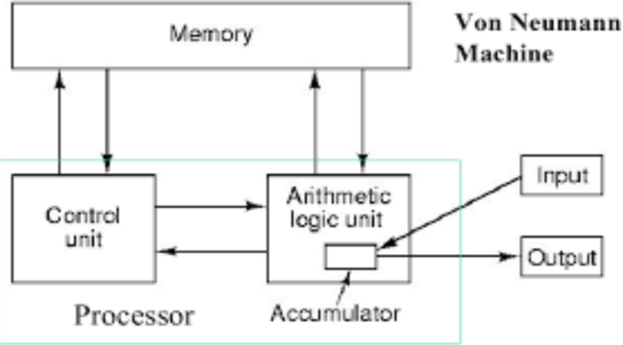
Modern Processors

Vipin Vasu

Outline

1. Stored Program Computer Architecture
2. General-purpose cache-based microprocessor architecture
3. Performance metrics and benchmarks
4. Low-Level Bench Mark
5. Transistors galore: Moore's Law
6. Pipelining
7. Superscalarity
8. SIMD
9. Conclusion

Stored Program Computer Architecture



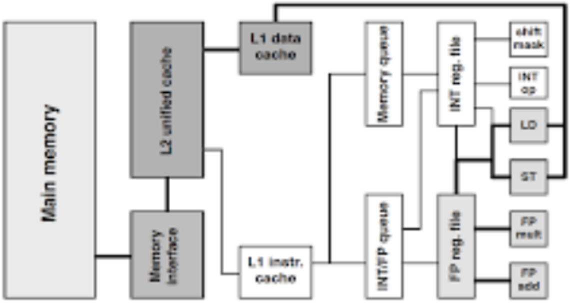
Stored Program Computer Architecutre

- Instructions and data must be continuously fed to the control and arithmetic units, so that the speed of the memory interface poses a limitation on compute performance.
- The architecture is inherently sequential, processing a single instruction with (possibly) a single operand or a group of operands from memory.(SISD)

General-purpose cache-based microprocessor architecture

- Microprocessors implement stored pgm....
- Modern processors have lot of componets but only a small part does the actual work -AU for fp and int operations.
- Rest are CPU regs,nowdays processors req all operands to reside in regs.
- LD(load) and ST(store) units handle instruction tranfer.
- Queues for instructions
- Finally Cache

General-purpose cache-based microprocessor architecture



Performance metrics and benchmarks

- Cpu components can operate at a peak performance
- Need to quatify this “speed”-DP and SP
- The performance at which the FP units generate results for multiply and add operations is measured in floating-point operations per second (Flops/sec). 2-4 DP in one cycle:clock freq 2-3ghz
- 4-12GFlops/Sec
- Data speed based on Main Memory and Cache tranfer speed.
- The performance, or bandwidth of thosepaths is quantified in GBytes/sec.

Low-Level Bench Mark

- A low-level benchmark is a program that tries to test some specific feature like, e.g., peak performance or memory bandwidth
- On standard microprocessors, performance grows with N until some maximum is reached, followed by several sudden breakdowns. performance stays constant for very large loops.
- In order to decide whether some CPU or architecture is well-suited for some application the only safe way is to prepare application benchmarks.

Transistors galore: Moore's Law

- Pipelined functional units
- Superscalar architecture
- Data parallelism through SIMD instructions
- Out-of-order execution
- Larger caches
- Simplified instruction set(CISC to RISC)

Pipelining

The most simple setup is a “fetch–decode–execute” pipeline, in which each stage can operate indepen-dently of the others

Superscalarity

If a processor is designed to be capable of executing more than one instruction or, more generally, producing more than one “result” per cycle.

- Multiple instructions can be fetched and decoded concurrently
- Multiple floating-point pipelines can run in parallel

SIMD

- The SIMD concept became widely known with the first vector supercomputers in the 1970s
- They allow the concurrent execution of arithmetic operations on a "wide" register that can hold, 2DP or 4SP.
- A single instruction can initiate four additions at once.
- Can be parellel or a single pipeline.

The End