A Wafer Scale Programmable Systolic Data Processor

Dave Landis, Joe Yoder, & Denny Whittaker
University of South Florida
Center For Microelectronics Research
4202 E. Fowler Ave, ENG 118
Tampa, FL 33620-5350
(813) 974-2096

Tim Dobbins
Honeywell, Inc
Space and Strategic Avionics Division
13350 US Highway 19 South
Clearwater, FL 34624
(813) 539-2510

ABSTRACT: *This paper describes the Programmable Systolic Data Processor (PSDP) which is being jointly developed by the University of South Florida Center for Microelectronics Research and the Honeywell Space and Strategic Avionics Division[1]. The PSDP will enhance DoD mission capabilities by extending signal and data processing speed/performance while reducing system size, weight, and power consumption. The characteristics of this architecture which make it opportune for building as a wafer scale system include: broad homogeneity, ease of redundancy, and limited physical interconnect requirements. The PSDP exploits the high interconnect bandwidth of WSI using a robust programmable systolic array processing architecture. Thus, it will provide unique on-board processing capabilities for DoD missions.*

I. Background

Wafer Scale Integration (WSI) has been proposed for many DoD critical mission applications [1-4]. Physical interconnection and packaging of a multi-wafer system are key to the full exploitation of WSI in military applications. Honeywell and CMR are working as a team to develop the PSDP architecture as well as an advanced stacked 3D WSI interconnect technology. This combination of architectural and interconnect innovations will intrinsically boost system performance and reliability, while lowering size, weight, and power consumption. Several years ago, Honeywell studied the currently available signal processing technologies to determine the best candidate for several satellite based missions. The major throughput driver was determined to be IR signal processing algorithms for surveillance satellites and probes. These algorithms are most efficiently executed using simple highly parallel processors. At that time, the best known commercially available processor fitting this model was the NCR Geometric Arithmetic Parallel Processor (GAPP) [5].

Numerous fine grain signal processors have been implemented in VLSI over the past few years [5-8]. However, they generally require a complex off-chip controller. If these designs were directly mapped to WSI technology, they would result in a heterogenous WSI implementation with separate processors and controllers. The complexity of such a design results in higher system cost because of yield and harvesting problems. Furthermore, heterogenous WSI designs typically encounter communication and throughput degradation as they scale to large numbers of nodes.

II. PSDP Architecture

A joint effort is underway at Honeywell and CMR to study homogenous multicomputer designs for wafer scale implementation. We are developing the PSDP architecture to achieve throughput advantages over its predecessors. Architectural innovations include distributed control and software allowing each processor to act as a self-controlled node containing its own programming. There are only three global signals distributed across the wafer: the clock and two mode signals for self-test and initial start-up of the machine. Such a design has the inherent scalability needed to extend fine-grain multicomputer architectures to large numbers of processors. This minimizes the risk of large system prototype development, experimentation, and performance enhancement by building on a well-defined base.

The PSDP uses a 4/8-bit programmable systolic processing element which communicates with its neighboring cells over a 4-bit data path. This communication can be performed concurrently with internal 8-bit ALU operations and 4 X 4-bit multiplications. Figure 1 shows a high level block diagram of the PSDP cell. It illustrates the internal cell architecture and identifies the data paths between the major functional blocks (RAM, ALU, Multiplier, Latches, and Multiplexers). The PSDP architecture takes advantage of an orthogonal memory structure which allows 8-bit data access as well as 32-bit instruction access. Distributed memory consists of a 2K-bit dual port static RAM (SRAM) located within each PSDP processor cell. An SRAM was chosen over a dynamic RAM because it has lower power and does not need to be refreshed. The memory is organized with a word width of 32-bits and with either 4-bit or 8-bit read/write capability. Each processing element in the PSDP array can

[1] This work is being supported by a grant from the Florida High Technology and Industry Council and by Honeywell, Inc.

