Radiation-Hard Associative String Processor—a High Density Scalable SIMD Architecture

G. FRIEDMAN
Hughes Aircraft Company, 6155 El Camino Real, Carlsbad, California 92009, USA
and
R. Mike LEA
Aspex Microsystems Ltd., Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

Abstract

A radiation-hard high performance Associative String Processor (ASP) is under joint development by Hughes Aircraft Company and Aspex Microsystems. This system merges the architecturally advanced processor developed by Aspex with the technological capability of Hughes to provide a fine-grain massively parallel processor (MPP) which is easily scalable and provides high throughput while being tolerant to radiation.

The ASP is being implemented in the Hughes 1.25 μm Silicon-On-Sapphire (SOS) Complementary Metal Oxide Semiconductor (CMOS) technology, thereby enjoying its inherent tolerance to radiation. In addition, specific process and circuitapproaches are being used to enhance its native hardness. The SOS ASP will operate above 40 MHz., and for a 16,384-APE (Associative Processing Elements) VLSI ASP module; a peak performance of 13.7 Giga-OPS (for a 12-bit addition) and an I/O data rate of 40 Mbytes/second is expected.

1 Introduction

Real-time systems sense and process data to derive a useful response to recognized entities. The processing of data can be decomposed into three subprocesses: 1) detection—the recognition of useful information, 2) analysis—the qualitative and quantitative scrutiny and transformation of this information into a simpler form, and 3) decision—the application of judgement based on the simplified information such that a useful response to the initial stimuli that was sensed can be made. This sensor/response paradigm is common to all kinds of systems, from biological to electronic. A common trait of this paradigm is the massive amount of data that is typically sensed. This voluminous quantity of data must be efficiently and accurately reduced in such a way that only pertinent useful information remains.

Figure 1 displays an electronic version of this system in which numeric data derived from input signals must be transformed into simpler symbolic data from which a response can be made. These types of systems are common in military and commercial space-based satellites, areas in which Hughes is particularly experienced. Engineering requirements of these types of systems are typically very high throughput, small size and weight, low power, and high reliability.

The sensor/response paradigm represented in Fig. 1 can also be applied to particle detection problems in high-energy physics (HEP). In this environment, massive quantities of data are sensed at very high rates. This information is passed through 1st and 2nd level triggers to rid the data set of less useful information, keeping only that data requiring further analysis. Due to the complexity of the data, sophisticated algorithms are required to effectively perform the 2nd level trigger. Additionally, this information must be processed at very high speed since on-line real-time processing is required. Finally, to improve the speed and efficiency of each experiment, these processors should

COMPUTING IN HIGH ENERGY PHYSICS '91

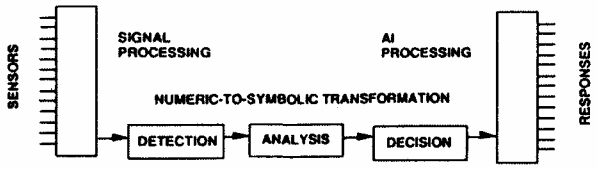

Figure 1: Processing Flow Diagram of Sensor/Response Paradigm

be placed physically close to the detector. However, due to the nature of the HEP experiments, high levels of radiation are encountered. Again, note the similarity to the space-based satellite environment. Therefore, a radiation-hardened high speed programmable processor must be used in order to satisfy the 2nd level triggering requirements of the HEP experiments.

This paper describes a joint effort between Hughes Aircraft and Aspex Microsystems to satisfy these conflicting and aggressive requirements for the HEP community. In section 2, the ASP architecture is described. Both Very Large Scale Integration (VLSI) and Wafer Scale Integration (WSI) technology tracks for the ASP architecture are discussed. These approaches are being monolithically integrated into the Hughes 1.25 μm radiation-hard CMOS/SOS technology. This technology, particularly its radiation hardened qualities, is described in section 3. Finally, some conclusions and some brief comments on future developmental activities are described in section 4 of this paper.

