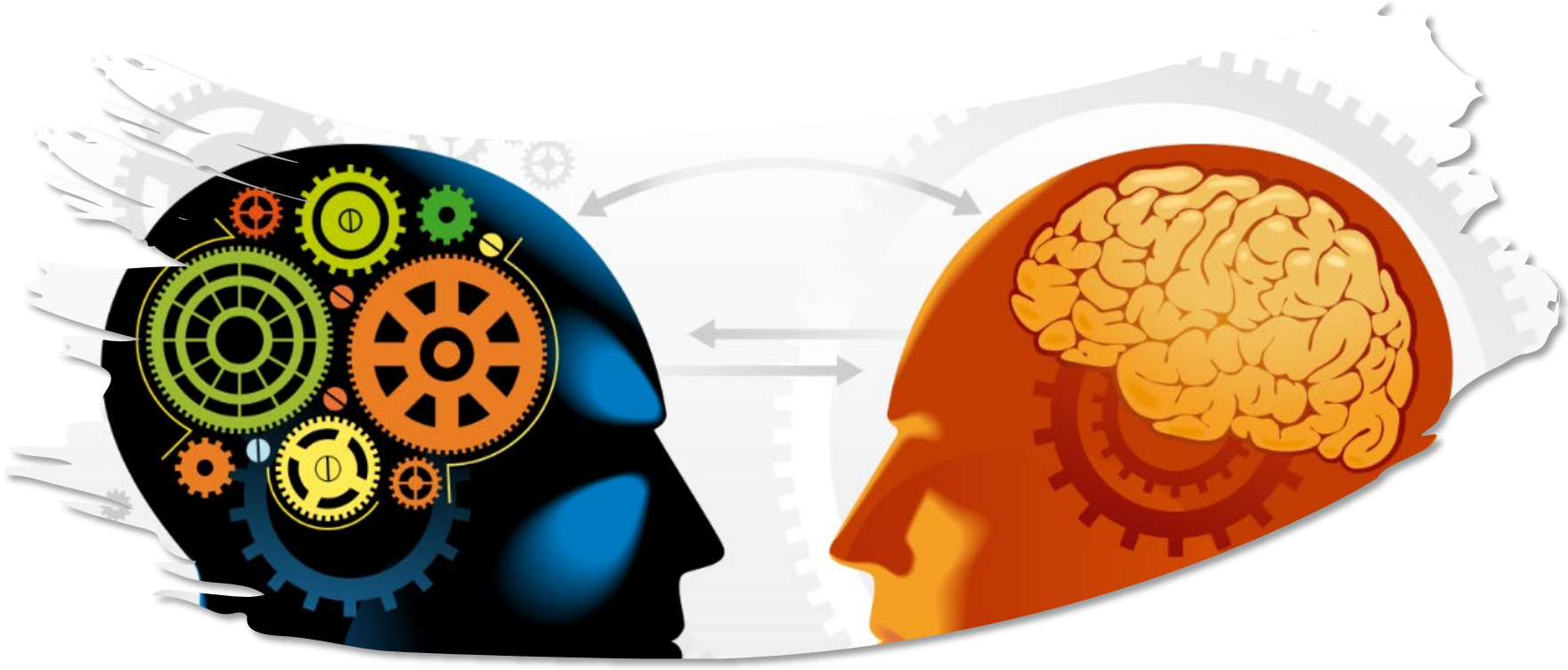


Building the Infinite Brain

COMP 590/790

Raghavendra Pradyumna Pothukuchi

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

✉ raghav@cs.unc.edu

Takeaways

Review B3: Ch. 2

What are some goals in designing the memory subsystem?

Capacity, program abstraction, protection, and sharing

What is virtual memory?

Abstraction to deliver the goals

What is a bottleneck in virtual memory systems?

Address translation

How can addressing bottlenecks be tackled?

Through fast paths: TLBs

How to choose the right memory hierarchy design?