0-7803-0109-9/91/0000-0252$01.00 © 1991 IEEE

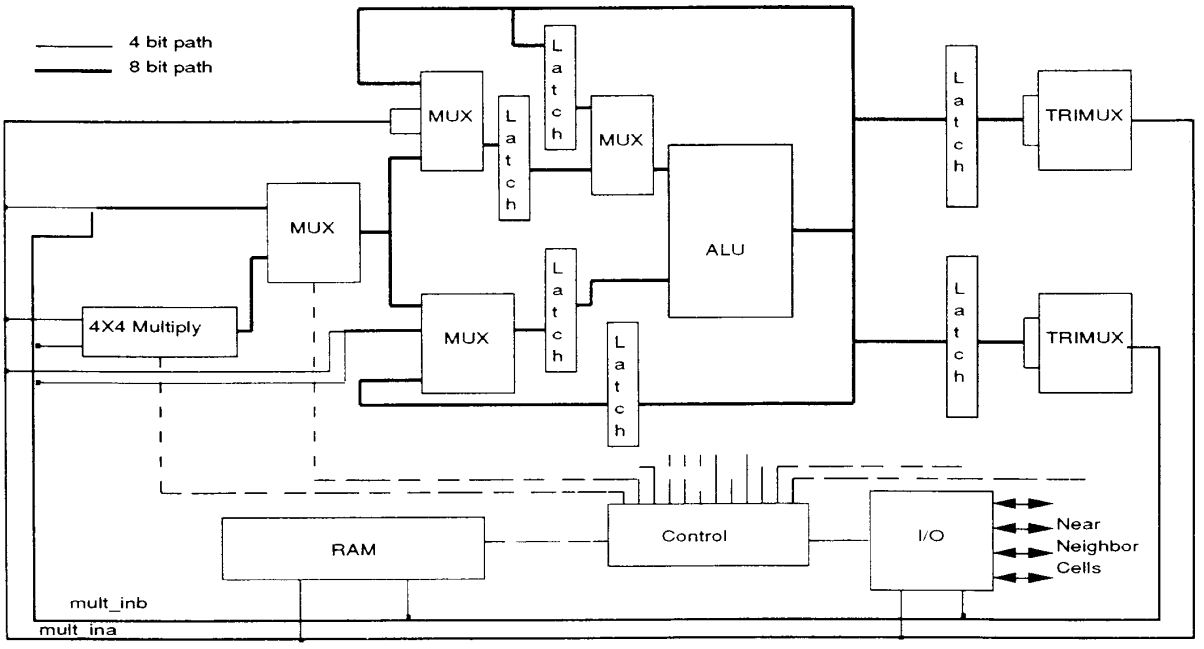

Figure 1. PSDP Processor Cell Architecture

be individually and uniquely programmed by downloading appropriate instructions into its RAM. Thus, both SIMD and MIMD machine architectures and algorithms can be implemented using the PSDP system.

The processor cell size is estimated at 5mm X 6mm (including inter-cell wiring area) using 1.2 micron CMOS design rules. The PSDP processor cell has a projected performance of 9 million 8-bit operations per second (Mops). Figure 2 shows a proposed planar prototype system which consists of five wafers each containing a 12 x 16 mesh of working processors. The 960 node multicomputer would operate at a performance of 8640 Mops (peak performance with 8 bit operations). Table 1 provides a detailed comparison between the proposed PSDP prototype (figure 1) and a Multi-Chip Module (MCM) design. The MCM uses commercial TI TMS320C30 digital signal processor IC's to perform the same signal processing functions as the PSDP. A practical system level figure of merit is the number of operations which can be performed for a given volume and power consumption. The table indicates that the WSI PSDP figure of merit (Kops/cm^3-watt) exceeds that of the MCM by more than an order of magnitude.

