A 4.75GOPS Single-chip Programmable Processor Array Consisting of a Multithreaded Processor and Multiple SIMD and IO Processors

Young-Don Bae and In-Cheol Park

Division of EE, Department of EECS, KAIST

Daejeon, Republic of Korea

Abstract

This paper describes a configurable platform chip integrating 9 heterogeneous processors, which is designed to enable rapid prototyping and verification without translating functional behaviors into hardware blocks. The chip consists of a 32-bit multithreaded RISC processor for fast context switching, four 32-bit SIMD processors for data-intensive applications, and four IO processors for handling I/O protocols. The prototype chip has been designed and fabricated in 0.25-μm CMOS technology and the die size is 8x8mm^2 including eighty-kilobyte internal memories.

Introduction

As the complexity of SoC (system-on-a-chip) grows, many system-level programmable chips are being announced to support short time-to-market and reconfigurability. To meet the increasing demands for high performance, many of the recent configurable devices integrate a high-performance microprocessor as well as programmable logic (1-3). Functional units comprising a SoC are designed and verified in a high-level language in the early design stage, but they should be translated into logic gates to use such a device. A desirable way for achieving easy prototyping and verification is to execute all the functions on a microprocessor (1). As a number of components are integrated on a SoC, however, the whole functionality is too complex to be executed on a microprocessor (4-6). Moreover, the input and output functions associated with complex protocols usually require much higher clock frequency in emulating those IPs on a microprocessor (7). One of promising solutions is to integrate multiple processors on a chip. Such a multiprocessor architecture is recently employed for high-end embedded systems to deliver very high MIPS performance and to process a number of simultaneous tasks.

In order to process complex system functionality associated with IO peripherals, the processor should handle concurrent IO events and data as well as system controls. Therefore, fast context switching is essential in the processor to process the concurrent tasks (1). As the context switching is slow in conventional embedded processors, often requiring several hundreds of clock cycles, another architecture such as multithreaded processors is required to allow rapid processing of concurrent events. In addition, to provide sufficient data processing capability for multimedia processing, it is necessary to exploit sub-word level parallelism by adopting SIMD instructions (8). Since the overhead required for manipulating input and output data is significant, we have to reduce the overhead by providing a hardware unit that is specially designed for the IO accesses.

In this paper, we present a configurable device, called single chip programmable processor array (SPPA), that consists of multiple heterogeneous processors. The functional units of a SoC device are mapped to the multiple processors by programming their behaviors on the processors, which enables rapid prototyping and verification without the need of translating the behaviors into hardware blocks. SPPA includes a multithreaded processor, four SIMD processors, and four IO processors to provide rapid task control, high-performance data-processing capability and programmable I/Os, respectively.

System Architecture

SPPA is designed to implement conventional system-on-chip applications by programming the functions on the multiple processors integrated on SPPA. As a typical system-on-chip consists of a main processor, multimedia processing accelerators and I/O peripherals, SPPA supports three kinds of processors to implement each element efficiently. First, a multithreaded processor (MT-RISC) provides rapid-context switching capability required for executing multiple tasks concurrently. Second, SIMD processors are supported for efficient software implementation of high-performance media applications such as audio and video encoders. Third, IO processors that are optimally designed for the generation and translation of input and output signals are integrated.

Fig. 1 shows the overall structure of SPPA containing MT-RISC with 4kB instruction and data caches, a DMA controller and a Task Management Unit, four 32bit RISC processors with SIMD instructions (SIMD processors) and 8kB program/data memories, four IO processors with 4kB program/data memories and 2kB buffers, a 128-bit on-chip bus, and a 32kB on-chip SRAM.

A. 32-bit Multithreaded Processor (MT-RISC)

MT-RISC can process instructions obtained from five different threads in its 5-stage pipeline simultaneously. Each stage can work for an instruction issued from a different thread. While typical RISC processors require more than 100

0-7803-8495-4/04/$20.00 ©2004 IEEE.

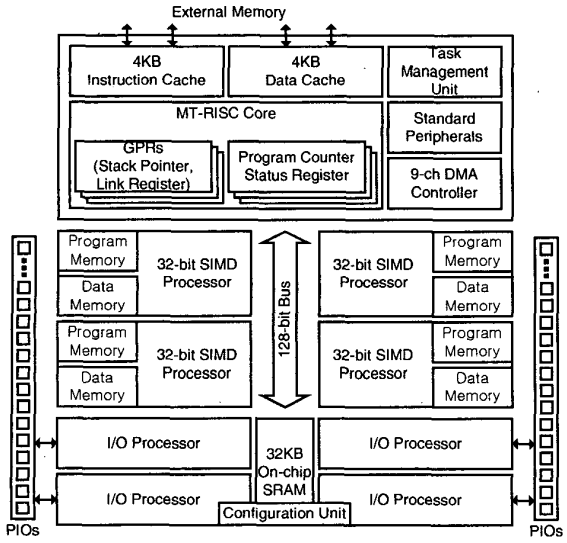

Fig. 1. Block Diagram of the SPPA.

