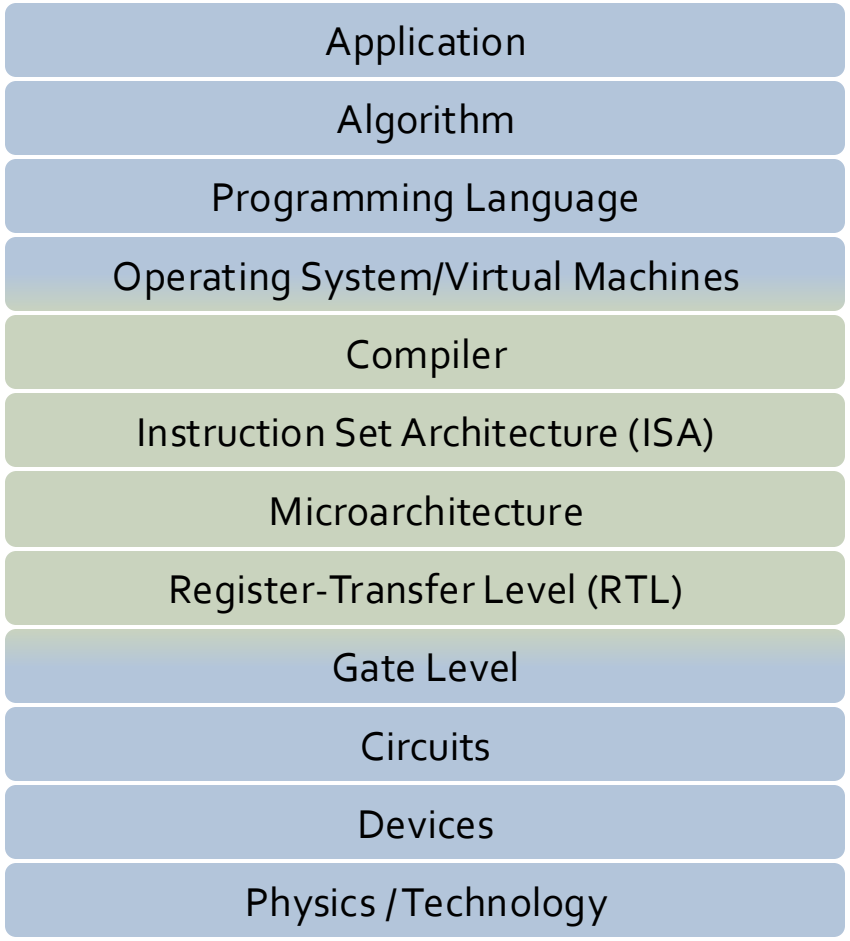
Computer Architecture
Finale

Ting-Jung Chang

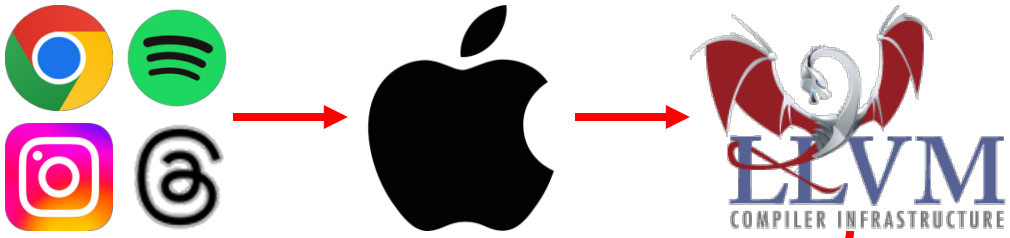
NYCU CS

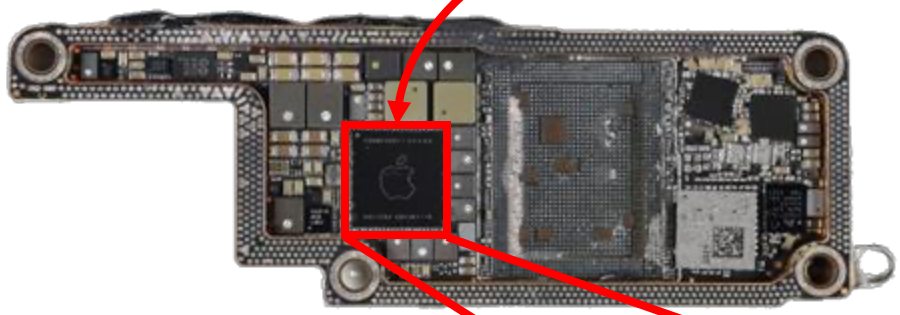
Abstraction Layers in Modern Systems

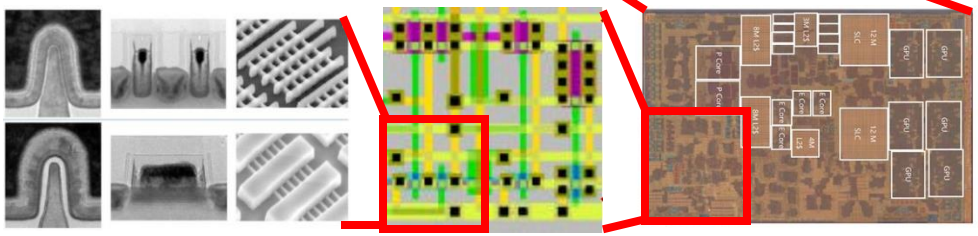


The Complexity of Modern Computer Systems

Application







Physics / Technology

Computer Architecture

~731,000,000 transistors

- Instruction Level Parallelism
  - Superscalar
  - Very Long Instruction Word (VLIW)
- Long Pipelines (Pipeline Parallelism)
- Advanced Memory and Caches
- Data Level Parallelism
  - Vector
  - GPU
- Thread Level Parallelism
  - Multithreading
  - Multiprocessor
  - Multicore
  - Manycore