2 The ASP Architecture and its Software, Technology Tracks, and Performance

ASP (Associative String Processor) modules (and support software) comprise highly versatile building blocks for the simple construction of dynamically reconfigurable low-MIMD (Multiple Instruction Multiple Data)/high-SIMD (Single Instruction Multiple Data) second-generation Massively Parallel Processor (MPP) systems [1], [2]. Indeed, based on state-of-the art microelectronics and packaging technologies, ASP modules constitute a family of packaged MPP configurations for the cost-effective implementation of highly compact application-specific high performance information processing systems.

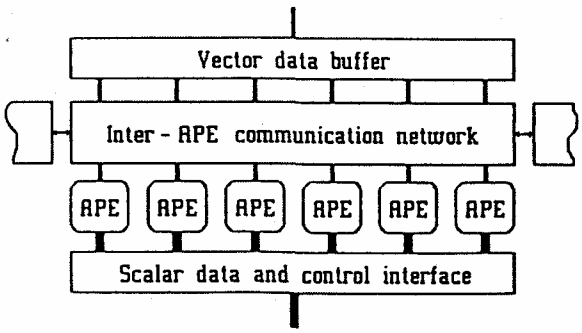

Figure 2: ASP Substring Schematic

Based on scaler-vector content matching rather than location addressing, ASP substrings comprise homogeneous fine-grain SIMD MPP structures, which, in operation, execute a form of set processing (*i.e.*, a sequence of scalar-vector and vector-vector processes) on relevant data. Moreover, application flexibility enables simple tailoring of parallel processing power to match user requirements.

2.1 ASP hardware

An ASP substring developed using VLSI technology is a programmable, homogeneous, scalable fine-grain SIMD MPP, incorporating a string of identical APEs (Associative Processing Elements) and a dynamically reconfigurable inter-processor communication network. ASP substrings can be linked to form longer substrings with simple connection of the LKL (LinKLeft) and LKR (LinKRight) ports, as indicated in Fig. 2.

All APEs share common 32-bit Data, 12-bit Activity and Control Busses, and a 1-bit feedback line (Match Reply, MR) maintained by the ASP substring controller. Each APE incorporates a 64-bit Data Register and a 6-bit Activity Register, a 70-bit parallel Comparator, a single bit full-adder, four status flags (viz., C to represent arithmetic Carry, M and D to tag Matching and Destination APEs, and A to activate selected APEs) and control logic for local processing and communication with other APEs, as shown in Fig. 3.

The APE Data Registers provide storage for vec-

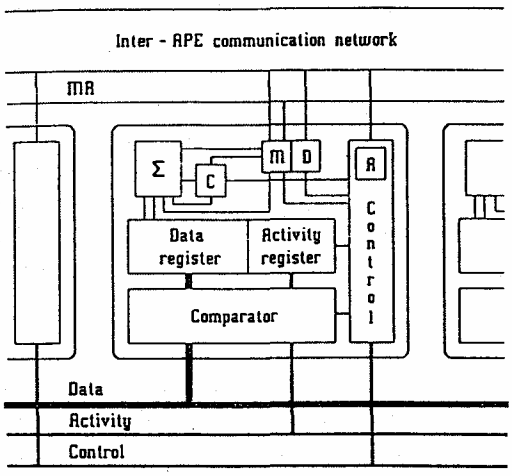

Figure 3: Associative Processing Element (APE)

tor data elements; whereas, the Data Bus supports scalar data which are directly broadcast or received by the ASP substring controller.

In contrast to more traditional SIMD parallel computer architectures, the ASP is based on scalar-vector content matching (therein the name associative) rather than location addressing. Thus, in operation, a subset of APEs are selected, for subsequent parallel processing, by comparing their (vector) data and activity content with the (scalar) states of the corresponding Data and Activity busses. Matching APEs are either directly activated or source inter-APE communications to indirectly activate other APEs. Active APEs then execute a form of set processing (*i.e.* a sequence of scalar-vector and vector-vector processes) on relevant data. The Match Reply (MR) line indicates which APEs match.

2.1.1 Inter-APE communications network

As indicated in Fig. 2, each ASP substring supports two styles of inter-APE communication, these being 1) bit-parallel single-APE communication via the shared Data Bus and 2) bit-serial multiple-APE communication via the Inter-APE Communications Network. Although there are many occasions when the former can be used to advantage, this subsection concerns the latter.

