Miniaturized, Low-Power Parallel Processor for Space Applications

W. J. Jacobi, N. J. Teneketges and L. A. Wadsworth

Space Computer Corporation, Santa Monica, CA 90404

A miniaturized, low-power parallel processor for space applications is under development by Space Computer Corporation for DARPA's Advanced Space Technology Program. The basic goal of this project is the reduction, by an order of magnitude or more, of on-board processor weight, size and power consumption for space-based sensor systems. Our approach for achieving this goal is to use low-power VLSI devices which maximize throughput per watt, together with three-dimensional hybrid wafer-scale integration and packaging technology. In its prototype version, a 12-node processor will have a peak throughput greater than 1.2 GFLOPS and occupy a volume less than 15 cubic inches.

1.0 INTRODUCTION

This paper describes the miniaturized, low-power SCC-100 parallel processor for space applications under development by Space Computer Corporation. This effort is sponsored by DARPA's Advanced Space Technology Program [1]. The basic goal of the SCC-100 project is the reduction, by an order of magnitude or more, of on-board processor weight, size and power consumption for space-based sensor systems. Such processors must employ parallel architectures with multiple processing elements in order to achieve the high throughputs required (up to tens of billions of operations per second). Our approach for achieving this goal is to use advanced VSLI implementations which maximize throughput per watt, together with three-dimensional hybrid wafer-scale integration (HWSI) packaging technology. Once the size of the processor is reduced, it is possible to use shielding techniques not practical with larger-volume equipment, thereby reducing or even eliminating the need for high-cost, radiation-hardened components.

Figure 1 shows a full-scale model of the prototype version of the SCC-100 space processor with 12 processing nodes. This processor, which has a peak throughput in excess of 1.2 GFLOPS, occupies a volume of less than 15 cubic inches. If a 300 mil thick radiation shield (not shown) is used to enclose and protect the processor assembly, the volume occupied is 35 cubic inches.

2.0 PROCESSOR REQUIREMENTS

A basic requirement for the space processor is *high throughput*. Depending on the sensor type, sensor parameters and processing functions performed, this throughput can range from hundreds of millions to billions of arithmetic operations per second [2,3,4]. The only way in which such a wide range of performance objectives can be achieved is by use of a parallel processing architecture with multiple processing elements or "nodes".

The second requirement is *flexibility*. This implies two important characteristics: programmability and modularity. It is essential that the processor be general-purpose rather than special-purpose in nature, with

