

MNOC: A Network-on-Chip for Monitors

S. Madduri, R. Vadlamani, W. Burleson, and R. Tessier
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst

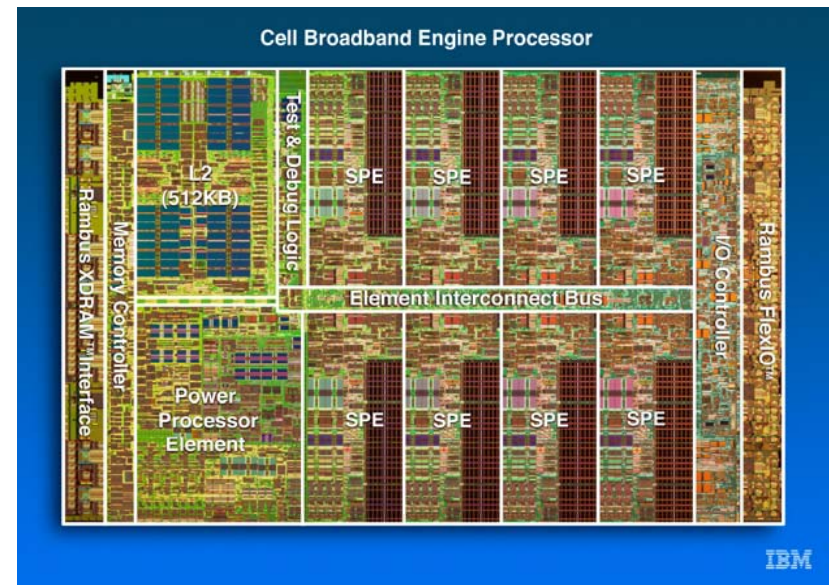
Outline

- Need for integrated monitors and system response
- Our Approach
 - Collections of integrated monitors
- Challenges for integrated monitors
 - Processing and use of monitor data
 - Monitor interconnection
- Preliminary results
 - Thermal data and critical code section correlation
- Current work

Proliferation of Embedded SoC

- Increased use of large multi-core systems
- Recent growth in use of real-time monitoring on-chip
- The interaction between cores has become a large concern
- Growth of multicores pertains to both FPGA and Reconfigurable SoC
- How can we make sure these systems operate correctly?

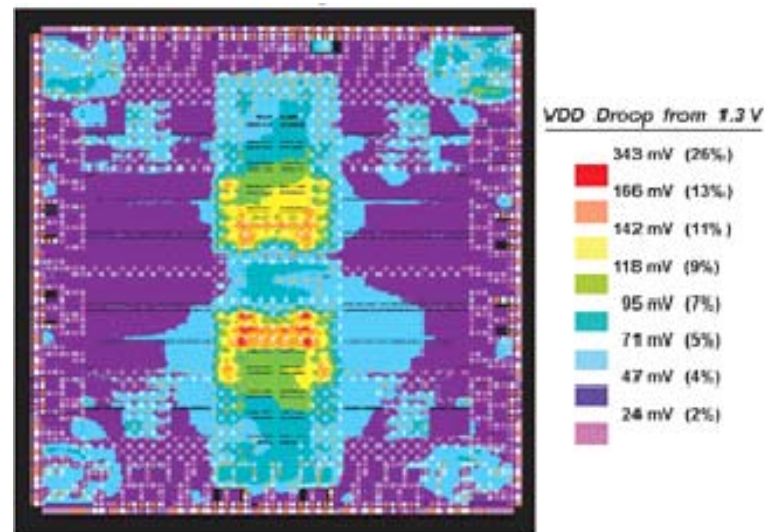
IBM Cell 9 core



Courtesy: IBM

Need for Integrated On-chip Monitoring

- Architectural monitors are important components of many SoCs.
- Specific uses of monitors in terms of soft-error failures, wear-out detection and security issues.
- Monitor interconnection may need to be some of the lowest-latency, cross-chip connections in a system-on-chip (SoC).
- Support for monitors and associated interconnect must be lightweight and consume minimal system resources.
- Collected data must be processed collaboratively.
- Goal: Develop integrated environment to collect, transport, and process monitor data for SoCs



