Lecture 6

Lecturer: Emery Berger   *Scribe: Dimitar Gochev*

6.1 Cramming More Components onto Integrated Circuits - Moore, 1965

Moore's law: The trend in computing hardware of doubling the number of transistors that can be placed inexpensively on an integrated circuit every two years. The law still holds today, but the proximity of components is making heat a major problem.

Historically there has been a strong correlation between the number of components and clock speed, but clock speeds have plateaued in recent years because consumers have started to care more about power consumption and battery life. The trend for the number of components is still continuing upwards. (See figure 6.1)

Rather than increasing the clock speeds, CPUs today include multiple additional components, especially cache memory, to reduce communication costs, thereby still providing improved speeds. These components include SIMD, ALU, FPU, GPU, L1, and L2 cache.

SoC - "System on a chip" - combining multiple chips into a single one. This tight coupling addresses the problem of slow bus speeds. It also makes the manufacturing process cheaper. However it lacks modularity of design.

6.2 Validity of the Single Processor Approach to Achieving Large Scale Computing Capabilities - Amdahl, 1967

Amdahl's law: Used to determine the maximum possible speedup of a system when only a part of it is improved. In parallel programming, the maximum speedup that can be achieved by multiple processors is limited by the inherently sequential portions of the program. If a program is 10% inherently sequential, then the maximum possible speedup is 10x.

ILP (instruction-level parallelism) is a measure of how many operations can be done in parallel. Typically the IPC (instructions per cycle) hovers around 2.

The recent trend of adding multiple cores to CPUs to improve their speed can be described as a desperate approach due to a lack of a better ideas. It is unreasonable to expect that developers will suddenly start to write only parallel programs.

Multiple cores also suffer from a cache coherence problem when two processors have a local view of the same memory location and they try to write to it. A ping-ponging cache lines phenomenon can occur when cache coherence problems necessitate reading from memory all the time.

Intel uses the MESI (modified, exclusive, shared, invalid) protocol to keep track of states of every line in

memory.

Limits to multi-core:

- Amdahl's law

- Real estate trade-off – What is most valuable: CPU, cache, GPU, etc?

- Symmetry – many of the same processor vs. heterogeneous processors

- Programming

- Power – cores have to be weaker if there are multiple ones

- Dark silicon (electricity propagation) – there is a limit to the speed of electricity; we'll end up with portions of the chip which are not powered

# of Components

Clock Speed