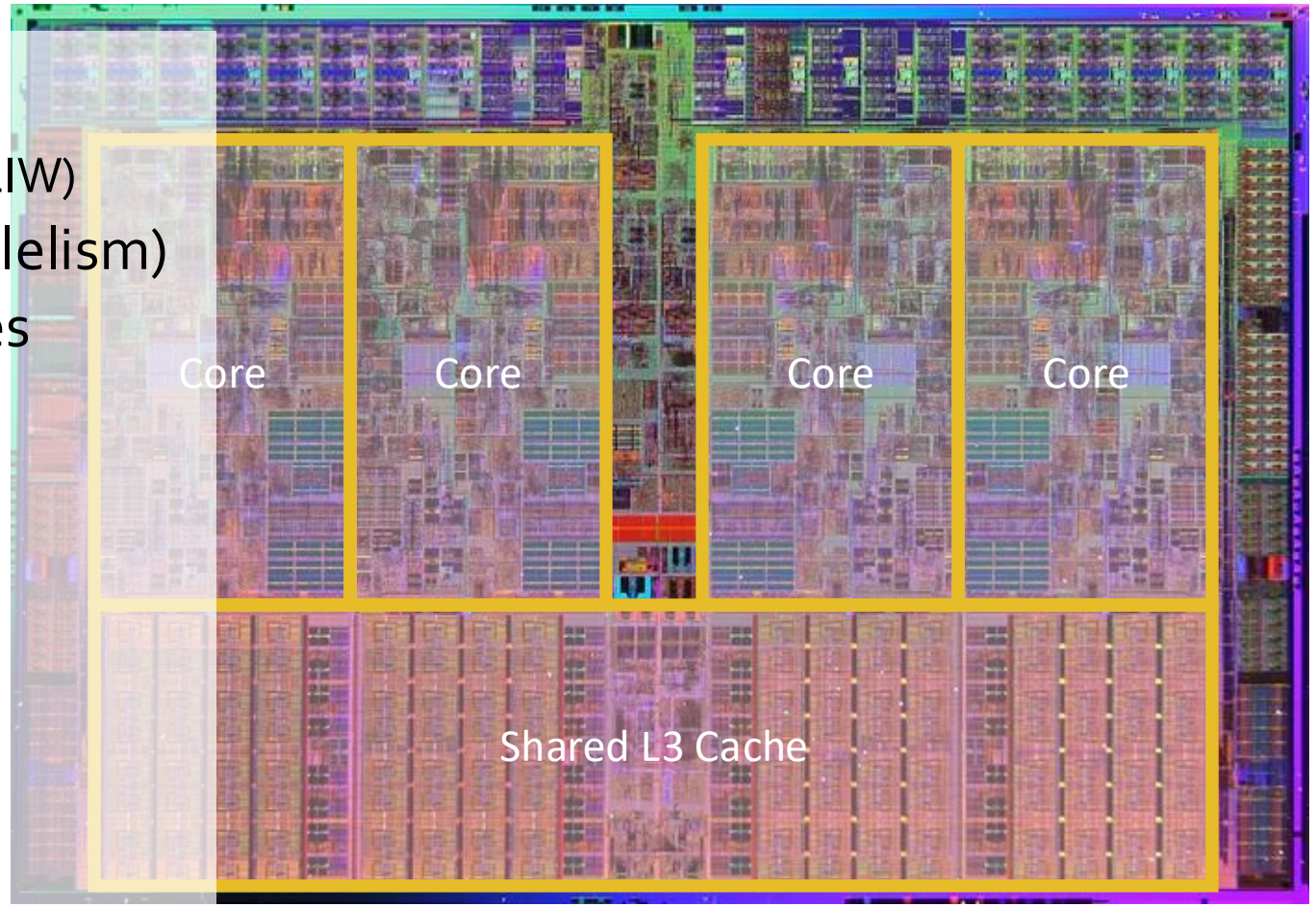