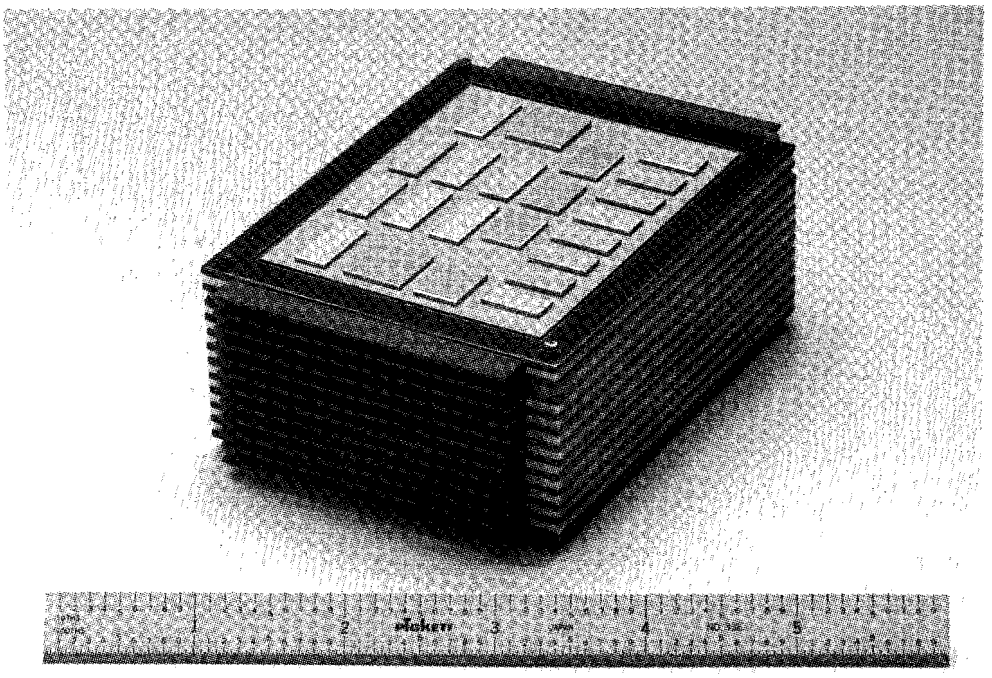

Figure 1. Full Scale Model of Miniaturized Space Processor with Cover Removed. (Throughput = 1.2 GFLOPS)

the ability to be programmed for a wide range of algorithms. Modularity can be achieved by the use of a scalable architecture implemented with varying numbers of identical nodes, depending on the throughput requirement.

The final requirement is *small size, weight and power consumption*. Low power consumption is particularly important, since the weight of the hardware required for power generation, power distribution, power conditioning and cooling can easily exceed the weight of the processor itself.

3.0 PROCESSOR ARCHITECTURE

A wide variety of parallel processing architectures have been developed over the past two decades. At one extreme, "coarse-grained" machines like the Cray X-MP supercomputer use a small number of very powerful processors. Each processor has a sophisticated architecture and is built using the fastest available circuit technology. This approach is unsuitable for space applications because of the large number of devices, the large size and weight, and the high levels of power required.

At the opposite extreme are "fine-grained" machines based on very simple, bit-serial processing elements. Although any one processor is of little value by itself, the aggregate computing power can become very large when many such processors (e.g., tens of thousands) are coupled together. Most machines of this type have a so-called SIMD (single-instruction, multiple-data) structure. Examples include the Connection Machine and the DAP (Digital Array Processor). They require custom VLSI chips and custom software, and are generally not suitable for real-time, high-data-rate applications.

A third approach combines an intermediate number (tens to hundreds) of medium-grained processing elements utilizing standard VLSI microprocessor chips. Each of the identical nodes contains one or several

microprocessors with local memory plus provisions for communicating with other nodes in a network. There is no central shared memory; processing tasks are executed separately and concurrently on different nodes. Most machines of this type utilize a MIMD (multiple-instruction, multiple-data) architecture, with 32-bit floating-point arithmetic. Examples include the NCUBE and the Intel Touchstone machines designed for scientific and engineering applications.

We have selected the medium-grained, MIMD approach for the SCC-100 space processor. A major advantage of this approach is the fact that it can utilize standard, off-the-shelf, state-of-the-art, low-power CMOS microprocessor and memory devices. Such devices are available in low-cost versions from a variety of merchant suppliers. These firms are making large investments to develop successive generations of devices with progressively improved performance, including associated support software which is indispensable for applications software development. For applications which require a very high degree of radiation hardness, it is possible to fabricate these devices with a radiation-hard semiconductor process such as CMOS/SOS (silicon-on-sapphire) or CMOS/SOI (silicon-on-insulator).

For maximum computational efficiency, the communications network topology should match the structure of the algorithms which are to be executed. Signal processing algorithms involve such operations as matrix arithmetic, digital filtering, correlation, the Fast Fourier Transform (FFT), sorting, and thresholding. Such algorithms share the key attributes of regularity, recursiveness and local communication. They have a simple linear structure ideally suited to pipelining. In pipelining, a task is broken up into sequential segments that can be executed one after the other. The segments are selected to have nearly equal execution times (say Δt). The total time for execution of the complete task, referred to as the latency, is then Δt times the number of segments. For most applications, a reasonably large value of this latency can be tolerated (on the order of seconds). Pipelining is an extremely effective method for achieving high throughput and in fact is very widely employed in signal processing architectures. A key requirement for such architectures is a high rate of data transfer along the linear dimension, i.e., a high communications bandwidth between successive computation blocks in the algorithm chain.