Tailor it to the access patterns and layout: general purpose to domain specific

Understand the relevant "systems" problems and identify solutions

For Today

- Quick review
- **Multicore and multithreaded processors**

Achieving Efficiency

Pipelining

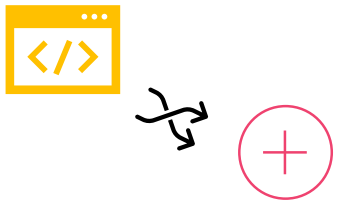
Concurrency among consecutive instructions

Superscalars

More concurrency among consecutive instructions

Out of order execution

Extending concurrency through tracking dataflow



CPU

Instruction level parallelism

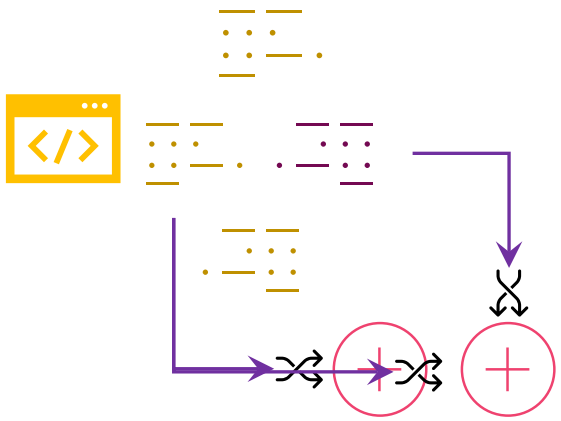
Going Beyond a Single Dataflow

Thread level parallelism

Existed before Tomasulo! | https://people.computing.clemson.edu/~mark/

Implicit

Parallelism from a sequential program



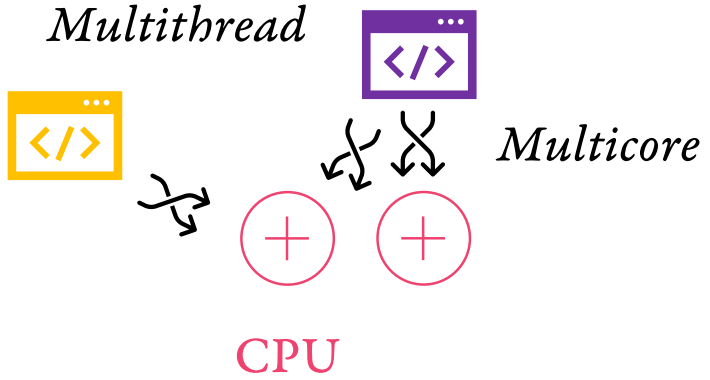
Explicit

Parallelism from multiple programs

What is a program?

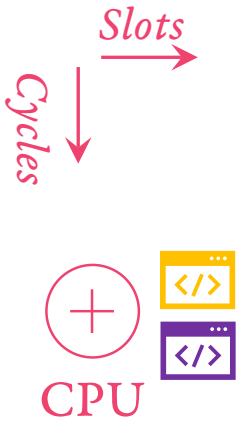
Independent instruction stream

Program counter, address space



"Concurrency"

Multithreading

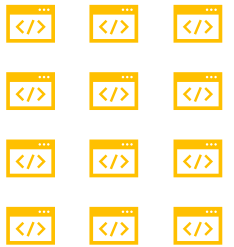


Scalar

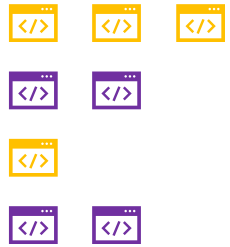
Fine (interleaved)

Coarse (block interleaved)