Table I. MCM TMS320C30 vs WSI PSDP

Attribute	MCM w/ TMS320C30's	WSI w/ PSDP
Cell Size	150mm^2	30 mm^2
Clock Rate	33Mhz	25Mhz
Power/Chip	1.5 Watts	.3 Watt
Card Config.	16 IC's/MCM 4 MCM/card	192 nodes/WSI 5 WSI/card
Cards/System	8 proc. 1 memory	1 proc. 1 memory
Nodes/System	400	960
System Power	700 Watts	150 Watts
Figure of Merit (Kops/cm^3-w)	193	2472

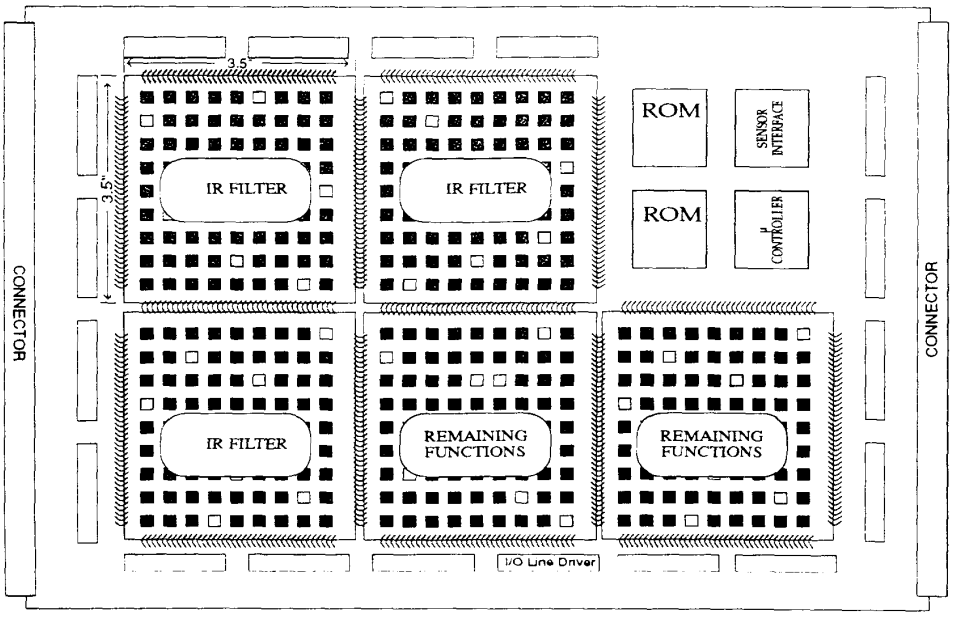

Figure 2. Proposed PSDP Planar Prototype System

III. Wafer Scale Implementation

Optimization of the PSDP processor architecture is required for wafer scale implementation. We are addressing power minimization and enhanced testability issues, as well as design optimization for defect tolerance and inter-cell connectivity. A preliminary study was conducted to estimate cell area and yield in order to evaluate internal redundancy options. Analysis was performed using the Stapper model [9], and also using models and data derived from a commercial process [10]. We found that the total area of the processor, excluding RAM, occupies less than 10% of the total chip area. As indicated by the top most curve of figure 3, the basic processor and ROM have a very high yield because they consume such a small area. Thus, the preliminary yield analysis indicates that redundancy within the processor cell is not necessary to achieve cell yield objectives. The extent to which the total cell yield is dominated by the two-port RAM is further indicated by the bottom two curves on figure 3 (which reflect RAM yield without redundancy). The figure also shows that by including RAM redundancy in the form of spare rows, cell yield can be dramatically improved. Thus, we conclude that redundancy for the RAM will be required to obtain a 90% RAM yield, which allows us to achieve our overall cell yield objective of greater than 80%.

A complete multicomputer system can be constructed using the programmable systolic processing element building block. The homogeneity of this architecture maximizes the potential use of all functional cells on each wafer. The PSDP operates with nearest neighbor communication (including virtual nearest neighbors) making the design easy to harvest into a 2D mesh of nodes on a wafer. Furthermore, the low cell I/O count minimizes the wafer area required for processor-to-processor interconnections. With a projected cell size of 5mm X 6mm (including interconnect wiring and power bussing), 238 processors can be fabricated on a 5 inch wafer. As previously described, a conservative cell yield of more than 81% is predicted with redundancy in the RAM. This will be sufficient to harvest a 12 x 16 2-D processor mesh per wafer.