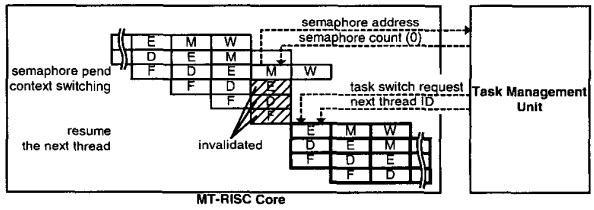

Fig. 2. MT-RISC Pipeline For Context Switching.

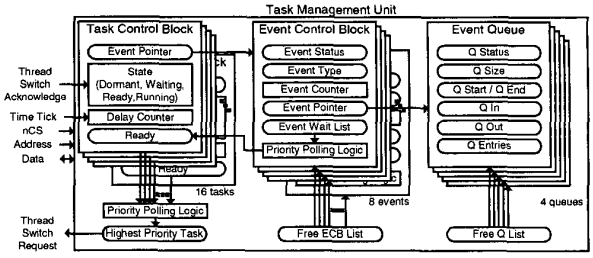

Fig. 3. Structure of the Task Management Unit.

clock cycles to perform a context switching, even without counting the scheduling overhead (9), MT-RISC requires zero clock cycle when a task tries to delay for a given number of clock ticks, and one clock cycle when a task is blocked during the synchronization process as shown in Fig. 2. The total processing power available is 275MIPS at 250MHz clock frequency. MT-RISC architecture supports up to 16 threads each of which is associated with a complete register set to store its thread state. Each register set has sixteen 32-bit general-purpose registers including stack pointer, link register and program counter, and status registers.

B. Task Management Unit

To support task control and inter-task synchronization, the Task Management Unit is integrated as shown in Fig. 3, which contains task control blocks (TCBs) to store the information of sixteen tasks, and event control blocks (ECBs) to store the information of eight events. Each event can be one of three types: semaphore, message mailbox and message queue. Through the profiling of several real-time operating systems (RTOSs), five special instructions are included in the ISA of MT-RISC for critical operations. Table I describes the detailed behavior of the instructions for multithreading. They significantly improve the response time by reducing the clock cycles required for the critical section of RTOS, provide atomicity needed to access shared resources exclusively, and save the code space. With the assistance of the Task Management Unit, MT-RISC always executes the highest priority task that is ready to run. The priority polling logic is used in the Task Management Unit to find the highest priority task among the tasks ready in TCBs and the tasks waiting for events in ECBs. As the priority polling logic and the timers in TCBs operate parallely, the number of clock cycles consumed for the scheduling is reduced from about 500 cycles to one cycle, compared to the software implementation.

TABLE I
INSTRUCTIONS FOR MULTITHREADING.

Instruction mnemonic	Description	Function
PEND *kind, register, message id*	Waiting for a message	Checks the Event Control Block, calls scheduler, and switches the context
POST *kind, register, message id*	Sending a message	
ACCEPT *kind, register, message id*	Get a message without waiting	
DELAY *task id, clock ticks*	Delay a task for a given clock ticks	Changes the Task Control Block, calls scheduler, and switches the context
RESUME *task id*	Resume a delayed task	

C. 32-bit SIMD Processors

Although MT-RISC is efficient for processing concurrent control flow, it is not fast enough to handle the function of today's complex SoCs. Multiple 32-bit SIMD processors are integrated for data-intensive functions such as audio and video encoding. For a short data type, a major portion of a 32-bit datapath computes sign-extension bits. Single Instruction Multiple Data (SIMD) is an efficient architectural model if the application has data regularity. The SIMD processor is based on the same basic instructions as MT-RISC, and provides nineteen SIMD instructions additionally. As described in Table II, arithmetic instructions, multiply, pack/unpack and permutation instructions manipulate four 8-bit or two 16-bit operations in the 32-bit datapath. Experimental results on several DSP algorithms show that a SIMD processor has more than two times higher performance than a conventional single-MAC DSP. The SIMD processor operating at 250MHz clock frequency provides 1GOPS (peak) performance.