A small fraction of the signal processing computations do not have the foregoing characteristics and cannot be performed in a simple sequential manner without data-dependent branching. Such computations cannot be pipelined or mapped onto a simple linear structure. Almost without exception, however, they involve slowly-changing variables such as processing parameters or thresholding levels. Thus they do not require a high throughput or communications bandwidth. A practical signal processor architecture must include a component which can execute these computations efficiently without pipeline interference.

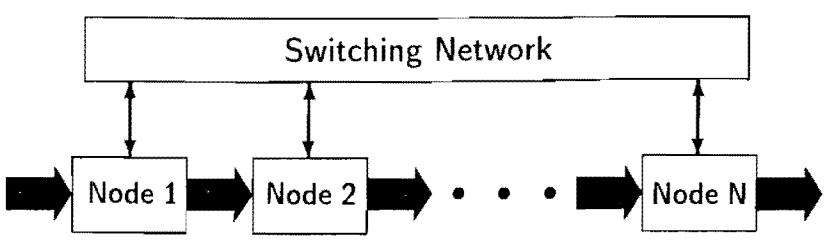

Figure 2. SCC-100 Architecture.

In accordance with the foregoing considerations, we have selected the architecture shown in Figure 2 for the SCC-100 space processor. It consists of a linear array of processing nodes with a very high-bandwidth

parallel communications channel between adjacent nodes (indicated by the thick arrows) that is ideally suited for pipelined computations. Each node also has a low-bandwidth serial link (indicated by the thin arrows) connected to a switching network that can connect any node to any other node. In this manner arbitrary, low-speed network topologies can be established for those computational structures which are not suitable for pipelining. The links can also be used for processor monitoring and control functions such as fault detection, network reconfiguration, program down-loading, etc.

This architecture is easily scalable to accommodate different numbers of nodes for different throughput requirements. It can also be easily modified to incorporate fault tolerance. For this purpose, spare nodes are added to the array and the switching network is duplicated to avoid single-point failures. Each of the operating nodes is continuously monitored to detect possible faults. When a fault is detected, the failed node is bypassed, one of the spare nodes is activated, and the program is re-loaded.

4.0 HARDWARE TECHNOLOGY AND IMPLEMENTATION

4.1 VLSI Implementation

Implementation of the the SCC-100 architecture shown in Figure 2 is accomplished with standard VLSI microprocessors and memories, together with application-specific integrated circuits (ASICs) for device interfacing and inter-node communications. In order to minimize power consumption, low-power CMOS devices which maximize throughput per watt are employed. A variety of very high performance, single-chip, programmable microprocessors which provide peak throughputs greater than 15 MFLOPS/watt are available (e.g. Intel i860, Motorola DSP96002, Texas Instruments TMS320C30, and Zoran ZR34325). The implementation chosen for the present version of the SCC-100 node requires three Zoran ZR34325 vector signal processors, each with a peak processing throughput of 37 MFLOPS, one Inmos T805 Transputer for housekeeping functions, and 2 MBytes of static RAM. Each processing node has a peak throughput greater than 100 MFLOPS with a peak power consumption of only 5 watts. A parallel system composed of 12 nodes will therefore have a peak throughput greater than 1.2 GFLOPS and a peak power consumption of 60 watts. It is possible to utilize a number of such systems, either in an extended pipeline or in a parallel-pipeline configuration to achieve still higher throughputs. Average system power may be further reduced by using power switching techniques to power down temporarily-inactive nodes in the system.

Protection from the effects of radiation on circuit function can be provided in several ways. With sufficent shielding, total dose requirements can be met with commercial CMOS components, especially in the case of low-altitude circular orbits and short-duration elliptical orbits. Given the small physical size of a 12-node SCC-100, the entire processor can be enclosed in a 300 mil thick radiation shield without a large weight and volume penalty. For more stringent orbits and in cases where nuclear weapons effects must be considered, special radiation-hardened components may be needed. In all cases, it is necessary to mitigate single-event upset (SEU) errors, through error detection and correction circuitry and/or through special software routines. The approach taken to mitigate SEU errors in the SCC-100 design is to detect and correct errors with the hardware, and to periodically scrub the memories of stored data through software routines.

