

MONARCH: A First Generation Polymorphic Computing Processor

Michael Vahey (Raytheon), John Granacki (USC-ISI), Lloyd Lewins (Raytheon), Drew Davidoff (Raytheon), Jeff Draper (USC-ISI), Craig Steele (Exogi), Gillian Groves (Raytheon), Matt Kramer (Raytheon), Jeff LaCoss (USC-ISI), Kenneth Prager (Raytheon), Jim Kulp (Mercury), Charles Channell (Raytheon)
mdvahey@raytheon.com, llewins@raytheon.com, granacki@isi.edu, draper@isi.edu

MONARCH Processor Vision

Computing workload continues to fill available computing capacity. Algorithms grow in complexity, I/O rates increase, number of I/O channels grows, and size of images or data bases grow to fill available memory. If only the available size, weight, power, and cost would shrink as fast as demand grows – but they don't, thus driving the need for sustaining and even disruptive innovation in all processing technologies.

Even as commercial processors have increased in processing performance, the amount of data provided by the sensor front ends in the analog-to-digital interface has placed an even greater stress on the interface to the processor, demanding wider and wider bandwidth of the input/output (I/O) ports.

Now imagine a highly integrated sensor suite with data from multiple sensors, Radar, EO/IR, SIGINT and Communications all fused together in a cognizant processor that adapts to the signal and threat environment. In this same suite, waveforms and algorithms rapidly morph and adapt without intervention from a human operator.

Future Requirements

While the next generations of processors for commercial products are addressing the necessary throughput for next-generation video processing, they do not adequately address efficiency for low power, high throughput computing and I/O interface requirements.

To address these future requirements and to further improve the processing throughput of modern Radar, E/O, Missile, Communications and SIGINT systems the MONARCH (morphable networked micro-architecture) processor has been developed, adhering to a polymorphic computing architecture (PCA). The MONARCH technology is differentiated from current and near-term commercial processors by the very nature of its architecture. Key attributes that differentiate MONARCH include:

1. A distributed processor architecture making the chip highly scalable and fault tolerant
2. Distributed inter-node communication for higher processing bandwidth and resource utilization
3. The ability to handle multiple computing models: Scalar Reduced Instruction Set Computer (RISC), Wide Word and Streaming
4. System-on-a-chip (SOC) computing minimizes cost and area associated with peripheral components and improves power and volume efficiency

5. Reconfigurable computational array (morphability), which dynamically tunes to meet application needs
6. Coupled with the highest efficiency (GFLOPS/Watt; GFLOPS/m³) processing power

Looking back 10 years

Ten years ago, parallel processing was beginning to displace the very high speed, pipelined custom processors used in DoD systems. Algorithms were beginning to take advantage of better programmability, floating point, and larger memories with Space Time Adaptive algorithms – allowing for better clutter rejection or operation in jammed environments. Input data rates were relatively low bandwidth by today's standards, being measured in 10's of millions samples per second. Programming was tied to vendor specific communication and math library functions, as the open methods like MPI and VSIP were embryonic.

2006 processing needs

By comparison, today parallel processing is omnipresent. Algorithm developers develop with a 64-bit Matlab floating-point assumption and presume that the deployed, embedded hardware will be able to support their design. Adaptive time, space, and frequency algorithms are pervasive. The front-end digital processing has moved forward with input data rates being small number of GBytes/s per channel with possibly dozens of channels. This has increased the computing needs to small numbers of TeraFlops of throughput, balanced with appropriate memory and I/O bandwidth to support raw data input, corner turns, and down linked processed data.

2006 processors vs. earlier architectures

MONARCH was conceived and designed to meet the high efficiency TeraFlop computing needs. It uses the largest 90 nm ASIC standard image that IBM offers measuring a full 18.76x18.76mm. It is a member of the *1 kilometer club* with more than 1.5 km of on-chip wiring. It has more than 280 million placeable cells.

Comparing to the collection of i860s or PowerPCs used in the past, MONARCH has moved onto a single chip what used to require more than a chassis to accomplish. The i860 used to be estimated at 80 MFLOPS peak throughput. MONARCH throughput is 64 GFLOPS peak and achieves >60 GFLOPS on balanced algorithms like FIR filters. Previous boards with Raceway had bandwidth of 160 MBytes/s versus MONARCH with an external I/O bandwidth of 43 GBytes/s per chip.

MONARCH Architecture

The MONARCH architecture is shown in Figure 1. :

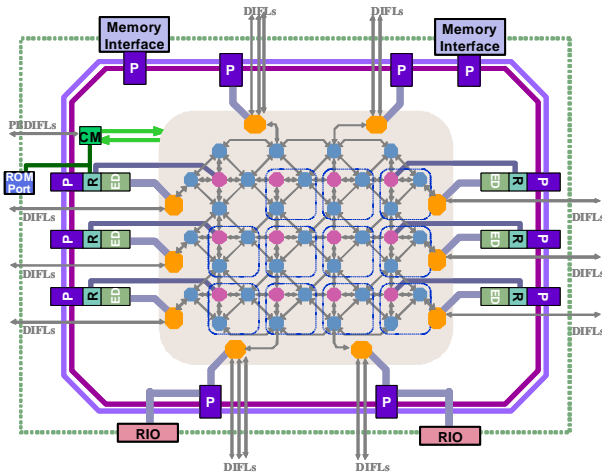


Figure 1: Monarch Chip Architecture.