Chip voltage droop profile for 223 million transistor ASIC

Courtesy: <http://www..mayo.edu>

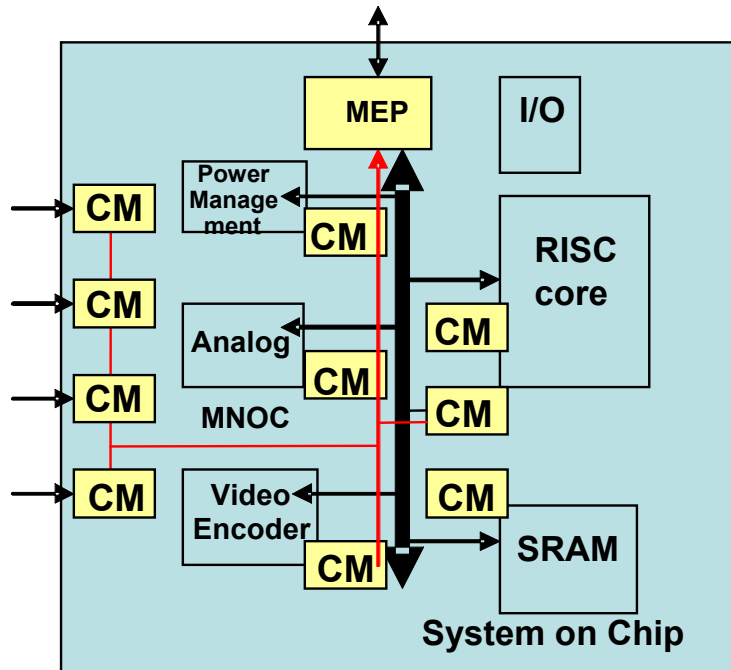
Survey of existing on-chip monitors

Sensor	Typical Bandwidth	Typical Power Dissipation	Typical Area/ Process width
Thermal monitors	20Kbps	2 μ W - 220 μ W	0.01mm ² /0.18 – 0.5 mm ² /1 μ m
Wear-out detectors	1Mbps	1mW	0.15% of chip area
Soft-error monitors	Very low and variable- based on Soft Error Rate- once in 5hrs – once in 28 yrs	7% of chip power	5- 7% of chip area
Processor performance monitors using performance counters (branch prediction, cache monitors, etc)	PC profiling rate – 4Kbps Cache monitor sampling-10Mbps Branch prediction : 0.1Mbps	Low	2-6 counters on chip, 2-4% of program execution time overhead
Intrusion detection (Processing monitor)	Low	Very Low	Overhead of around 10% in terms of memory
Jitter monitor	5 Mbps	Low	0.05 mm ² (0.25 μ technology)

Current Limitations of Monitor Usage

- Difficult to configure monitors
- Difficult to interconnect monitors and collect data
- Monitor data interacts with main computing system in an ad hoc fashion
 - Non-standard interfaces
- Difficult to ensure monitor data is secure
 - An issue for both FPGAs and ASIC SoCs
 - Includes external interfaces
- Diversity of monitors makes standardization difficult
 - Likely to become a bigger issue in the near future