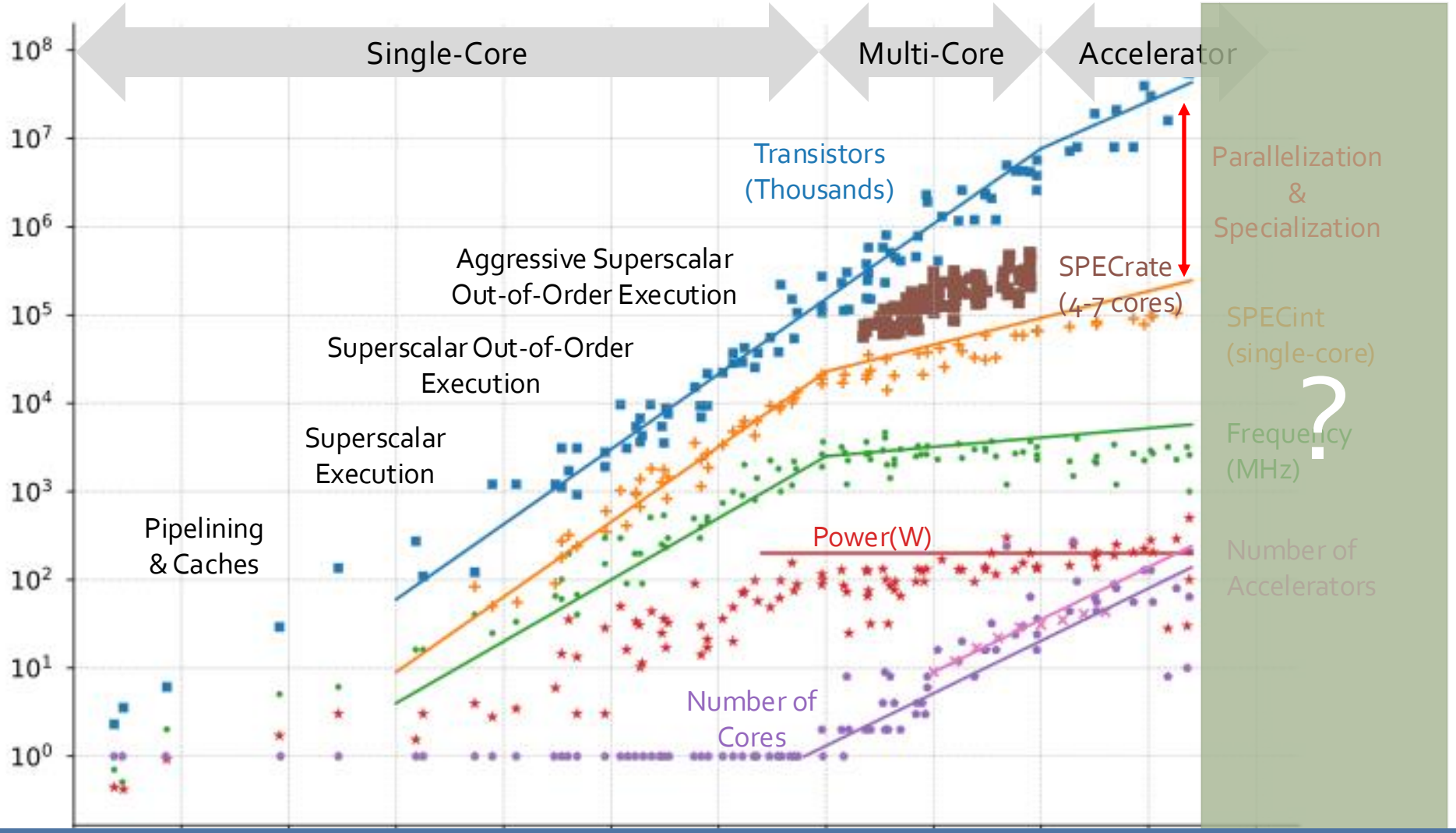


