bitlines (WBLs) are grounded in the absence of writes in order to minimize bitcell leakage currents. Furthermore, individual local bitline (LBL) and global bitline (GBL) clocks are designed for the two halves of the array in order to save dynamic power. The rest of this paper is organized in the following manner. Bitcell and peripheral circuit design is discussed in Section 2. Simulation results are analyzed and explained in Section 3, together with applicability of the proposed bitcell to highly-ported RFs. Finally, conclusions are drawn in Section 4.

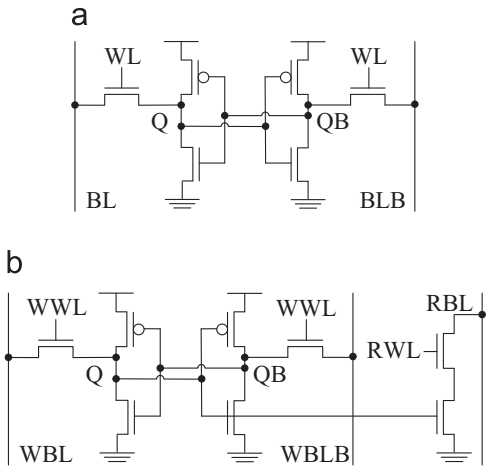

Fig. 1. (a) The conventional 6T SRAM bitcell. (b) The conventional 8T SRAM bitcell [12].

2. Bitcell and peripheral circuit design

2.1. The conventional register file

The conventional dual-V_t RF is designed with low-V_t transistors used on a port-enhanced version of the conventional 8T bitcell in [13]. Higher threshold voltage devices are employed only on leakage-sensitive nodes, specifically the read select transistor on the read stack (M7 in Fig. 2a) and the GBL pull down transistor on the clocked GBL pull down stack (M9 in Fig. 2b) to suppress leakage currents and to improve robustness of the dynamic LBL and GBL [19,24,25]. The use of minimum-sized WBL access transistors (M1 and M6) and latch transistors (M2–M5) ensures zero write bit errors under worst-case write conditions, owing to the differential nature of write operations. A minimum-sized read select transistor (M7) however results in a non-zero read bit error rate under worst-case read-after-write conditions. Read port transistors (M7 and M8) are therefore optimally sized in a progressive manner to satisfy (i)

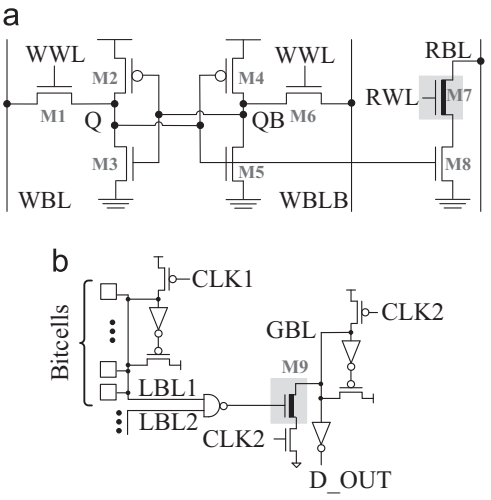

Fig. 2. (a) The conventional dual-V_t 1-read, 1-write-ported RF bitcell. W_{M1-M6}= 120 nm, W_{M7}=140 nm, W_{M8}=190 nm, $L_{M1-M8}=L_{MIN}$=60 nm. (b) Interface of the dynamic local and global bitline on a typical dynamic read out path. A thin line in the channel area indicates low-V_t transistor. A thick line in the channel area indicates standard-V_t transistor.