Chapter 9
Multiprocessor

Parallel Computers

• Definition: “A parallel computer is a collection of processing elements that cooperate and communicate to solve large problems fast.”

• Questions about parallel computers:
  – How large a collection?
  – How powerful are processing elements?
  – How do they cooperate and communicate?
  – How are data transmitted?
  – What type of interconnection?
  – Does it translate into performance?

Parallel Processors

• The dream of computer architects since 1950s: replicate processors to add performance vs. design a faster processor

• Led to innovative organization tied to particular programming models since “uniprocessors can’t keep going”

• Argument instead is the “pull” of opportunity of scalable performance, not the “push” of uniprocessor performance plateau?

What level Parallelism?

• Bit level parallelism: 1970 to ~1985
  – 4 bits, 8 bit, 16 bit, 32 bit microprocessors

• Instruction level parallelism (ILP):
  ~1985 through today
  – Pipelining
  – Superscalar
  – VLIW
  – Out-of-Order execution
  – Limits to benefits of ILP?

• Process Level or Thread level parallelism
  – Pentium4 HT (Hyper-threading)
Why Multiprocessors?

- Complexity of current microprocessors
  - Do we have enough ideas to sustain 1.5X/yr?
  - Can we deliver such complexity on schedule?
- Slow (but steady) improvement in parallel software (scientific apps, databases, OS)
- Emergence of embedded and server markets driving microprocessors in addition to desktops
  - Embedded functional parallelism, producer/consumer model
  - Server figure of merit is tasks per hour vs. latency

Multiprocessor Classifications:

- Proposed by Michael J. Flynn, a professor at Stanford, in 1966

  • Based upon the number of concurrent instruction (or control) and data streams available in the architecture
  • Known as “Flynn Taxonomy”

SISD (Single Instruction Single Data)

• A sequential computer which exploits no parallelism in either the instruction or data streams.
• Examples of SISD architecture are the traditional uniprocessor machines

SIMD (Single Instruction Multiple Data)

• A computer which exploits multiple data streams against a single instruction stream to perform operations which may be parallelized.
• Example, an array processor or GPU.
MISD (Multiple Instruction Single Data)
- multiple processors on a single data stream
- Unusual due to the fact that multiple instruction streams generally require multiple data streams to be effective
- No commercial product
- Redundant parallel processing can be classified into this category i.e. air/space craft flight control computer

MIMD (Multiple Instruction Multiple Data)
- Most multiple CPUs scheme today
- Multiple processors simultaneously executing different instructions on different data

Memory Organization
- Centralized shared-memory multiprocessor or Symmetric shared-memory multiprocessor (SMP)
  - Multiple processors connected to a single centralized memory – since all processors see the same memory organization → uniform memory access (UMA)
  - Shared-memory because all processors can access the entire memory address space

Commercial SMP motherboard
- Integrated Audio with CODEC connector
- Integrated LAN controller (E举办的)
- Dual Channel Memory Bus
- Supports two Intel Xeon processors; FSB 533/667MHz
- Serial ATA
- USB 2.0 ports
- 400-pin 2.5V DDR DIMM socket
- Supports up to 4GB unbuffered DDR 266/333MHz, 800MHz or non-ECC modules
- 240-pin 2.5V DDR DIMM socket
- SiS usable slots
Memory Organization

- For higher scalability, memory is distributed among processors → distributed memory multiprocessors
- If one processor can directly address the memory local to another processor, the address space is shared → distributed shared-memory (DSM) multiprocessor
- If memories are strictly local, we need messages to communicate data → cluster of computers or multicomputers
- Non-uniform memory architecture (NUMA) since local memory has lower latency than remote memory