4.2 Silicon-Based HWSI Technology

Hybrid Wafer Scale Integration (HWSI) technology is employed to interconnect the various components that comprise a processing node in the system. The use of HWSI provides the size, weight and power reductions that allow a processor system capable of billions of operations per second to be used in small satellite applications. Recent developments in the area of multichip module packaging have shown that the HWSI technology is well suited to highly complex circuits and provides maximum packing density of VLSI chips [5-7].

The approach we will utilize for the SCC-100 node is a silcon-based HWSI technology. The approach employs bare IC chips directly attached to interconnection substrates that are fabricated from silicon wafers. The technology is a true hybrid approach in that multiple device technologies may be mixed on the same substrate (e.g., MOS and bipolar, silicon and GaAs) as well as discrete components (resistors, capacitors, etc.). In our present design only low-power CMOS ICs are employed to minimize system power consumption. In addition to providing the interconnections among devices, the silicon substrate, by way of its high thermal conductivity, can be employed as part of the heat transfer path from the chips to the heat sink. Another advantage of silicon substrate technologies is that crystalline silicon wafers can be highly doped to the point that they are conductive with only a slight resistive loss through the thickness of the substrate. This conductivity allows the construction of an integrated capacitor within the substrate to decouple I/O signals from the power and ground structure, resulting in cleaner signal waveforms.

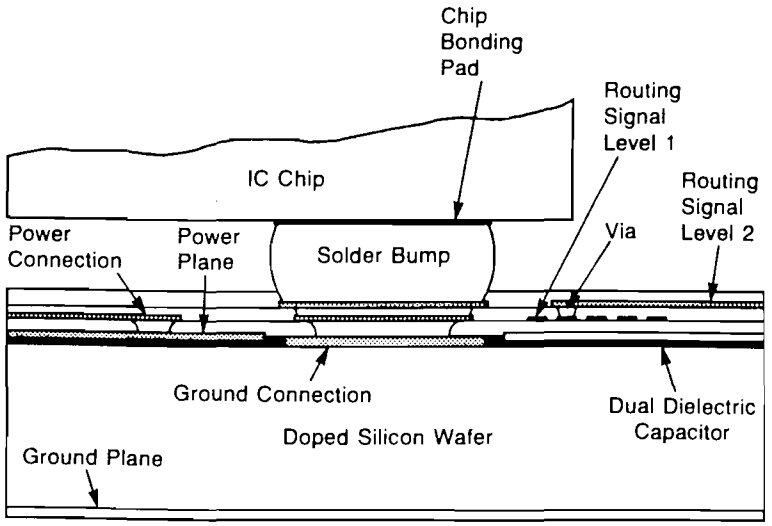

Figure 3. Crossection of HWSI Substrate with Flip-Chip Bonded IC.

The basic silicon substrate structure with a flip-chip bonded IC is shown in cross-section in Figure 3. The substrate starts out as a highly-doped 20 mil thick crystalline Si wafer. SiO_2 and Si_3N_4 are deposited on the top side of the silicon to form a pin-hole-free dielectric that will provide an integrated decoupling capacitor within the substrate. Both sides are then plated with copper to form a power plane on the top side of the substrate and a ground plane on the bottom side. The nominal thickness for the power/ground planes is 2 microns. Connections from the chip side of the wafer to the ground plane below are made by patterning openings in the metal power plane and etching through the dual dielectric. Signal interconnect levels are created by spinning on polyimide to form the dielectric layer above which copper conductors are plated to thicknesses of 2 microns. Standard interconnect line widths are 20 microns with 25 micron line-to-line spaces. Additional signal levels are added in

the same manner by spinning on polyimide and plating Cu lines. Vias between metallization levels are photodefined, etched and plated with solid nickel, forming planar surfaces within the substrate stackup. After the last metallization layer is fabricated, another polyimide layer is added to provide a level of surface protection to the interconnects below. This "cap" level of polyimide is opened at the bonding pads for chip I/O attachment, either flip-chip solder bonding, wire-bonding or tape automated bonding (TAB).