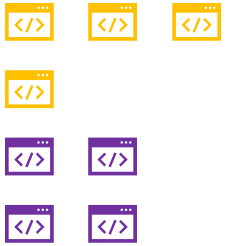
Superscalar



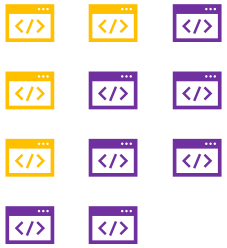
Fine

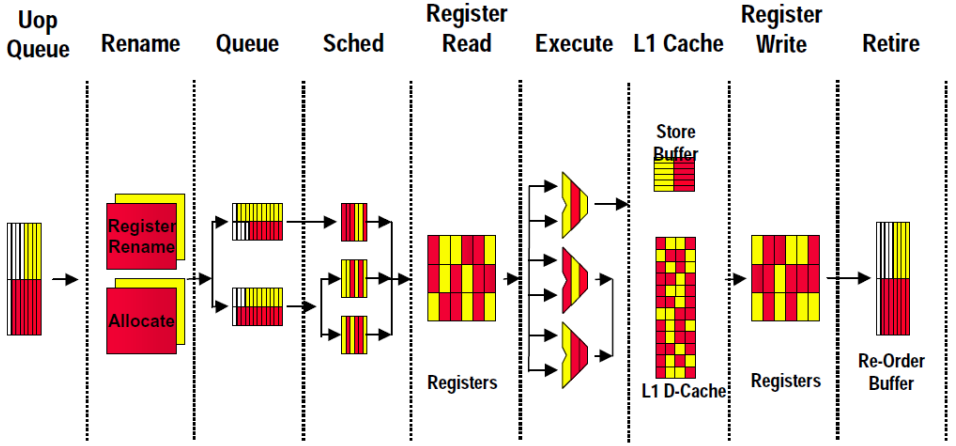


Coarse

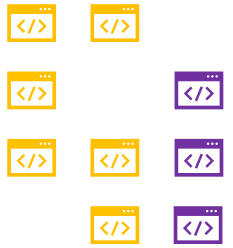


Simultaneous





Very Large Instruction Word (VLIW)
Static

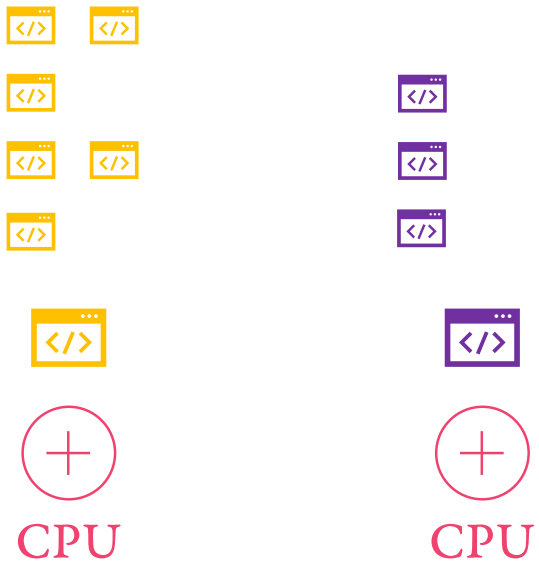


"Hyper-Threading Technology Architecture and Microarchitecture", Intel Technology Journal Q1, 2002

Chip Multiprocessors (Multicore)