CMR has developed a VHDL behavioral model of the PSDP processor cell to verify the functionality of the basic cell architecture. In addition, we are developing system models for mesh connected arrays of PSDP cells. Beginning with 2X2 and 4X4 blocks, we are attempting to scale up to a complete 10 node by 10 node architectural VHDL model. These models are being used to perform system level evaluation and trade-off studies. Results obtained from analysis of the 10 x 10 model will be extended to a generic N x N system, as illustrated in Figure 4. The VHDL cell model has been transferred to a Honeywell CAE system and is being used for high level system trade-off analysis.

IV. Interconnect and Packaging

One successful method for providing defect avoidance at the wafer level has been developed and demonstrated at the MIT Lincoln Laboratories [11,12], and transferred under DARPA support to CMR. This technique uses a laser to configure wafer level interconnections after wafer probe tests have identified defective components. Both additive and deletive interconnections can be made with this technology using laser diffused link structures on the wafer. Alternative methods for restructuring of WSI designs are being investigated at CMR, including focused ion beam and laser enhanced deposition technologies. These next generation link technologies are centered on "Back End Of the Line"

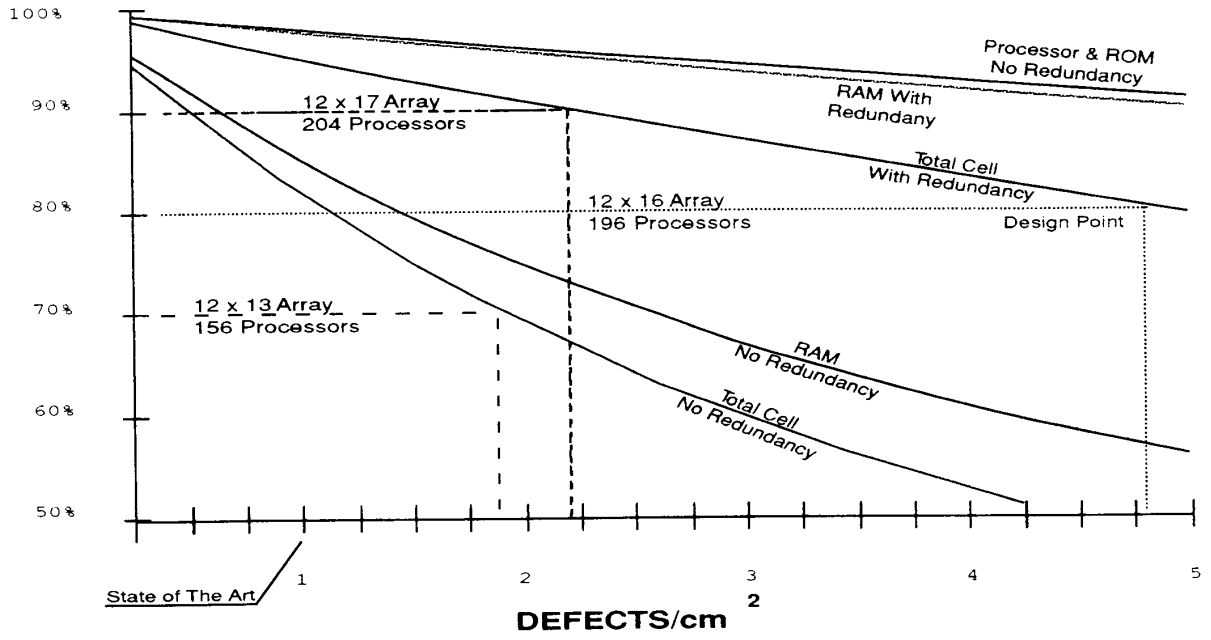

Figure 3. PSDP Projected Cell Yield

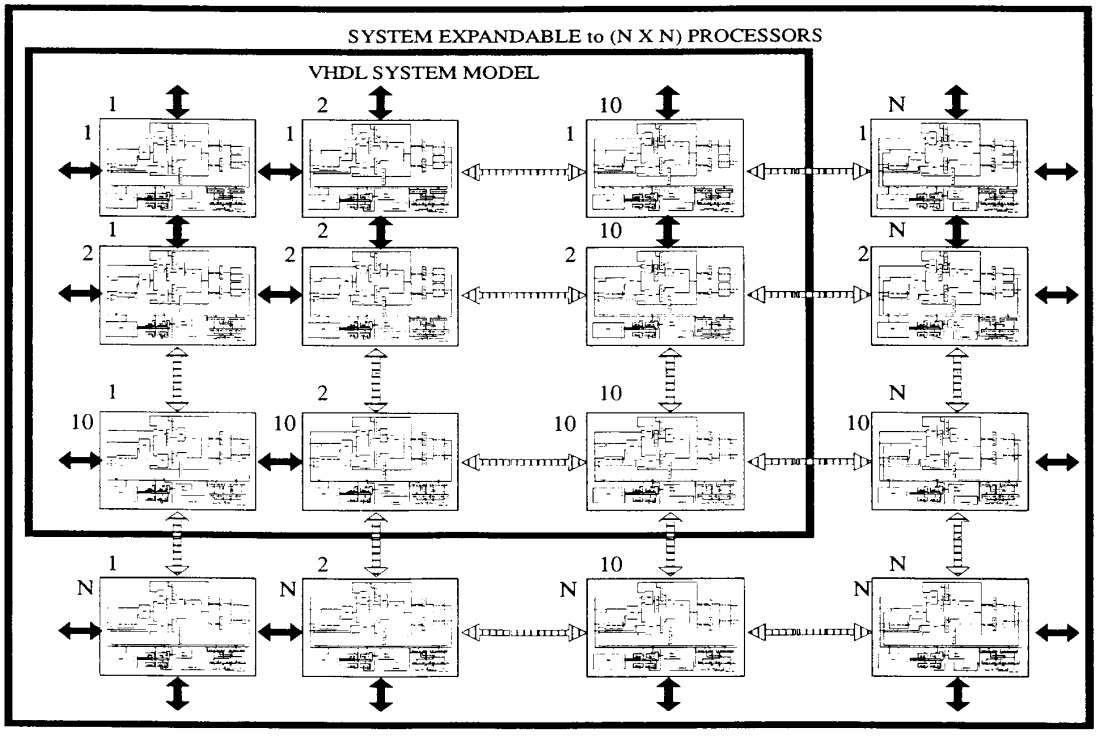

Figure 4. N X N node PSDP Processor Array Model

link processes which are not affected by most transistor and processing parameter variations. Focused ion beam (FIB) system generated micro-connects have been demonstrated by CMR at the VLSI chip level in cooperation with Micrion Corporation.