Photo of Intel Nehalem Processor, Original Core i7, © Intel

In this Course…

- Early simple machines, complex pipelining, out-of-order execution, VLIW, branch prediction, caching, prefetching, memory management, vectors/SIMD, multithreading, memory consistency, cache coherence, synchronization, interconnect…
- Just an introduction to main concepts in modern computer architecture, could easily spend a semester course on any one topic!



How to go beyond?

Computer Architecture Today

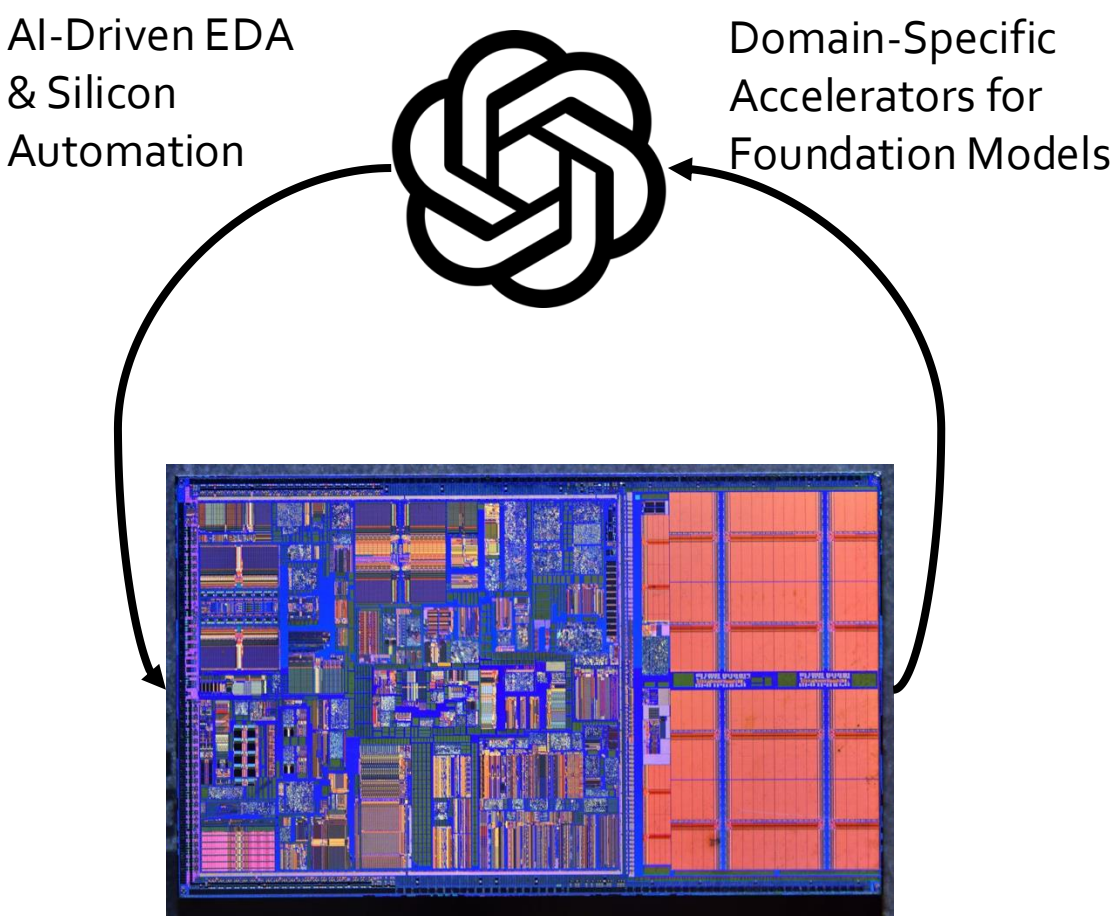
- Explosion of interest in custom architectures due to end of transistor scaling
  - Alibaba, Apple, Amazon, Bytedance, Meta, Google, Huawei, Microsoft, Qualcomm, Tencent, Tesla, design and build their own processors and SoCs!

- But need to learn about application domains
  - Cannot just work with precompiled binaries anymore!

Big Tech's Homegrown Chips

Company	Chip	Launched	Generation
Amazon	Trainium	2022	3
Amazon	Inferentia	2019	2
Google	TPU	2015/2017	7
Microsoft	MAIA	2023	1
Meta	MTIA	2023	2
Amazon	Graviton	2018	4
Google	Axion	2024	1
Microsoft	Cobalt	2023	1