Secure System Level



**CM = Configurable
Monitor**

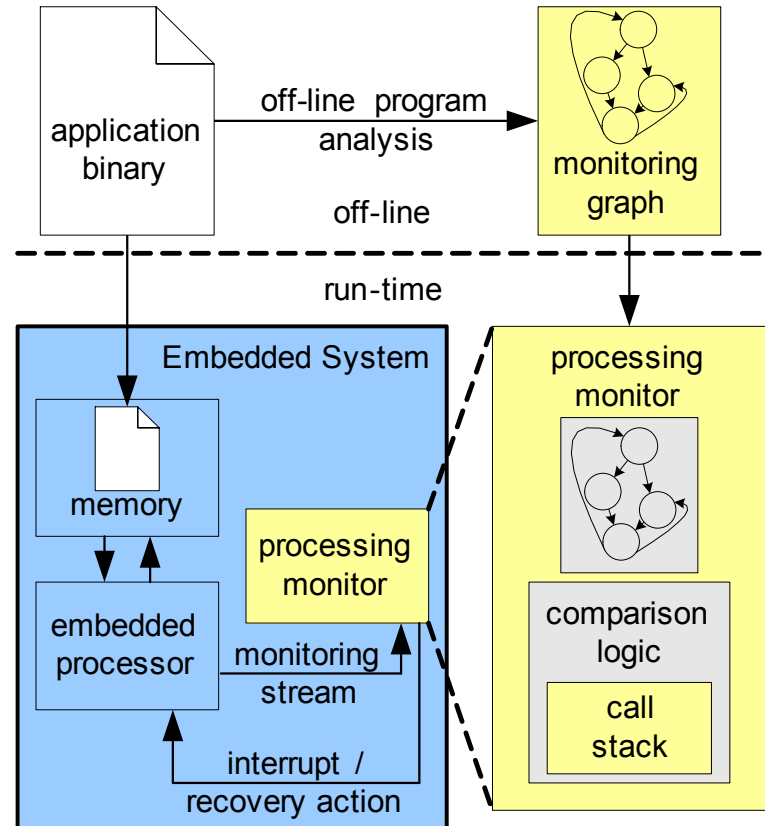
**MNOc = Monitor
Network on Chip**

**MEP – Monitor
Executive Processor**

- Configurable computing is seen as a system
 - Monitoring of the activity of the system to detect irregular sequence of computation
 - Dynamic reconfiguration of the critical parts of the system

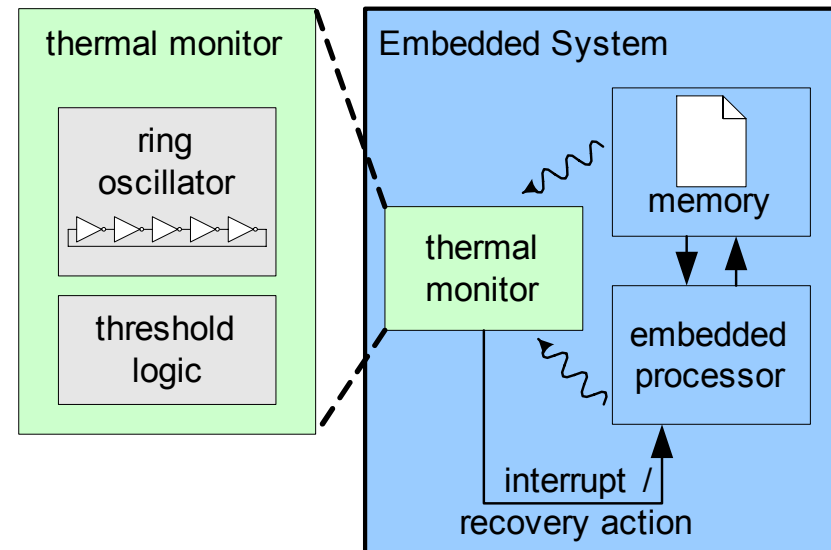
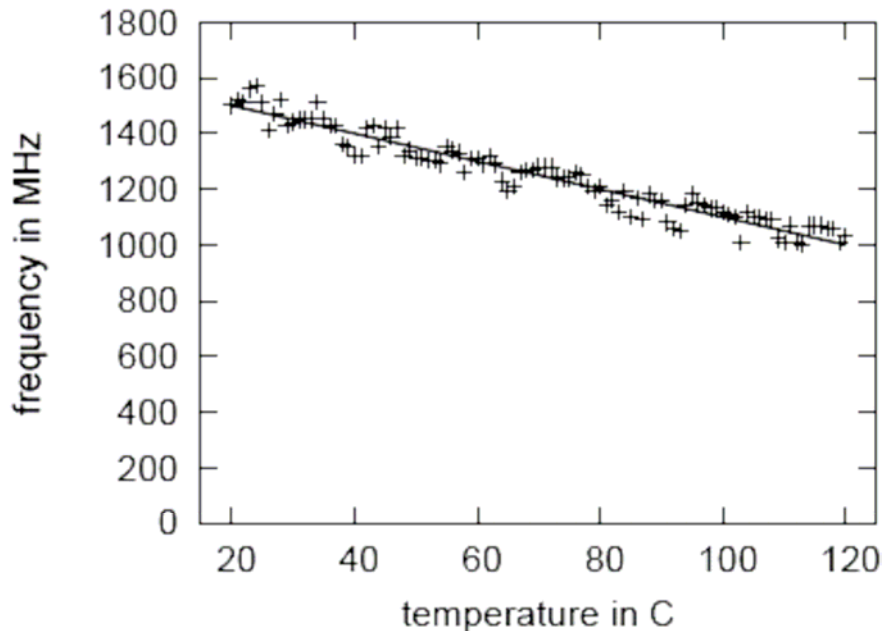
Processing Monitor

- Offline binary analysis
 - Monitoring graph extraction
- Online validation of processing
 - Information stream from processor
 - Comparison to monitoring graph
 - Requires call stack for returns
 - Interrupt/recovery on deviation
- Choices on what to monitor
 - Address
 - Vulnerable to code replacement
 - Opcode
 - Vulnerable to changes in registers
 - Control flow
 - Vulnerable to code replacement within basic block



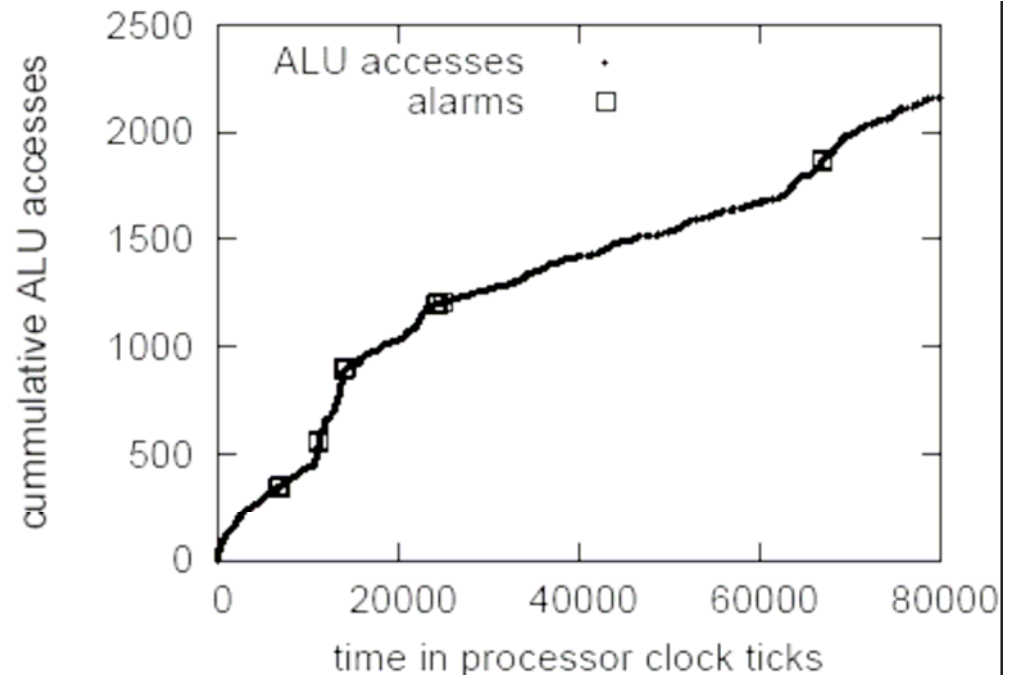
Thermal Monitor

- Ring oscillator
 - Odd number of inverters in loop
 - Delay across inverter is temperature dependent
 - Larger number of inverters mitigates power supply noise