A key element in the vertical stacking of die or 3D-Stacked Wafer Scale Integration (SWSI) involves the etching or drilling of holes through wafers to produce the necessary Z-axis vias. These holes must be insulated with a dielectric and then plated or filled with a conducting material. Previous work by CMR has successfully demonstrated laser drilling of 1-2 mil diameter holes in silicon. This was accomplished with a NdYAG TEM00, CW q-switched laser at 1.06um using approximately 12 watts. Experimental evaluations are currently in progress to determine the necessary ground rules so that nearby active devices are not adversely affected by this process. An alternate approach under consideration uses photograph exposure of resists, followed by plasma etching of vias. Deposition of the insulating films will be performed either by RF plasma deposition, by ECR plasma deposition, or by reactive sputter deposition. The metal conductors will be deposited by sputter deposition or by evaporation and will be etched using a reactive ion etch plasma system. The interspersing of multi-layer dielectric circuit layers will be investigated as a means to allow greater flexibility in I/O routing. Wafers or die with vertical conductive interconnects can provide the foundation for global communications using backside interconnections.

The through-the-wafer interconnect technology just described will be used to develop a prototype stacked wafer scale system. This Z-axis interconnect capability provides fundamental system level advantages because memory can be significantly expanded with very little increase in "distance" from the processor to the memory. A processor memory pair (PMP) provides the short memory distance by placing a memory wafer (or wafers) directly below the processor wafer with Z-axis interconnections through the wafers. The short distance from a processor to a large memory, along with the ability to fabricate memory and logic from their optimum process technology, can result in up to a 100 times improvement in performance. Alternatively, the Z-axis interconnect technology can be used to construct a 3D mesh using the self-contained 2D PSDP processor cells. This allows the basic architecture to scale to much large numbers of nodes within a given system volume.