MONARCH uniquely combines multiple processing paradigms into a single morphable chip – that shares the underlying hardware ingredients to build at run time a chip tuned to the mission software needs [1]. Each MONARCH chip is a self contained heterogeneous parallel processor. It contains six simple 32-bit RISC processors, each with an associated 2 MByte DRAM; a Field Programmable Computer Array (FPCA); and a high bandwidth interconnect. Each RISC has a conventional five stage pipeline, instruction cache, timers, address translation unit, and exception unit. Unusually, the RISC has no data cache, but instead relies on a simple 8 word load/store buffer and very high bandwidth to the on-chip DRAM [2].

The FPCA is key to the MONARCH chip's high performance, efficiency and morphability. The FPCA contains 96 Multiplier-ALUs, 124 dual-port memories, 248 address generators, and 20 DMA engines all connected through a rich, dynamically switched, interconnect. The architecture of the FPCA has been optimized for signal processing algorithms, for example FFT and FIR, using 16 and 32-bit integer, and 32-bit IEEE floating data. The FPCA supports streaming data with hardware support for dataflow synchronization, and uses a novel distributed programming paradigm. The FPCA is currently programmable in assembly code, but development of a higher level compiler is underway. The resources of the FPCA can also be flexibly morphed into six conventional 256 bit wide SIMD processors each tightly coupled to a respective RISC processor (similar to Altivec). Each SIMD engine supports 8, 16 and 32-bit integer data and 32-bit IEEE floating point data. The morphability of the FPCA allows the machine to adapt to the needs of the algorithms running on it.

Finally, the chip includes two high bandwidth off-chip DDR2 DRAM ports, including Reed-Solomon error correction; two 4-lane serial RapidIO ports; and sixteen full duplex High Speed Serial (HSS) links to form a distributed interconnect fabric for TeraFlop computing and direct connect to multiple GByte/s I/O devices

The MONARCH Field Programmable Compute Array (FPCA) contains 12 Arithmetic Clusters and 31 Memory Clusters. Each cluster is connected to its neighbors through a very high bandwidth non-blocking interconnect. Each Arithmetic Cluster provides 8 Multiply-ALU blocks (integer and floating point) and 8 Shift/Saturate Compare (SSC) logic blocks (total of 96 Multiply-ALU, 96 SSC). Each Memory Cluster provides 4 dual-port 256x32 memory blocks (configurable to adjust number of memories and word depth) and 4 address generator pairs. Additionally, the FPCA provides 10 IO Clusters. These IO Clusters provide 16 full duplex High links. Each link provides 4 channels, each transmitting/receiving 8b10b encoded data, for a logical link transmit/receive of 32-bits. Each logical link provides approximately 2.6 GBytes/s bandwidth (1.3 GBytes/s for each transmit and receive link). The total IO bandwidth via the MONARCH HSS links is approximately 43 GBytes/s. Attached to the FPCA, MONARCH provides 12 Node Bus Ring connection networks connected via the Packet Buffer's (PBUF) PIRX network. The Node interconnect itself is a high bandwidth point-to-point bi-directional ring architecture providing 43 GBytes/s bandwidth at each device, and protected global accessibility for all connected resources.

The MONARCH chip is being developed under contract to DARPA IPTO (Defense Advanced Research Projects Agency Information Processing Technology Office) and under the auspices of their PCA Program administered by AFRL Wright Patterson. Raytheon is the prime contractor on the XMONARCH development, and is partnered with USC/ISI, Exogi, Mercury Computers, IBM and Georgia Tech for specific intellectual property. IBM will fabricate the chip in their Cu08 (90nm) CMOS ASIC process and Mercury will build a prototype module and processor system for demonstration programs.

Looking forward – 2016 needs

Disruptive technologies like field programmable compute arrays will be essential to achieving PETAFLIPS of throughput with acceptable power. Even greater emphasis will need to be placed on reducing power per op, providing on-chip security, and dealing with single event transients. Algorithms and programming models will include more focus on cognitive control for robustly stable processing for mission assurance as ubiquitous, totally networked computing will be at the core of all defense systems.

References

- [1] John J. Granacki and Michael D. Vahey: "MONARCH: A High Performance Embedded Processor Architecture With Two Native Computing Modes", In Proceedings of High Performance Embedded Computing Conference, March 2002.
- [2] Jeff Draper, et al, "The Architecture of the DIVA Processing-In-Memory Chip", Proceedings of 16th ACM-SIGARCH International Conference on Supercomputing, June 2002: