The First Superscalar 29K™ Family Member

Brian McMinn

Hot Chips VII
August 14, 1995

Presentation Outline

- Product Overview
- Block Diagram
- Operation and Pipeline
- Instruction Tracing and Debug
- Relationship to K86™ Family
- Conclusion
Product Overview - Goals of Design

- Software compatible with existing 29K™ Family members
- Fast execution of existing binaries and use of existing compilers
  - Superscalar implementation
  - Out of order execution
  - Register renaming
  - Speculative execution across branches
- Constrained cost & power
- Hardware assist for software debug required
- Speeds to 100 MHz internal, 33 MHz external
- Pin compatible with AM29040™ Microprocessor

Product Overview - Family Features

- Large general purpose register set
  - 128 local registers in on board stack cache
  - 64 global registers
- Fast and predictable interrupts
  - Simple interrupts handled in freeze mode
  - Complex interrupts possible after saving some state
  - Exceptions taken in program order despite out of order execution
- Hardware support for big and little endian systems
- Dual MMU structure
  - Two TLB’s with one prioritized over the other
  - Each TLB has independent page size from 1 Kbyte to 16 Mbyte
  - Supports huge virtual space AND flexibility for small segments
Operation - Cache Load

- **8K Byte Instruction Cache**
  - One bit branch prediction cleared when block is first loaded
  - One valid bit per 4 words

- **Cache loaded from memory in 4-word blocks**
  - Block aligned fetch rather than word of interest first
  - Predecode logic encodes read/write count and size of result

- **On subsequent execution of block (cache hit)**
  - Branch prediction information may redirect fetch unit
  - Potentially speculative fetch and dispatch continues
  - Branch prediction information sent to decode
    » will be validated when branch is executed
Operation - Decode

- Decode cache block as a unit
- Allocate reorder buffer entries
  - Maintains architectural state
  - Eliminates false dependencies
  - Increases parallelism for non-superscalar code
- Check data dependencies, locate operands
  - Read from the register file
  - Forwarded from results not-yet-written stored in reorder buffer
  - Represented by a “tag” when not yet available
    » Based on the tag, the operand will be forwarded directly to the requiring function block when it becomes available
    » Use most recent value if more than one copy

Operation - Dispatch

- Dispatch instructions in order
- Allocate reservation station entry
  - Two reservation station entries per function block
  - FIFO allocation within each function block
  - Place instruction and operands (or tags) in reservation station
- Can be held by
  - One instruction per function unit per clock cycle
  - Reservation station of a unit may be full
  - Lack of register file read ports (4 ports implemented)
  - Reorder buffer full
  - Special cases which require serialization
Operation - Execute

- Reservation station entries serviced in FIFO order
- Wait for forwarded data (if required)
- Various function units are completely independent
  - Instructions may execute out of order
  - Single cycle latencies except for two cycle pipelined multiply
- Arbitrate for a result bus
  - 3 result busses available
  - execution will block if results can’t be returned
- Return both the result and the corresponding tag
- Tags compared in each reservation station for forwarding to next cycle

Operation - Writeback

- Drive result from function block to reorder buffer
- Based on previous tag compare result may be forwarded to other reservation stations
- Recover from incorrect branch predictions
  - Redirect instruction fetch
  - Update prediction information in cache
- Speculatively complete loads that hit in the data cache
Operation - Retire

- Results retired from Reorder Buffer in program order
- Up to 4 reorder buffer entries can be retired at once
  - Two writeback paths to register file allow two results to be written
  - One branch and one store can also be retired
- Handle exceptions in program order
- Update architectural program counter
  - Invalidate speculative results for incorrect branch prediction
- Release store operations to cache and load-miss operations to external interface

Operation - Loads and Stores

- 8 Kbyte Snooping Data Cache
  - Physical addressing, MOESI protocol, buffered copyback
- Load-hit is a single cycle operation
  - Completes immediately, but results considered speculative
- Load-hit can bypass deferred stores
  - Hardware dependency check prevents read after write conflict
- All stores and all loads that miss in the data cache wait for retirement of corresponding instruction
  - No speculative external data transactions
- Reorder buffer fills with subsequent results while waiting on external load or store
Superscalar Tracing - The Problem

- Desire to provide a full trace of internal instructions
- Don't want to
  - Perturb the system by slowing down the processor
  - Create a special bond out chip
  - Require ICE to track instruction cache contents
- During each external clock cycle, can retire
  - up to 16 instructions
  - up to 4 taken branch instructions
- Snooping data cache
  - Task list and control can be passed via snooping
  - There is no requirement to EVER initiate an external transaction!

Traceable Caching™ System
Traceable Caching™ System

- **Tracing Processor is off the shelf standard part**
  - Placed in trace mode during system reset
  - Runs in lockstep with Master Processor
  - Pins which Tracing Process does not need for input are redefined to contain information about internal state

- **Trace analysis hardware**
  - Captures internal information for display and analysis

- **Compression**
  - Trace is compressed to fit available bandwidth
  - No need to look at program text to interpret trace

- **Flexible on chip breakpoint hardware reduces need to trace data activity**

Relationship to K86™ Family

- **An early model of this product was the basis of the K86 Family**
  - Some of the same people contributed to both the architecture and the design of both chips

- **Targeted at much lower cost system**
  - Significantly lower transistor count and area
    - Provides integer multiply, but not floating point unit
    - No interaction between instruction and data caches
    - Much simpler instruction decode
  - Slightly lower performance
  - Much lower sticker price

- **Doesn’t boot Windows ‘95 (or even DOS)**
A Superscalar 29K Family Member

- Maintains compatibility with existing 29K compilers and executables
- Retains full in-order programmer's model even though execution can be out-of-order
- 100 MHz internal speed in 0.4 micron technology
- Low power consumption
- Designed for high performance with low sticker price rather than highest performance at any price