TABLE II
SIMD INSTRUCTION SET

Category	Function	# of instructions
SIMD arithmetic	16-bit addition and subtraction 8-bit addition and subtraction selection	8
Fractional multiply	Dual 16-bit multiply Four 8-bit multiply	8
Misc.	pack/unpack permutation	3
Total		19 instructions

D. IO processors

The IO processor translates a sequence of IO signals into a regular data stream that the microprocessor can read easily, and generates IO signals from the data stream written by MT-RISC. Although conventional RISC processors are designed to process a few data types aligned on word boundary, I/O data streams such as network packets are usually not aligned to word size in order to reduce the total size of data to be transferred, meaning that a data field is not always aligned to word boundary and its size varies from a few bits to thousands of bytes. Before processing a data field that is not aligned, a sequence of manipulation operations such as shifting, masking and concatenating must be applied to convert the data field to a word-sized one, increasing the clock frequency to several times of the data rate. As shown in Fig. 4, each IO processor is associated with an Intelligent Buffer that does the data alignment as well as works as a buffer. The processor can access an arbitrary-sized field from the Intelligent Buffer by specifying its size ranging from 1 bit to sixteen bits in a load or store instruction. Every field read from the Intelligent Buffer is automatically aligned to the word size, which saves clock cycles and reduces code size too. Each IO processor operating at 125MHz is associated with 16 programmable input/output pads (PIOs) that can be configured either input or output pins. According to the configuration specified by the number of pins, direction and data rate, the Intelligent Buffer reads data from the PIOs or writes data to the PIOs.

E. On-chip Communication and Memories

The 32kB on-chip SRAM can be used as a scratch-pad memory for MT-RISC and SIMD processors or as message queues to transfer commands or information between processors. As shown in Fig. 3, the event queue in the Task Management Unit contains start and end pointers of the message storage area where the next message will be inserted and extracted. The message queue is a circular buffer that can accommodate 2 to 8 thousand entries depending on the size of messages ranging from 8 to 32 bits. There are 8 ECBs, and four of them support mutual exclusion to have four message queues shared by several tasks. In addition, the 9-channel DMA controller and the 128bit on-chip bus are included to deliver 32Gbps bandwidth by enabling four independent data transmissions between on-chip SRAM and the local memories of SIMD and IO processors.

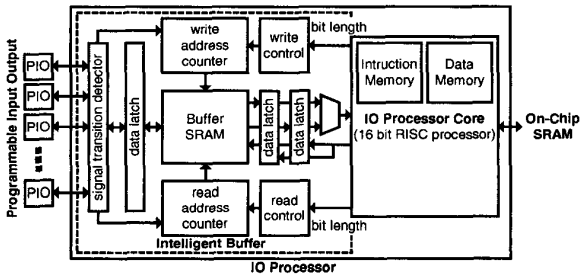

Fig. 4. IO Processor Block Diagram.

Performance Evaluation

Recently, a number of multithreaded processors are announced for embedded real-time applications (1, 9-10). Compared to the state-of-the-art embedded multithreaded processors, MT-RISC provides the fastest context switching capability and supports fine-grained multithreading to execute different threads on a cycle basis. In addition, a task management unit is integrated to assist a real-time OS.

Fig. 5 shows the performance enhancement resulting from the SIMD instructions employed in the SIMD processors. The experimental results over several DSP algorithms indicate that the proposed SIMD processor outperforms the conventional single-MAC DSPs.

TABLE III
INTERFACE IPS ON A COMMERCIAL FPGA.

IP	# of Logic Elements	Usage (Altera EP20K200(11))	Estimated area (0.18μm process)
USB Controller	2149	39%	48.7 mm^2
SDRAM Controller	1100	20%	24.9 mm^2
UART	660	12%	14.9 mm^2
I2C Bus Controller	480	9%	1.9 mm^2

Table III shows the area efficiency of the IO processors, which is obtained by performing experiments for frequently used interface IPs. If implemented in APEX20K, a recent product of Altera, each IP consumes up to 40% of logic elements (50mm^2 of chip area), while the proposed IO

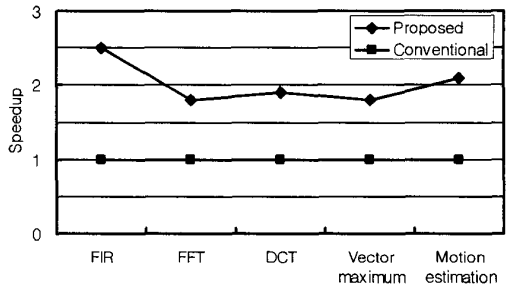

Fig. 5. Speedup for DSP algorithms..

processor can handle the same feature with only 2.5mm^2 of chip area including program and data memories.
The resulting performance of SPPA is 4.75 GOPS, which is more than four times the performance of the state-of-the-art chip (2).

Chip Implementation

The prototype chip is fabricated in 0.25μm CMOS technology with 4 metal layers. A microphograph of the chip is shown in Fig. 6. The chip operating at a single power supply of 2.5V occupies 8x8mm^2 including pads. Simulation and verification results show that the chip can operate at 250 MHz clock frequency in the worst case. Technology and device characteristics are summarized in Table IV.