HWSI technology provides an excellent electrical medium for the interconnection of high-speed electronic devices. With extremely high interconnect capability, metallized signal paths between devices can be made very short. This feature, coupled with a low dielectric constant material (typically 3.5 for polyimide), ensures very low propagation delays across the substrate. Electrical parasitics can also be made very low, with typical line capacitances ranging from 1.0 to 1.8 pF/cm and typical line inductances ranging from 2.0 to 3.5 nH/cm. Characteristic impedance can be easily tailored to the application in the range of 30 to 100 ohms by varying line width and dielectric thickness. With proper design techniques, HWSI modules can be built to support systems operating at synchronous frequencies of 100 MHz and beyond.

4.3 HWSI Node Design

Based on the modularity of the SCC-100 design, a processing node was chosen as the functional unit to be designed as an HWSI substrate. A single processing node is an ideal application for the HWSI module in that it encompasses a large amount of circuitry, yet has a relatively simple (and low I/O) interface to other processing elements. The number of interconnects on the HWSI is approximately 1400, which can be accommodated with the design rules of 20 micron wide lines and 25 micron wide spaces, while the off-substrate I/O is less than 200. The HWSI substrate interconnects 23 large ICs including the three vector signal processors, the housekeeping processor, the two communication ASICs and the memory ICs. Additional decoupling capacitance is provided through the use of discrete 0.1 microfarad capacitors attached to the HWSI substrate. The placement of the components on the substrate is shown in Figure 4.

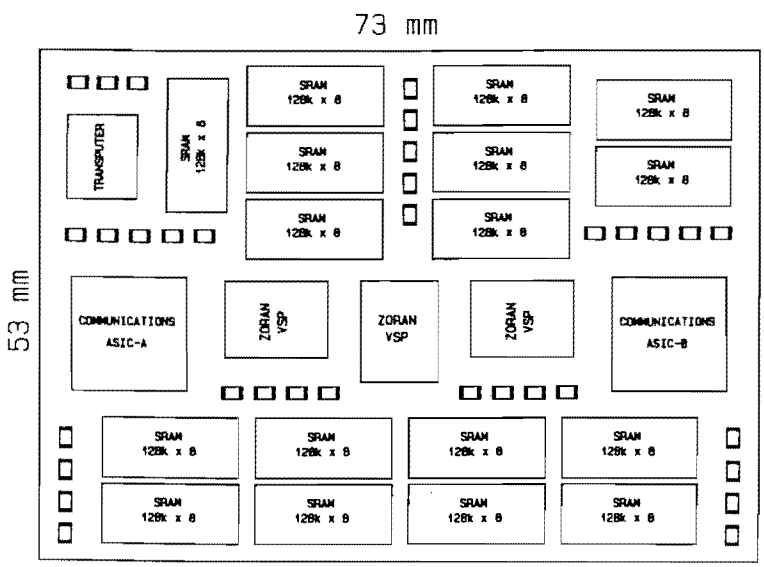

Figure 4. Component Placement on HWSI Substrate.