The Inter-APE Communications Network implements a globally controlled, easily scalable, tightly-coupled, and dynamically-reconfigurable APE interconnection strategy, which supports the cost-effective emulation of common network topologies [viz., arrays (*e.g.*, vector, matrix, and binary n-cube), tree networks, graph (*e.g.*, semantic) networks, address permutation (*e.g.*, exchange, shuffle, and butterfly) networks, and shifting networks]. Most significantly, the APE interconnection strategy supports unlimited modular network extension, via the LKL and LKR ports, as indicated in Fig. 2, to enable simple tailoring of parallel processing power to match user requirements.

In contrast to the networks adopted by first-generation MPP architectures, the Inter-APE Communications Network was not designed primarily for the transfer of actual data between APEs. Instead, communication is restricted to the high speed transfer of activity signals (or M-tag patterns) between neighboring or selected remote APEs (viz., those matching the selection criteria). Since APEs can be easily activated by content-addressing and their data content processed in-situ, the time consuming movement of data can be reduced to an absolute minimum. Thus, the interconnection strategy adopted for ASP substrings supports a high degree of parallelism for local communication and progressively lower degrees of parallelism for longer distance communication. In order to preserve continuity at the two ends of the Inter-APE Communications Network, the LKL and LKR (shown in Fig. 2) allow activation or M-bit signals to be injected and sensed by the ASP controller and act as the left and right neighbors of the leftmost and rightmost APE in the ASP substring, respectively.

Thus, to summarize, ASP modules offer considerable application flexibility, maintaining high efficiency (in computation and communication) over a wide range of signal and data processing applications, due to the ability to 1) configur ASP modules which are well matched to their functional application requirements, 2) pipeline and overlap the sequential input-output and transfer (between ASP substrings) of vector data (via the Data Communication Network) with parallel processing (within ASP substrings) with local and global memories, 3) overlap the sequential input-

output and processing of scalar data (in local or global ASP control units) with parallel processing (in ASP substrings), 4) map different application data structures to a common string representation within ASP substrings (supporting content-addressing, parallel processing, and a dynamically-reconfigurable inter-processor communication network), 5) eliminate processor (location) addressing, for the purposes of achieving unlimited architectural scalability and minimizing the sequential and parallel processing overhead, and 6) minimize inter-processor data movement, with high speed activity transfer between processor subsets and in-situ processing.

In contrast to first-generation MPPs, the ASP has been specifically developed as a silicon efficient parallel architecture. Indeed, the ASP architecture is particularly well matched to the exciting opportunities and exacting constraints of VLSI and WSI fabrication; size and cost reductions and being achieved by 1) maximizing processor packing density towards levels which are more usually associated with memory components, 2) incorporating a highly compact inter-processor communications network on-chip, and 3) independence of the pin input-output (I/O) requirement from the number of on-chip processors (*c.f.*, a two-dimensional processor array where the pin I/O requirement grows as the square root of the array size).

Moreover, the ASP is highly amenable to defect/fault-tolerance; owing to its construction from a large number of identical APEs, lack of location dependent addressing, and simple inter-APE interconnection. Consequently, as reduced feature sizes and increased ship areas drive VLSI chip fabrication technology towards ULSI (Ultra Large Scale Integration) circuits and the prospect of WSI devices, the ASP architecture offers scalability and higher cost-effectiveness. Accordingly, the ASP architecture has been designed to be technology independent; such that it can benefit, without modification, from the inevitable improvement in microelectronics technology.

2.2 ASP software

Although beyond the scope of this paper, the application and system software components of any parallel computer are of equal importance to that of the hardware implementation. Consequently, some aspects of the nature of the ASP software are briefly mentioned, such as that the ASP application programs can be written entirely in familiar block structure high level language, such as Pascal, Modula 2, C, or ADA, under a familiar operating system, such as Unix, VMS, or MS-DOS. Also, these programs permit calls to external utilities.

