RAW computation – A survey

Sudhindra Rao

Outline

- Motivation
- Contemporary techniques
  - FPGA, Numesh
- Raw
- Compiler for Raw – Maps
- Raw design
- Performance
- Further reading and applications
Motivation

- Reconfigurable computing
- Billions of transistors on chip by 2010
- Numesh, iWarp and FPGA
- Configurable silicon for applications
- Another idea - Raw

Configurable silicon

- Small memory
- Virtual wires
- Custom logic
- Leads to faster processing of customised applications with less power consumption
- FPGA – clock frequencies?
Numesh communication pattern and architecture

- Modular with high performance communication
- Reduce runtime routing decisions
- Faster than other dynamic networks – as they are predictable

Raw architecture and comparison

- Simple replicated tile
- Programmable interconnect
- Multi-sequential statically-scheduled instruction streams
- Multi-granular operations
- Configurability
Memory structure

Optimising designs for Raw
Dynamic load

Memory access latencies
Maps – Compiler for Raw

- Single centralized memory system – Von Neumann bottleneck
- Raw distributes memory
- Maps uses communication mechanisms to generate efficient memory accesses for low latency and parallelism

Hot pages

- Predicts and selectively virtualizes memory references
- Virtualization is a purely software idea hence completely predictable
- Eliminates cache tag lookups – upto about 90%
- Instead uses register comparisons by reusing translated virtual page descriptions from earlier nearby memory references
Some results - Speedup

![Speedup Chart]

Figure 4: Base comparison

Chip area breakdown for various applications

![Chip Area Breakdown Diagram]

Figure 7: Breakdown of chip area for processing, memory, local communication and global communication that give optimal machine configurations for a budget of 1 trillion logic transistor equivalent area.
Further reading and applications

- Reconfigurable machines
- Reconfigurable Parallel machines

References: