

# Concluding – Advanced Computer Architecture 2024 and beyond

- This brings Adv Comp Arch 2023-2024 to a close
- 2023 marks this course's 28<sup>th</sup> year
- How do you think this course will have to change for
  - ➡ 2024-2025?
  - ➡ 2028?
  - ➡ 2035?
  - ➡ The end of your career?
- Which parts are wrong? Misguided? Irrelevant?
- Where is the scope for theory?

# Advanced Computer Architecture

## Chapter 11

Wrapping up

Directions for improving the course

Theoretical computer architecture

November 2023

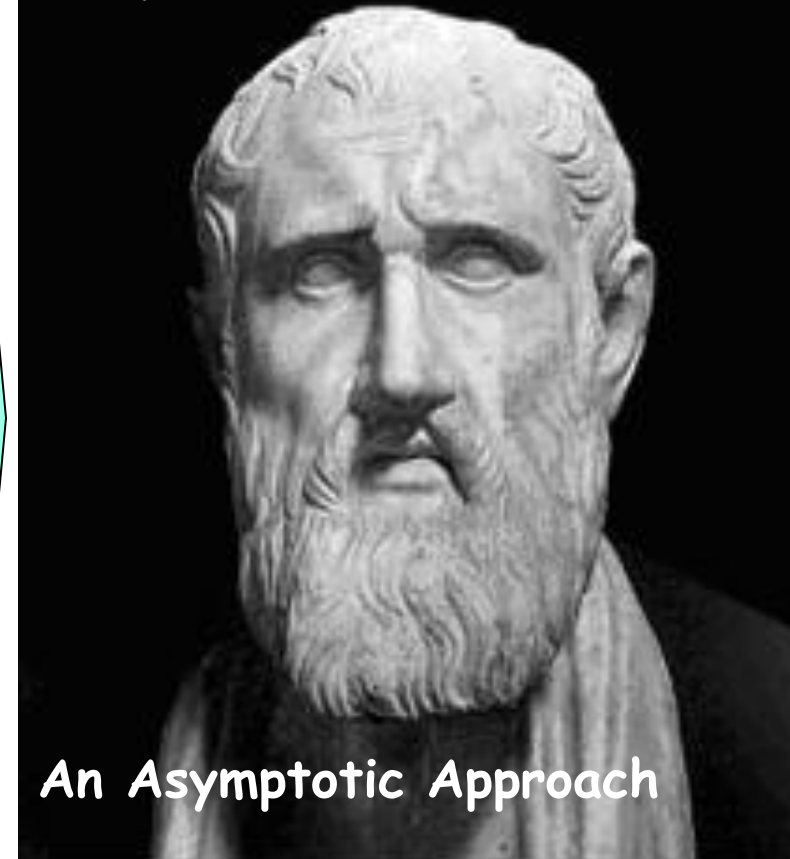
Paul H J Kelly

# Computer architecture – the future?



- A manifesto
- For computer architecture at the end of Moore's Law
- Where we confront fundamental physical constraints
- Where we have to account for fundamental costs
- Where architectural efficiency is paramount
- Where algorithms (in software *and* hardware) are paramount

Computer Architecture



An Asymptotic Approach

- ✧ The role of theory in computer architecture
  - Computing at the end of Moore's Law
  - Asymptotics versus reality
- ✧ Latency hiding in sequential machines with pipelined memory
  - Under what conditions can you hide latency, so performance is independent of RAM size?
  - Decoupling, address depth
- ✧ Latency hiding in parallel machines
  - Can you do this in a parallel machine?
- ✧ Models of computation for sequential computing
  - Counting FLOPs isn't enough: can we reason abstractly about the metrics that matter?
  - Uniform memory hierarchy: distinguishing cache-efficient algorithms
  - Cache-oblivious algorithms

- Models of computation for parallel computing
  - VLSI models; Area-time tradeoffs
  - BSP, Parallel memory hierarchy (PMH)
  - PRAM emulation; Ranade's machine (combining, randomisation, two-phase random routing)
- Caches: LRU stacks, cache obliviousness, AC/DC and the Bellman equation?
- Competitive strategies: spinlocks, paging, victim caches
- Some key algorithms: sorting, FFT, prefix scan, sparse matrix-vector multiply, geometric multigrid, parallel graph search
- Communication-avoiding algorithms
- Physical fundamentals: “plenty of room at the bottom”, noise, reliability, reversibility
- Frontier questions
  - Why is the physical universe such a bad platform for simulating the physical universe?

# Topics we should try to include...<sup>6</sup>

- ✚ Transactional memory and lock elision (and speculative cache updates)
- ✚ Datacentre architecture
- ✚ More on cache-coherency protocols
- ✚ More on memory system architecture – stacked, processor-in-memory, non-volatile
- ✚ More on predictors – I-prefetch, D-prefetch, aliasing predictors
- ✚ More on power – principles, mitigations
- ✚ Dark silicon
- ✚ Architectures for neural networks
- ✚ More on graphics aspects of GPU architecture
- ✚ More on performance optimisation methodology and tools
- ✚ Compiler topics eg loop optimisation, instruction scheduling
- ✚ More on security? CHERI? Control-flow integrity? Enclaves?
- ✚ More on more side-channel vulnerabilities
- ✚ **Less of...?**
- ✚ **Better explanation of?**
- ✚ **Your ideas?**