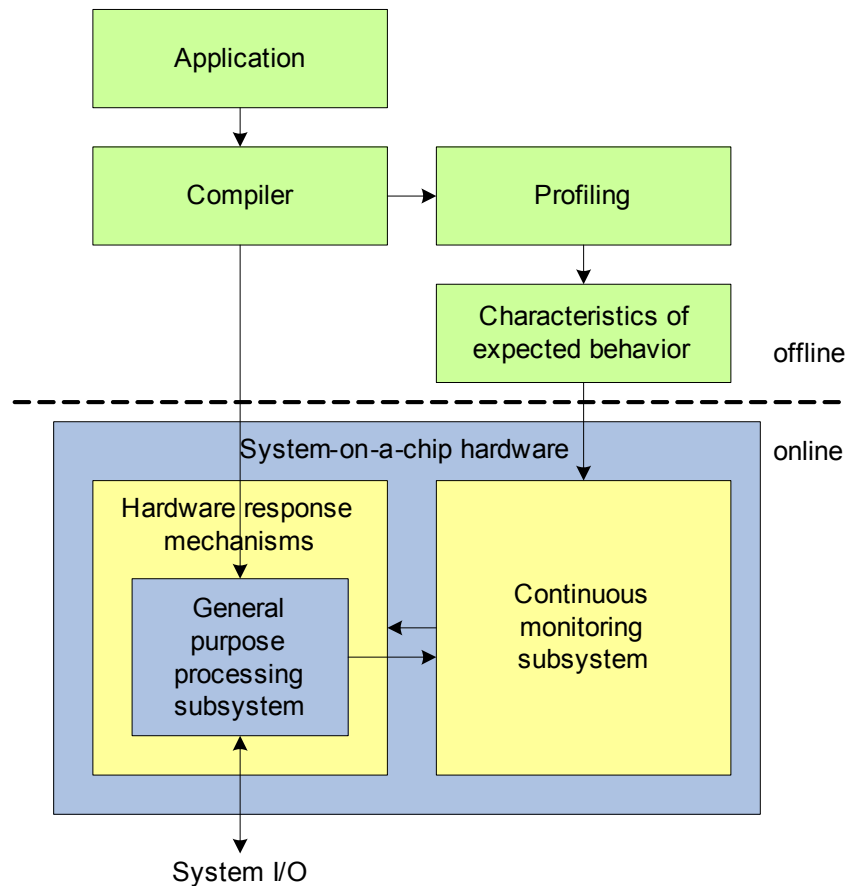
Thermal Monitor

- Hardware access counters
 - Tracks hardware utilization over time
 - High utilization in short interval can cause thermal problem
- Implementation overhead
 - Ring oscillator
 - Less than 0.5% of processor area
 - Less than 0.5% of processor power
 - Hardware access counter
 - Access counter
 - Sampling possible