Strategy to scale hardware

Each core/processor/CPU can be simple, or complex, like multithreaded



Utilizing multicores

Software (semi) invisible

Thread-level speculation

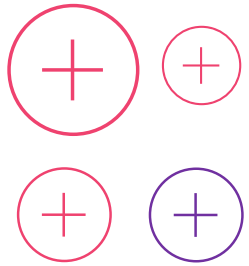
Software visible

Parallel programming

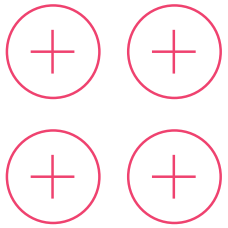
Multicore Organization

Compute

Asymmetric

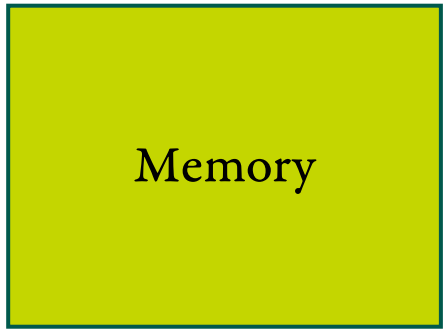


Symmetric

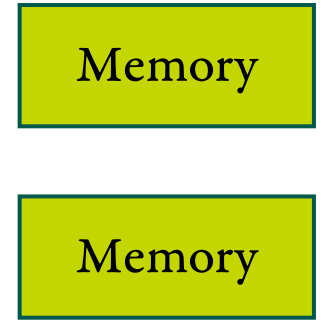


Memory

Centralized

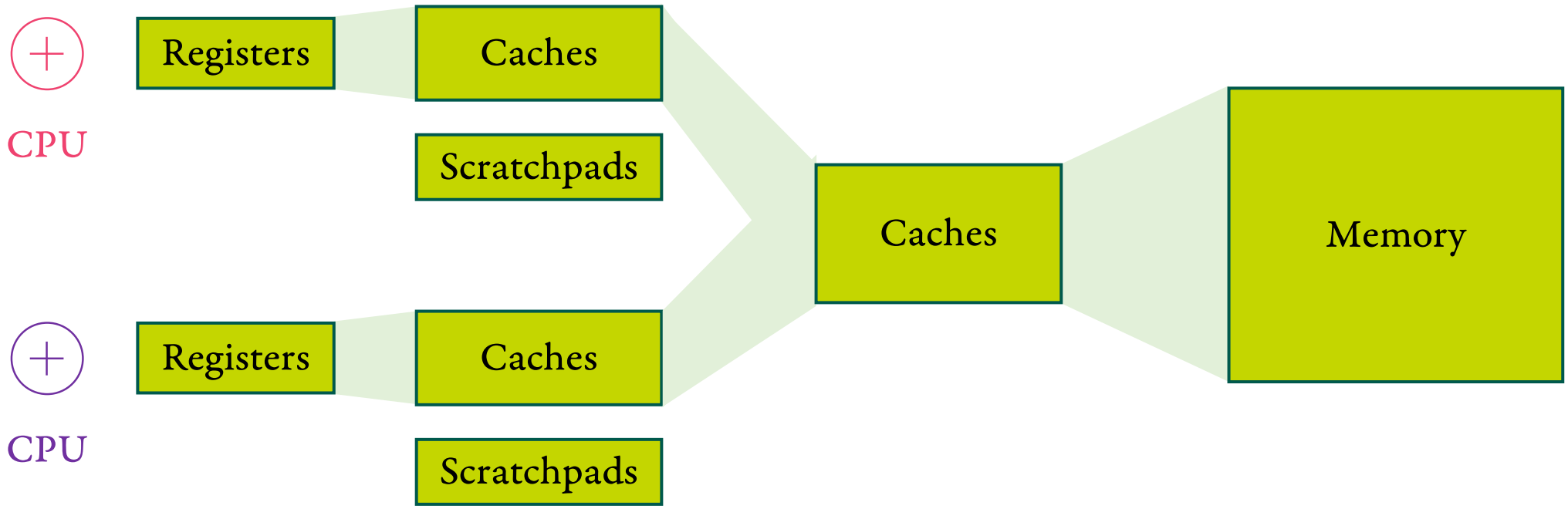


Distributed



Uniform or Non-uniform memory access

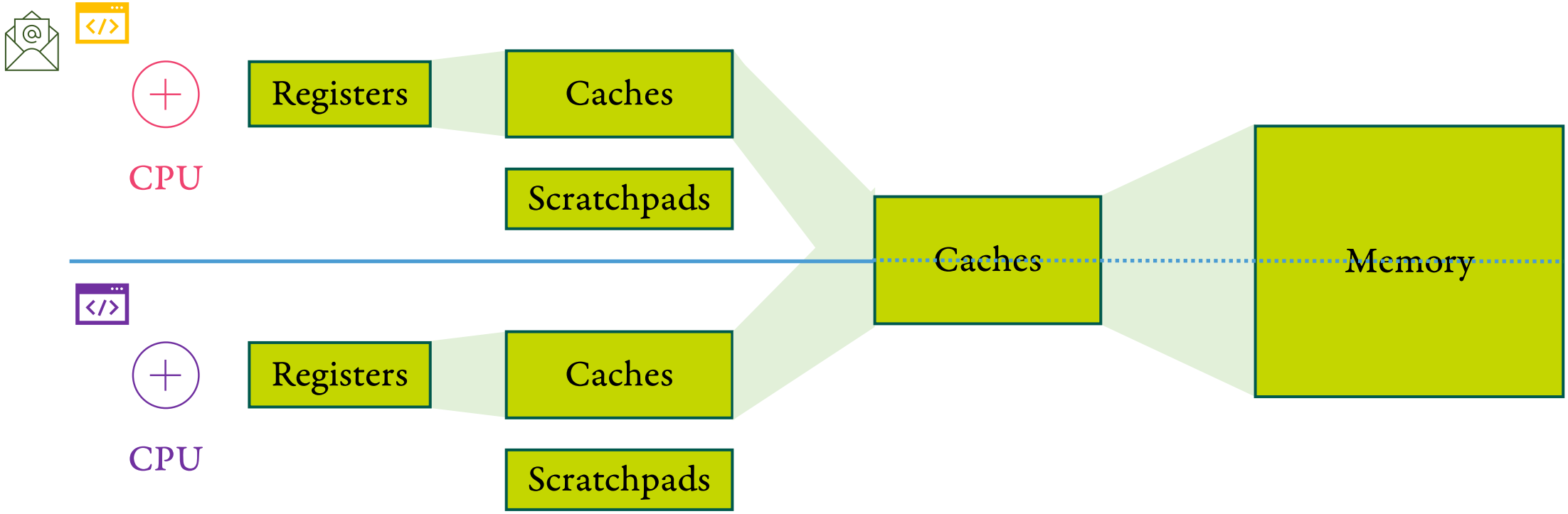
Multicore Organization



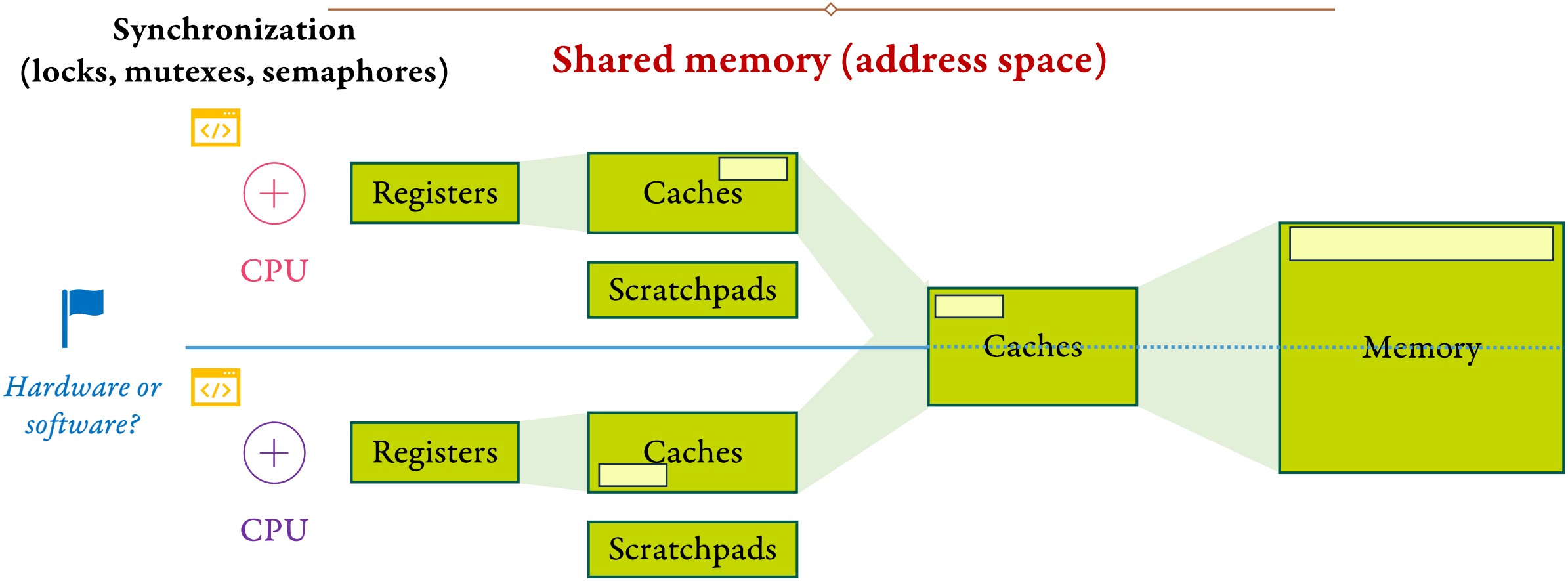
Utilizing Multicore Parallelism

Send and receive messages (inter process communication)

No address space sharing



Utilizing Multicore Parallelism



How to coordinate memory accesses?

Other parallel architectures, and scaling beyond a multicore chip?

Takeaways

What are the ways to improve throughput beyond single instruction stream?

Multithreading, chip multiprocessing, and combinations

What is the difference between multithreading and chip multiprocessing?

Multithreading: multiple instructions on the same core/processor/CPU
Chip multiprocessing: using multiple cores on the same chip