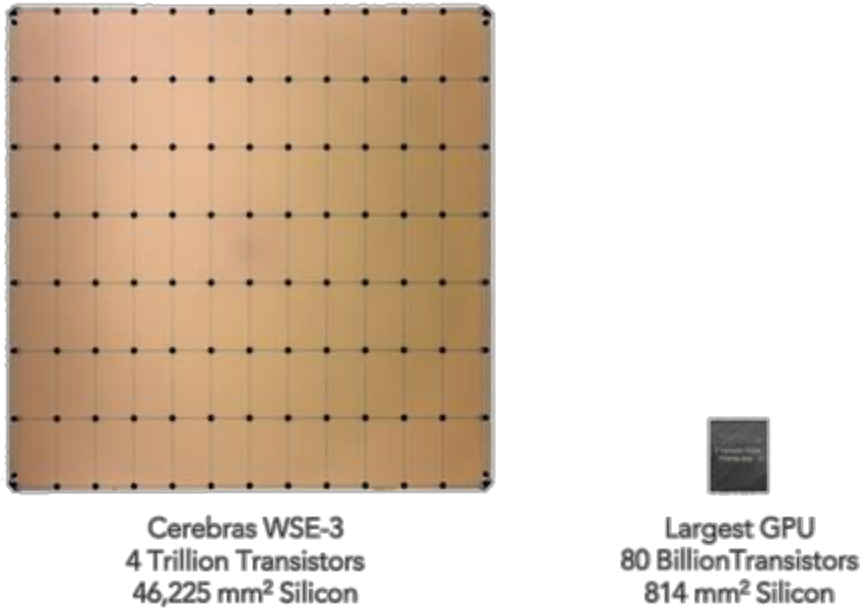
AI ↔ Chip Design



- Remarkable achievements
  - Design quality evaluation
    - Pow er, timing, area, routability, etc.
- Functional reasoning
  - Arithmetic word-level abstraction, SAT, etc.
- Optimization
  - Design space exploration, etc.
- Generation
  - RTL code, verification, etc.
- …

Field has Advanced

- Apple Announces The Apple Silicon M1: Ditching x86

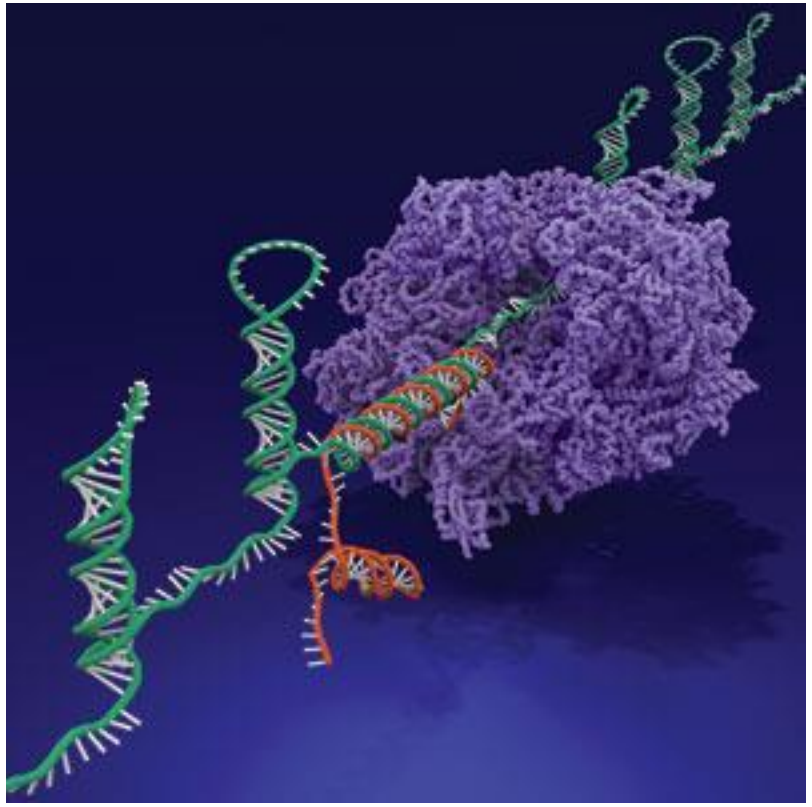


[Cerebras Wafer-Scale AI Processor]

[Selene AI Supercomputer]

New Frontiers

[IBM Q]



[IEEE Spectrum:
Biocomputer and Memory
Built Inside Living Bacteria]

Thanks!

Acknowledgements

- These slides contain material developed and copyright by:
  - Arvind (MIT)
  - Krste Asanovic (MIT/UCB)
  - Joel Emer (Intel/MIT)
  - James Hoe (CMU)
  - John Kubiatowicz (UCB)
  - David Patterson (UCB)
  - Christopher Batten (Cornell)
  - David Wentzlaff (Princeton)
- MIT material derived from course 6.823
- UCB material derived from course CS252
- Cornell material derived from course ECE 4750
- Princeton material derived from course ECE 475