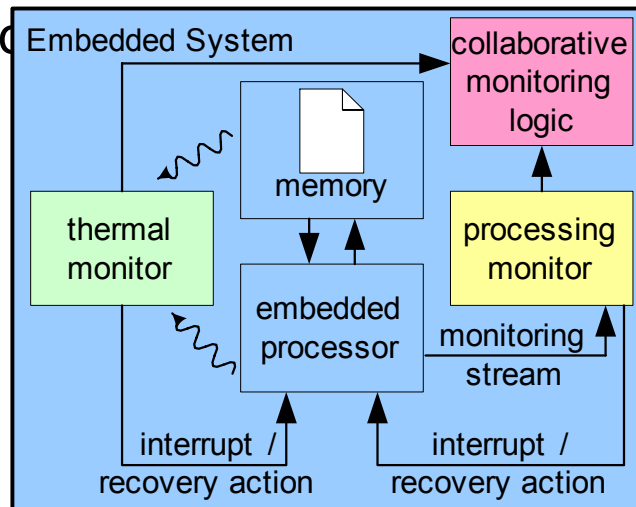
Collaborative Monitoring

- Information from different monitors can be correlated
 - Alarms based on joint information



Collaborative Monitoring

- Sensitivity tradeoff for individual monitors
 - High sensitivity trigger many false positives
- Monitors can work together to better assess data
- Example: processing / thermal monitor
 - Code regions with high access frequencies can be identified
 - Thermal threshold of time



Interconnected Monitors Design Issues

- Diversity of monitors likely to increase in coming years
- Monitor interfaces currently ad hoc.
- Performance constraints motivate separating monitor/main processing network
- Monitor interconnection criterion:
 - 1. low latency**
 - 2. low overhead**
- Need to aggregate, collate, and react to monitor data quickly
- Monitors need to be dynamically configured based on changing system requirements
- Global controller (possibly distributed) needed to coordinate monitor data collection and reactions

Monitor Interconnect Design Challenges

- Minimize interconnect latency
 - Difficult to pre-allocate monitor bandwidth
 - Difficult to tradeoff latency, area, and power limitations
 - Circuit level techniques needed
- Provide flexible interface for monitors
 - Difficult to generalize since monitors are diverse
- Minimize resources needed for monitor control/data processing
- Maintain effective response to monitor data

Distinction from current approaches

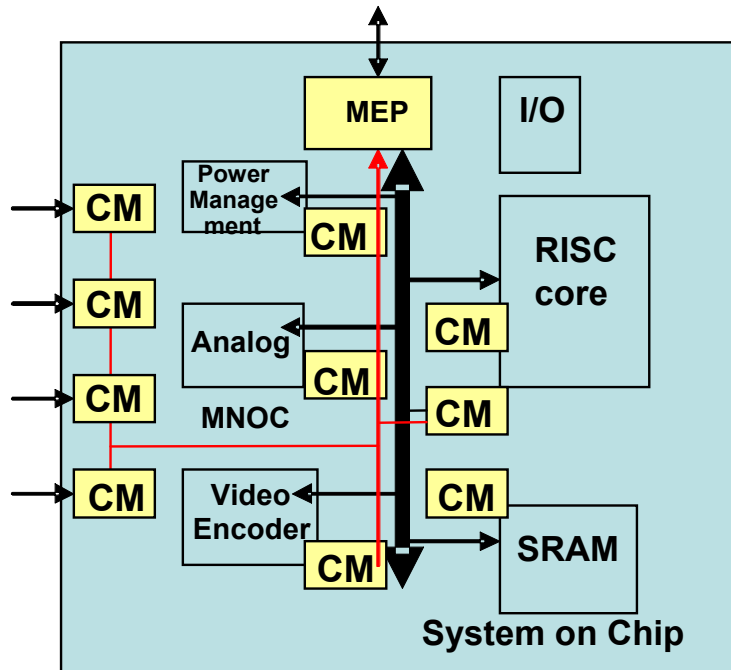
- Unlike NOCs designed for data transfer that typically use L2 caches or more exotic ring caches, MNOC will prioritize *latency* rather than bandwidth
- Develop protocols required for inter-monitor transfer based on dynamic system priorities.
- Study the impact of different sensor interface designs and corresponding responses at various time scales
- Architectural design and VLSI implementation of several CMs, the interconnect, and the controller along with monitors for power, voltage, and processor usage.
- Lightweight & robust monitor interconnect and interfaces