2.3 ASP technology tracks

Both VLSI and WSI technology tracks are under development. These approaches use the scalable and programmable nature of the ASP architecture. In addition, the WSI technology track further exploits fault-tolerance, so as to be consistent with the technological constraints of WSI Presently, applying either a monolithic or hybrid 1 μm CMOS WSI technology, 8,192-processor WASP (WSI ASP) devices are possible, enabling the assembly of 65,536-processor ASP modules achieving 10 Tera-OPS/ft^3, 1 Giga-OPS/Watt and 10 Mega-OPS/$ in cost-effectiveness.

As mentioned previously, a CMOS/SOS ASP (VASP64/H1) with 64-APEs is under development. This circuit utilizes fully static logic and has the characteristics described in Table 1. The SOS ASP circuit is pin-to-pin compatible with a VLSI ASP developed in bulk CMOS (VASP64/E1) and will operate at a much higher clock frequency (40 MHz.) and is much more radiation hardened Therefore, this integrated circuit will be used in those applications, such as high-energy physics, requiring higher performance, real-time processing and significant tolerance to radiation.

2.4 ASP Performance

Detailed evaluation of the numeric and symbolic processing performance of ASP modules is beyond the scope of this paper. Instead, an overview of ASP performance for common operation is given below.

All ASP programs start with the allocation of virtual APEs to the nodes of the data structure to be processed. Virtual APEs incorporate one APE or a contiguous group of APEs to construct a

64 - APEs (Associative Processing Elements)
90,000 transistors
40 MHz. clock rate
500 milliwatts dynamic power dissipation
295 X 276 mils (edge-to-edge)
1.25 μm CMOS/SOS double level metal technology

Table 1: CMOS/SOS VASP Circuit Characteristics—WASP64/H1

effective Data Register of the required bit-length. Thus, ASP substrings, with a total of N APEs and with n APEs allocated to each virtual APE, could be configured as 1) N/n virtual APEs, each incorporating a 64 by n-bit Data Register. This configuration supports bit-parallel scalar-vector matching and assignment operations and bit-serial scalar-vector and vector-vector arithmetic, logic, and relational operations and 2) N/n virtual APEs, each incorporating 64 n-bit Data Registers. This configuration supports bit-parallel vector-vector arithmetic, logic, and relational operations.

To provide an approximate indication of the performance of the VASP64/H1 operating at a clock rate of 40 MHz, it can be assumed that each step of bit-serial and bit-parallel operations can be executed in 100 ns., since each operation requires four clock periods (time slots). Therefore, the approximate limit of performance for an ASP substring comprising N virtual APEs can be estimated with the following simplified formula:

$$\text{ASP Performance} < \frac{\text{10-N}}{\text{number of steps per operation}} \text{Mega-OPS}$$

Such peak performance forecasts are very misleading, since it is relatively rate for all processors to be active simultaneously. Indeed, this performance estimate refers only to the low level arithmetic operations of numeric-to-symbolic data conversion. A more realistic performance indication can be gained from application benchmarks, which include all processing from data input to the output of the required response. To this end, the ASP architecture has been evaluated in terms of the Abingdon Cross benchmark, the two US DARPA Image Understanding benchmarks set in 1986 and 1988 [3]–[5], and the European CERN LAA (particle Physics track analysis) benchmark in 1989 [6]. Although beyond the scope of this paper, the results indicate leading performance advantages for ASP modules compared with contemporary parallel computers.

3 The Hughes 1.25 μm CMOS/SOS Technology

Hughes has been a leader in the development of radiation hard digital CMOS/SOS circuits since 1973, first with a family of 3.5 μm polysilicon gate devices and more recently with 1.25 μm polycide gate and double level metal polysilicon gate devices. Submicron CMOS/SOS, both 0.8 μm and 0.25 μm, is under development, providing increased density and speed (clock rates from 200 to 400 MHz. are expected) and extremely small speed/power products. The 1.25 μm SOS technology has been used to develop pipelined arithmetic circuits which operate at clock rates of (a tester limited) 150 MHz. at 5 volts while exhibiting a power dissipation characteristic of 0.58 watts/gate/MHz. at 3.3 volts. The reliability of the Hughes 1.25 μm devices is up to 10X the requirements for space applications and the technology is fully space qualified. Accelerated burn-in of SOS parts has verified reliability levels less than 10 FITS (failures per 10^9 device hours).