Conclusion

This paper has presented a configurable device, called single chip programmable processor array (SPPA). The chip is equipped with a 32-bit multithreaded processor that provides the single-cycle context-switching capability, four SIMD processors each of which gives up to 1 GOPS performance, and IO processors that is easy to program and area-efficient. The Task Management Unit is integrated to assist the real-time OS by reducing the overhead of scheduling and synchronization to one cycle. The SPPA delivers more than four times the performance of the state-of-the-art chip.

Acknowledgement:

This work was supported by the Korea IT Industry Promotion Agency through the project IT-SoC, by the Ministry of Science and Technology and the Ministry of Commerce, Industry, and Energy through the project System IC 2010, and by the IC Design Education Center (IDEC).

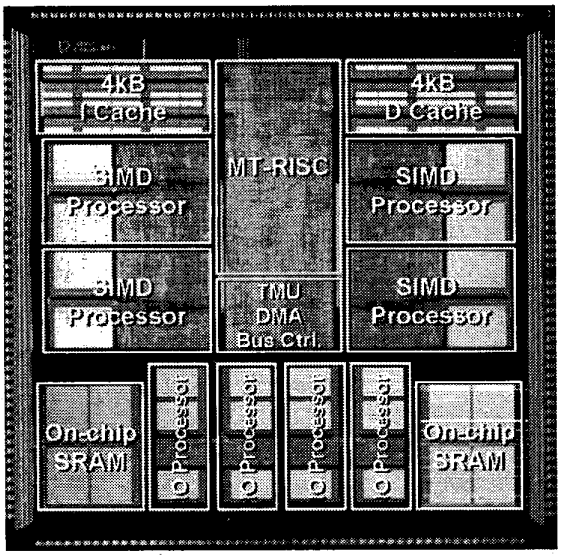

Fig. 6. Die photo of the SPPA.

TABLE IV
TECHNOLOGY AND DEVICE CHARACTERISTICS

Technology	0.25μm CMOS 5-ML
Clock frequency	250 MHz (MT-RISC, SIMD Processors) 125 MHz (IO Processors)
Memory	I$: 4kB, D$: 4kB (128-bit wide) On-chip SRAM: 32kB (32-bit wide) Program Memory: 32kB (32-bit wide) Buffers: 8kB (16-bit wide)
Chip size	8x8mm^2 (pad limited)
Customisable I/O	64 general-purpose inputs/outputs
Power supply	2.5V(core), 3.3V(external, 5V tolerant)

References

(1) Young-Don Bae, Seung-Il Park and In-Choel Park., " A Single-Chip Programmable Platform Based on a Multithreaded Processor and Configurable Logic Clusters," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 10, 2003, pp. 1703-1711.

(2) M. Borgatti, F. Lertora, B. Foret, and L. Cali., " A reconfigurable system featuring dynamically extensible embedded microprocessor, FPGA, and customizable I/O," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 3, 2003, pp. 521-529.

(3) H. Zhang, V. Prabhu, V. George, M. Wan, M. Benes, et al., " A 1-V heterogeneous reconfigurable DSP IC for wireless baseband digital signal processing," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 11, 2000, pp. 1697-1704.

(4) C.P. Ravikumar, "Multiprocessor architectures for embedded system-on-chip applications," in Proc. of International Conference on VLSI Design, pp. 512-519, Jan. 2004.

(5) J. Nickolls, L.J. Madar III, S. Johnson, V. Rustagi, K. Unger, M. Choudhury, "Calisto: A low-power single-chip multiprocessor communications platform," *IEEE Micro*, vol. 23, no. 2, 2003, pp. 29-43.

(6) P.G. Paulin, C. Pilkington, E. Bensoudane, M. Langevin, D. Lyonnard, "Application of a multi-processor SoC platform to high-speed packet forwarding." In *Proc. of Design, Automation and Test in Europe Conference and Exhibition*, pp. 58-63, Feb. 2004.

(7) M. Hirai, T. Mochida, T. Hashimoto, E. Fujii, and T. Kiyohara, "Software control of I/O subsystem on media core processor," *IEEE Trans. On Consumer Electronics*, vol. 44, no. 3, 1999, pp. 939-943.

(8) V. Lappalainen, T.D. Hamalainen, and P. Liuha, "Overview of research efforts on media ISA extensions and their usage in video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12 , no. 8, 2002, pp. 660-670.

(9) Erick Norden, "A Multithreading Extension for Low-Power, Low-Cost Applications," In *Proc. of Embedded Processor Forum*, June 2003

(10) Ubicom Corp., "A Next-Generation Packet Processor for Wireless Networking", April 2003.

(11) Altera Corp., "APEX 20K Devices: System-on-a-Programmable-Chip Solutions," http://www.altera.com.