In order to reduce the size of the substrate, flip-chip bonding of the ICs was considered. Flip-chip technology provides the maximum packing density on HWSI substrates by placing ICs as close to one another as is physically possible. Conventional peripheral approaches such as wirebond and TAB require a fairly large area surrounding the chip to make connections to the substrate. Typically a 2.5 to 3 mm spacing is necessary between chips for the wirebonding tip or TAB bonding tool to gain access to the substrate. Flip-chip spacing is mainly limited by the accuracy of the robotic assembly tool that aligns the chip to the substrate. Chip-to-chip spacings as low as 0.5 mm are achievable with existing robotic assembly technology and have been demonstrated [5]. In addition to the density advantage, flip-chip bonding has superior electrical performance to conventional attachment techniques. The parasitic inductance of flip-chip solder bumps is under 5% of the inductance of either wirebonds or TAB leads due to the solder bump's inherently smaller size (50 micron length compared to 1000 micron length). Since common switching noise is directly proportional to the I/O inductance, flip-chip technology greatly reduces the effect of multiple signal lines switching simultaneously on system performance.

The difficulties with using flip-chip technology are the acquisition of bare die wafers for solder processing, the cost of post-processing the bare die wafers and the fabrication of fixtures for testing die in bare wafer form. Since the SCC-100 substrate employs 17 memory ICs, a tradeoff between substrate size and overall cost was made by utilizing flip-chip bonding only for the memory devices, and wirebonding the other six ICs. This resulted in a substrate measuring 53 mm by 73 mm for the processing node.

4.4 HWSI Package Design

The HWSI substrate can be packaged in a number of ways. Custom single-substrate packages can be designed and fabricated in commercially available technologies such as multilayer plastic or multilayer ceramic. Substrates are mounted into the package cavity face up and are wire-bonded to the package I/O pads. Most of these packages allow for heat sinks to be bonded directly to the substrate or provide for heat transfer from the substrate to the heat sink through metallized thermal vias in the package. Most high I/O (greater than 300) package designs are pinned for mounting either in zero-insertion-force (ZIF) sockets or directly onto printed circuit boards.

To minimize the physical size and weight of the system, as well as reduce interconnect length between HWSI modules, a three-dimensional interconnect structure is required. It is this packaging approach that is being pursued for space applications, where physical volume and weight allocations are at a premium. Such 3-D structures employ special high-density flexible connector mechanisms to interconnect the HWSI substrates. The use of these connectors allows the separation of individual substrates from the 3-D stack for testing and/or replacement. Individual substrates can be wirebonded to a holder made from printed circuit board material that has a rectangular section routed out to accommodate the substrate. These holders can then be interconnected together through a flexible elastomeric material as shown in Figure 5. The commercially available elastomeric material conducts in the Z-direction and not in the X- or Y-directions. The elastomeric material is held in place between the printed circuit holders by placing the entire stacked holder assembly in compression. This compression is easily achieved by bolts that pass through all the substrate holders with alignment provided by either the bolts themselves or additional alignment pins. The resulting assembly can be enclosed in a housing with I/O connections from the stack made through flexible ribbon cables.

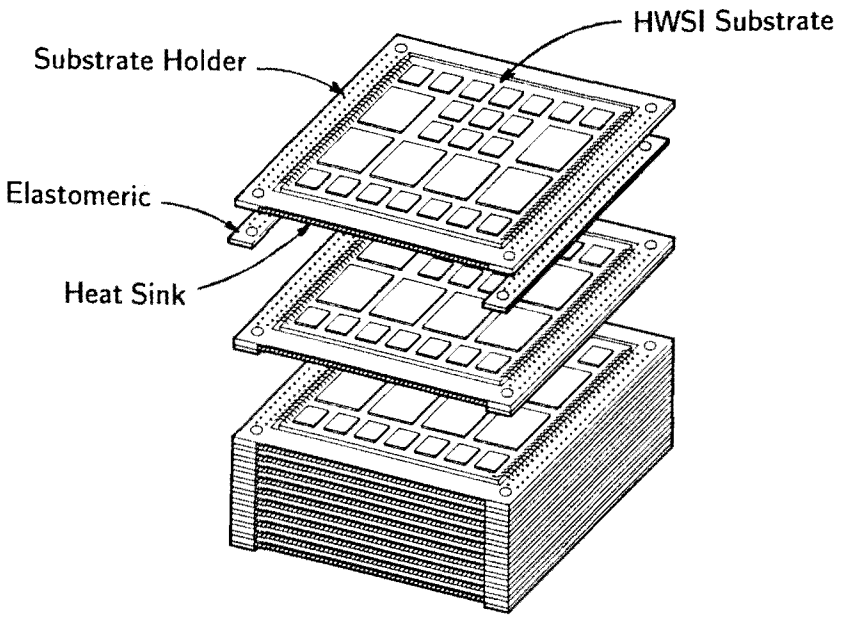

Figure 5. 3-D Stack Technology.