What are the aspects of multicore organization?

Compute (symmetric or asymmetric) and memory (centralized or distributed)

How are hardware and software parallelism related?

Abstractions are different—software (threads, processes), hardware (threads, contexts)
Hardware: implicit (single instruction stream) or explicit (multiple instruction streams)
Software: sequential (via automatic parallelism) or parallel (shared memory or message passing)

Image Credits (Educational, Fair Use)

- Title image: VLADGRIN, https://www.istockphoto.com/vector/human_-machine-gm147409511-16840728 (Educational fair use)
- Infinite brain: Science wonder stories, May 1930, Illustrator: Frank R Paul, Editor: Hugo Gernsback
- Brain color, ICs, cloud server, black rat: No attribution required (Hiclipart)
- Hand with spoon: public domain freepng
- Signals: https://www.nature.com/articles/nrn3724
- Thought clouds: F. Willett et al./*Nature* 2021/Erika Woodrum, https://med.stanford.edu/neurosurgery/news/2022/bci-award. https://www.the-scientist.com/news-opinion/brain-computer-interface-user-types-90-characters-per-minute-with-mind-68762
- Picture of scientists: https://www.cs.auckland.ac.nz/~brian/rutherford8.html (original: Pierre de Latil), Bush (Carnegie Science), Others (Wikipedia, National Academies, IEEE, and university profile images)