Types of On-chip Interconnect

- Current on-chip interconnect approaches range from buses to networks-on-chip
- MNOC will build on existing technologies with optimizations for latency, low overhead

Interconnect type	Characteristics	Possible issues/optimizations
On-chip bus	<ol style="list-style-type: none"> 1. Simple 2. Limited scalability 	<ol style="list-style-type: none"> 1. Probably not appropriate for MNOC
Statically-scheduled network	<ol style="list-style-type: none"> 1. Predictable performance 2. Difficult to prioritize transfer 	<ol style="list-style-type: none"> 1. Could be integrated with virtual channel dynamic network
Single channel dynamic network	<ol style="list-style-type: none"> 1. Low overhead 2. Straightforward interface 	<ol style="list-style-type: none"> 1. Difficult to prioritize latency versus bandwidth
Virtual channel dynamic network	<ol style="list-style-type: none"> 1. Allows for priority tradeoffs 2. More resource overhead 	<ol style="list-style-type: none"> 1. Allows for latency/throughput tradeoffs 2. Parameterize based on selected monitors 3. Combine with circuit level optimizations

Securing the Monitor Gateway



**CM = Configurable
Monitor**

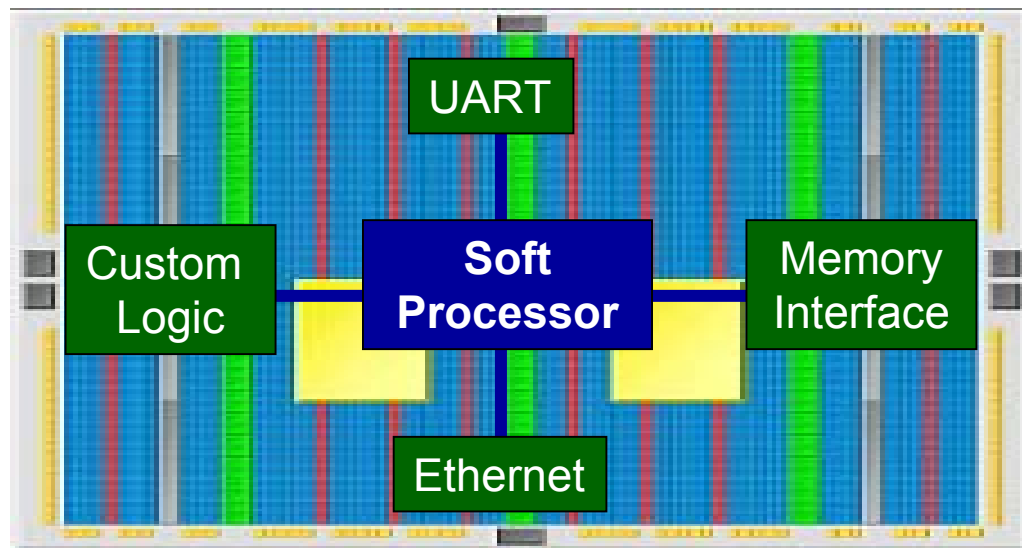
**MNOc = Monitor
Network on Chip**

**MEP – Monitor
Executive Processor**

- External user needs capability to access monitor information
 - Low overhead security needed
 - Dynamic reconfiguration of the critical parts of the security provides flexibility
- Gateway also needs to provide control for monitor reconfiguration
- Complexity of monitors makes this issue a challenge

Secure Reconfigurable Processors

- Soft processors provide a special opportunity for monitoring



- How can we ensure monitor usage is secure?
 - Data encrypted for gateway
 - Tampering either evident or not possible