V. Summary & Future Work

Work is continuing to develop the optimal design point for a wafer scale programmable systolic data processor. Design tradeoffs are being made considering architectural, interconnection, and packaging constraints. In addition, we are including design for test considerations during the preliminary stages of architectural development and cell design. This will allow us to verify aspects of our test methodology when the PSDP processor cell design has been completed and fabricated as a test chip. The complexities of test for wafer scale systems are being addressed with a four point approach: 1). development of a structured test methodology based on use of the IEEE 1149.1 (JTAG) Standard Test Interface, boundary scan, and built-in test [13], 2). incorporation of test structures in selected regions of the wafer to allow for wafer acceptance testing based on parametric measurements, 3). application of scan test at wafer level through probing of the 4-pin IEEE 1149.1 test access port, and 4). performance testing of the restructured wafer scale system using functional and manufacturing test patterns. Our test hardware, a Hewlett-Packard 82000 Test System, has a tester-per-pin architecture supporting up to 384 pins at test speeds of up to 200MHz. Incorporation of the IEEE 1149.1 test access port greatly simplifies the initial wafer level structural testing by providing an easy 6-point probe (including power and ground pins) to be used for cell verification [14]. This minimizes the difficulties of large pin count probing commonly required for VLSI chips and WSI systems.

References:

[1] C. Adams, "Chasing the Elusive Gigaflop In a Soup Can", *Military and Aerospace Electronics*, February 1991, pp. 26-29.

[2] M. J. Little, et. al., "The 3-D Computer", *Proc. 1989 Intl. Conf. on Wafer Scale Integration*, January 3-5, 1989, pp. 55-64.

[3] D. L. Allen, et. al, "MUSE, A Wafer-Scale Systolic DSP", *Proc. 1990 Intl. Conf. on Wafer Scale Integration*, January 23-25, 1990, pp. 27-35.

[4] J. Brewer, "Practical Considerations for WSI Based Systems", *Proc. 1991 Intl. Conf. on Wafer Scale Integration*, January 29-31, 1991, pp. 321-327.

[5] Manufacturers Data Sheet, NCR 45CG72, NCR Geometric Arithmetic Parallel Processor, 1987.

[6] R. Hughey and D. Lopresti, "Architecture of a Programmable Systolic Array", *Proc. 1988 Intl. Conf. on Systolic Arrays*, May 25-27, 1988, pp. 41-49.

[7] D. Foulser and R. Schreiber, "The Saxpy Matrix-1, a general purpose systolic computer", *IEEE Trans. Computers*, vol. 20, July 1987, pp. 35-43.

[8] K. Chen, et. al, "PASIC, A Sensor/Processor Array for Computer Vision", *Proc. 1990 Conf. on Application Specific Array Processors*, Sept. 5-7, 1990, pp. 352-366

[9] C. H. Stapper, "Yield Model for Fault Clusters within Integrated Circuits", *IBM Journal Research and Development*, Vol. 28, 1984, pp. 636-640.

[10] J. P. Cunningham, "The Use and Evaluation of Yield Models in Integrated Circuit Manufacturing", *IEEE Trans. Semi. Manuf.*, Vol.3, No.2, May 1990, pp. 60-71.

[11] P.A. Wyatt & J. I. Raffel, "Restructurable VLSI - A Demonstrated Wafer Scale Technology", *Proc. 1989 Intl. Conf. on Wafer Scale Integration*, Jan. 3-5, 1989, pp. 13-20.

[12] A. H. Anderson & R. Berger, "RVLSI Applications and Physical Design", *Proc. 1989 Intl. Conf. on Wafer Scale Integration*, Jan. 3-5, 1989, pp. 39-45.

[13] IEEE Standard Test Access Port and Boundary-Scan Architecture, *IEEE Std 1149.1-1990.* May 21, 1990.

[14] D. Landis, "A Self-Test Methodology for Restructurable WSI", *Proc. 1990 Wafer Scale Integration Conf.*, Jan. 23-25, 1990, pp. 258-264.