1 Performance Evaluation

- flops is a standard to measure of performance; number of floating-point operations per second (flops); megaflops = mflops = 1000 flops
  - gigaflops
  - teraflops
  - petaflops (range of current top 500 supercomputers)
  - exaflops
  - numbers reported are usually peak theoretical flops or tests on Linpack benchmark (good implementations achieve approx. 80% peak theoretical)
  - gives no evaluation of interconnect performance; example NASA-Ames 1985, Intel processors for CFD (computational fluid dynamics) application, 80% of the processors were idle waiting for data
- MIPS = millions of instructions per second (non-scientific computing, banking, business, etc)
- Perfect club (1990s), University of Illinois Urbana-Champagne, I/O benchmark that failed around 1995
- HPC challenge benchmarks
- Intel Paragon
- SPEC benchmarks; unfortunately the code is proprietary

1.1 Some numbers as of May 2010

- Intel Westmere 6 core proc
  - 2.66 GHz - $1200
  - 2.80 GHz - $1500 (25% increase cost, 2% faster)
  - 2.93 GHz - $1750 (46% increase cost, 10% faster)
  - 3.33 GHz - $2000 (67% increase cost, 25% faster)
- AMD Opteron 6000 8 core proc
  - 2.3 GHz - $600
  - 2.4 GHz - $850 (42% increase cost, 4% faster)
- AMD Opteron 6000 12 core proc
  - 2.3 GHz - $1150
  - 2.4 GHz - $1600
- NVIDIA S2070 (GPU)
  - 2.3 GHz - $3450 (programming model is not clear)

1.2 Amdahl's Law

- From 1970s: If 10% of a program is serial, and the other 90% can be perfectly parallelized, and the program runs in 100 seconds, then the runtime is

$$\text{Time} = 10 + \frac{90}{p}$$

  So no matter how much we increase p, the best runtime is still 10 seconds.
- From 1980s: For most applications, the percentage of time spent in serial portion $T_S \to 0$ as the problem size increases.