The 1.25 μm SOS technology is an all ion implanted, all dry etched polysilicon gate double level metal process. The technology is formed by the etching of completely isolated silicon islands on a 0.5 μm silicon-on-sapphire epitaxial layer. The thin active layer of silicon on the single crystal sapphire (Al_2O_3) provides perfect isolation between the transistors and significantly reduced device and interconnect parasitic impedances. This structure results in superior transient and single event upset radiation hardness and immunity to latch-up. Neutron hardness is very high ($> 10^{15}$ n/cm^2) as are all MOS majority carrie technologies, while total dose hardness can be optimized with the same processing techniques used on bulk CMOS. Total dose hardness levels above 1 Mrad (Si) for the 1.25 μm process are achieved. Back channel leakage due

to total dose radiation is controlled by the use of specially optimized SOS material.

The major challenge in hardening CMOS/SOS is in the total dose hardness which is generally dominated by gate oxide quality. Total dose behavior has been extensively studied by wafer testing on an Aracor X-ray source and Co^{60} irradiation of packaged parts. N-channel data from devices with gate oxides of 400 Åshow that up to about 2×10^6 rads (S_iO_2) device behavior is dominated by threshold shifts and that mobility drops are minimal ($\sim$ 10%). SEU rates as low as 10^{-10} upsets/gates-day (where one gate is composed of four CMOS transistors) are exhibited in SOS due to its limited film thickness and small area of vulnerability.

In addition to the inherent radiation tolerance of the SOS technology, circuit approaches can be used to further enhance the hardening of specific circuits. Design techniques, such as using only static logic instead of dynamic logic, adding RC impedances to critical nodes for improved SEU immunity, power bus sizing optimized for transient radiation, substrate tie-downs to minimize threshold shifts, and restricting the maximum number of serially stacked transistors, are used to enhance the radiation hardening of specific circuits.

4 Conclusions and Future Plans

In this paper, a high performance VLSI Associative String Processor VASP64/H1 is described. This parallel processor is being designed to satisfy the specific performance and environmental conditions required to perform real-time on-line 2^{nd} level triggering close to High-Energy Physics (HEP) particle detectors. The circuit exploits the radiation-hardened and high speed performance characteristics of the Hughes CMOS/SOS 1.25 μm technology required by the HEP community. The circuit will be a plug-in replacement for the present bulk CMOS VASP circuit; providing greater speed ($\geq$ 40 MHz. clock rate) and hardness to radiation.

Higher levels of integration and improved ASP throughput will be achieved beyond the VASP64/H1 by applying WSI technology and three-dimensional wafer stacking. These technology improvements, merged with the specialized fault-tolerant and scalable nature of the ASP architecture, will provide unprecedented levels of cost-effectiveness (viz. up to 100 Tera-OPS/ft^3, 1 Giga-OPS/Watt, and 10 Mega-OPS/$) and reliability.

References

[1] R.M. Lea, "The ASP: a cost-effective parallel microcomputer," *IEEE Micro* (1988) 10.

[2] R.M. Lea, "ASP modules: cost-effective building blocks for real-time DSPsystems," *J. of VLSI Signal Processing* (1989) 61.

[3] A. Krikelis and R.M. Lea, "Performance of the ASP on the DARPA architecture benchmark," *Proc. Frontiers 88, 2nd Symp. on the Frontiers of Massively Parallel Computation* (1988) 483.

[4] A. Krikelis, I. Kossioris, and R.M. Lea, "Performance of the ASP on the DARPA architecture benchmark II, " *Proc. DARPA Image Understanding Benchmark Workshop* (1988).

[5] C. Weems, E. Riseman, A. Hanson, and A. Rosenfeld, "The DARPA image understanding benchmark for parallel computers," *J. of Parallel and Distributed Computing* II (1991) 1.

[6] S. Lone, R.K. Bock, Y. Ermolin, W. Krischer, C. Ljuslin and K. Zografos, "Fine-grain parallel computer architectures in future triggers," *CERN report CERN-LAA RT/89-05* (1989).