A three-dimensional stack implementation of the SCC-100 system comprising 12 nodes with a performance greater than 1.2 GFLOPS will occupy a volume of 15 cubic inches. The weight of the 12-node system excluding its enclosure will be 0.8 pounds. If a 300 mil thick aluminum enclosure is employed for reduction of total dose radiation at the device level, the system volume increases to 35 cubic inches and the system weight becomes 2.5 pounds. The total power dissipation for this system based on the current node design is 60 watts. A 12-node system can be further reduced in size by reducing the chip count on the substrate through the design of custom processing chips and higher density memories. Excluding the aluminum enclosure, the estimated volume, weight and power for such a system are 5 cubic inches, 0.3 pounds and 36 watts, respectively.

5. SUMMARY

A miniaturized low-power high-performance parallel processor optimized for space-based sensor systems has been developed. In particular, the processor was designed for image processing applications requiring throughputs of hundreds of millions to billions of operations per second. The processor, consisting of 12 nodes in a parallel pipeline architecture, delivers a peak throughput in excess of 1.2 GFLOPS. A key goal of the effort has been the reduction of weight, size and power consumption by an order of magnitude or more. The use of three-dimensional HWSI technology enables the system to be packaged in less than 15 cubic inches, weigh less than one pound and consume only 60 watts, achieving a throughput/power ratio greater than 20 MFLOPS/watt. Future work will focus on system designs that achieve greater computing power with substantially greater reductions in size, weight and power consumption.

6. ACKNOWLEDGEMENTS

We wish to acknowledge the support of Col. Ed Nicastri and Major Al Wheatley of DARPA. The work reported in this paper was funded in part under DARPA contract DACA76-90-C-0002, administered by the U.S. Army Engineer Topographic Laboratories.

7. REFERENCES

[1] L. Flynn, R. Bonometti, E. Nicastri, R. Sudol and A. Wheatley, "Creating the Future of Microspace Technology," *Proceedings of the 4th Annual AIAA/USU Conference on Small Satellites*, Logan, Utah (August 27-30, 1990).

[2] B. Raab, "Modular Small Satellite Design for Responsive Tactical Applications," *Proceedings of the 4th Annual AIAA/USU Conference on Small Satellites*, Logan, Utah (August 27-30, 1990).

[3] T. F. Utsch, T. H. Hopp, A. H. Sarrafian and S. Kilston, "Design Concepts for Space-Borne Multi-Mission Sensors for Tactical Military Needs," *Proceedings of the 4th Annual AIAA/USU Conference on Small Satellites*, Logan, Utah (August 27-30, 1990). Originally published in *Proceedings of the Society of Photo-Optical Instrumentation Engineers, Infrared Technology XVI*, Vol. 1341, July 1990.

[4] J. Farrell, "Reserves: A Responsive, Multi-Mission Tactical Satellite Architecture," *Proceedings of the 4th Annual AIAA/USU Conference on Small Satellites*, Logan, Utah (August 27-30, 1990).

[5] C.J. Bartlett, J. M. Segelken and N. Teneketges, "Multichip Packaging Design for VLSI-Based Systems," *IEEE Trans., CHMT-10*: pp. 647-653, December 1987.

[6] C. L. Chao, et al., "Multilayer Thin-Film Substrate for Multichip Packaging," *Proceedings of the 38th Electronic Components Conference*, pp. 276-281, 1988.

[7] J. Reche, "Fabrication of High Density Multichip Modules," *IEEE Trans., CHMT-13*: pp. 565-569, September 1990.