ECEn 528
Study Guide – Background and Metrics

- Read Chapter 1 of H&P
  - Things to focus on
    - Architectural goals for different marketplaces
    - Dimensions of an ISA. We will not cover this material in class, but I will assume that you are familiar with this terminology.
    - Technology trends
    - Performance measurement: benchmarks, metrics, summary statistics
    - Parallelism, locality, common cases
    - Amdahl's law
    - Speedup
    - Processor performance equation

  - Clarifications
    - Computer organization within a processor is often called *microarchitecture*. Computer organization outside of a processor is sometimes called *system architecture*.
    - The correct mean to use (arithmetic, geometric, harmonic) when summarizing benchmark results is a hotly contested question. The arguments usually hinge on utility and statistical "goodness." H&P's statements about the geometric mean are **not** accepted by many. However, geometric mean is the most commonly used mean.
      - The argument I find most convincing is that if you have a bimodal or highly skewed distribution (which are the situations in which the means differ by large amounts), **no** mean is actually meaningful.
    - Remember speedup as "old" vs. "new" times, **not** slower vs. faster (because sometimes your "enhancement" actually slows things down)
    - H&P claim that $Speedup_{enhanced}$ is always greater than 1 on p. 39; this is **only** true when the "enhancement" actually helps (many enhancements actually slow down performance in some cases.)

  - Answer the following questions:

1. Why has performance improvement slowed down over the last five to ten years and what has the response of computer architects been?

There are substantial issues with power dissipation and a lack of exploiting parallelism efficiently. Architects stopped trying to simply make processors faster and relying on instruction-level parallelism, and have now turned to data-level, thread-level, and request-level parallelism (note that these are explicit forms of parallelism and currently require programmer attention!).

2. What are the major classes of computers?

Personal mobile devices, desktops, servers, clusters/warehouse-scale computers, and embedded

3. Assume that performance increased by 52% per year (as it did until recently). You have an idea which will improve the performance of the product you are currently designing, but which will delay its shipment date (due to additional design, test, and qualification time) by 8 weeks. How much more performance must your idea provide to keep up with the industry?

52%/year = 1%/week, so the idea must provide an additional 8% performance

4. What are the three main principles of computer design? (hint: only one says “principle” in its name)

1. Take advantage of parallelism
2. Principle of locality
3. Focus on the common case

5. What does it mean to have a speedup of less than 1?

Performance actually went down with the "enhancement!"

6. What does Amdahl's law imply about the limits to performance improvement?

Law of diminishing returns; the overall speedup from an enhancement can never be greater than the speedup in the specific area

7. When is it valid to compare the CPIs of processors?

When the clock cycle time and instruction count remain the same