- Flowchart: Pause08 – flaticon.com; Digital brain: Smashicons – flaticon.com; Quantum processor icons created by Paul J. - Flaticon
- Server rack: upklyak – freepik.com
- Arm, Lotus: Adobe stock
- Quantum processor: Rigetti computing
- Images of implanted users: Top: Case Western Reserve University (https://thedaily.case.edu/man-quadriplegia-employs-injury-bridging-technologies-move-just-thinking/), Bottom: Jan Scheuermann (University of Pittsburgh/UPMC; https://www.upmc.com/media/news/bci-press-release-chocolate)
- Images of wearable BCIs: Cognixion, NextMind
- Types of BCIs: "Brain–computer interfaces for communication and rehabilitation,
- Illustrative BCI: Neuralink
- Electrodes: "Electrochemical and electrophysiological considerations for clinical high channel count neural interfaces", Vatsyayan et al.
- Form factors: Neuropace, Medtronic, Bloomberg, "Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS" by Vansteensel et al., Blackrock Neurotech
- Jose Delgado's video: Online, various sources (CNN, Youtube)
- Video of Kennedy and Ramsey: Online, various sources (Youtube, Neural signals)
- Code snippet inspiration: ECE 252 slides at Duke (Dan Sorin et al.)
- Apple processor pipeline: https://dougallj.github.io/applecpu/firestorm.html