Next Steps

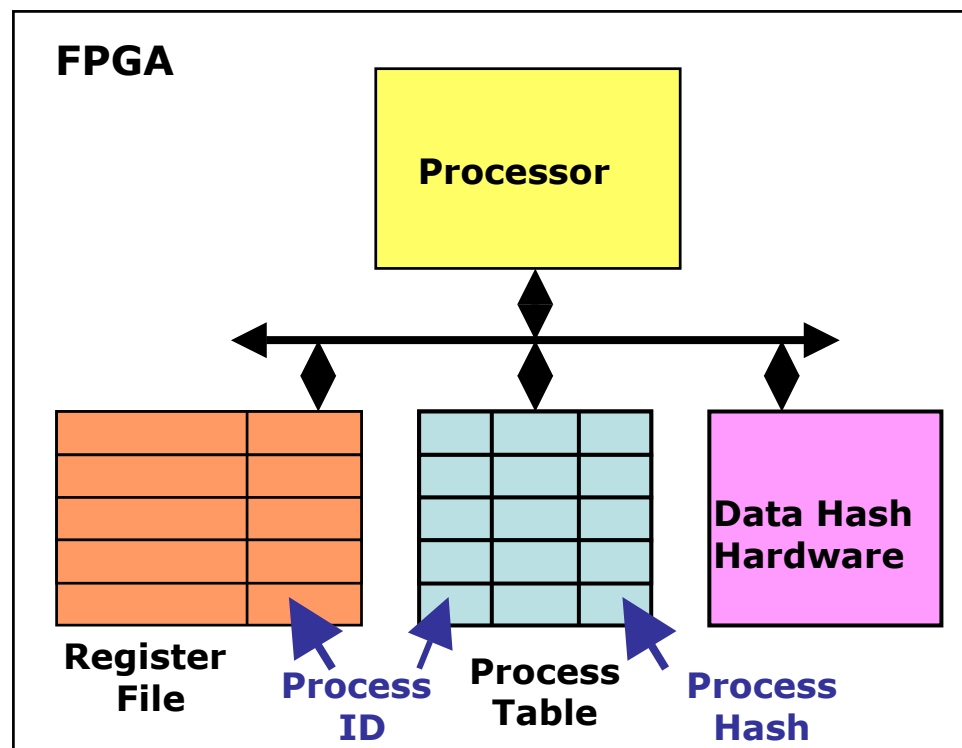
- Architecture development for monitor network on chip currently under way
- Interface to monitors and appropriate analog techniques being evaluated
- Continued assessment of diverse monitors
- Architecture and interface for the monitor executive processor
 - Encryption interface needs to be defined
- High-level modeling and use of monitors to be evaluated

Conclusion

- Need for better use and implementation of collections of on-chip monitors
- Previous interconnect approaches are insufficient. Need for new approaches
- Monitors can work together to better assess processing
- Monitor executive processor needed to coordinate use of monitor data
- Project under way

Secure Context Swap for FPGA Processors

- Hardware for secure context swaps another security enhancement
- Follow previous examples from secure processors (e.g. Aegis, XOM)
- Memory and logic availability in FPGAs simplify implementation



Monitoring Graph Example

- Example
 - MiBench on SimpleScalar simulator
- Monitoring graph
 - Chained basic blocks
 - Different information within basic blocks

sample object code	monitoring graph		
	address	opcode	control flow
...
020004d0 str r0, [sp]	020004d0	str	*
020004d4 str r0, [sp, #4]	020004d4	str	*
020004d8 ldr r1, [pc, #1c4]	020004d8	ldr	*
020004dc sub r4, r11, #2080	020004dc	sub	*
020004e0 ldr r3, [pc, #1c0]	020004e0	ldr	*
020004e4 sub r4, r4, #8 ; 0x8	020004e4	sub	*
020004e8 ldr r2, [r11, -#2136]	020004e8	ldr	*
020004ec mov r0, r4	020004ec	mov	*
020004f0 bl 02091aa0	020004f0	bl	bl 02091aa0
020004f4 mov r0, r4	020004f4	mov	*
020004f8 mov r1, #0 ; 0x0	020004f8	mov	*
020004fc bl 020905dc	020004fc	bl	bl 020905dc
...