Introduction to Parallel Programming

Marty Kandes, Ph.D.

Computational & Data Science Research Specialist
High-Performance Computing User Services Group
San Diego Supercomputer Center
University of California, San Diego

SDSC HPC/CI Training Series
Friday, April 8th, 2022
2:00PM - 4:00PM PST

Today

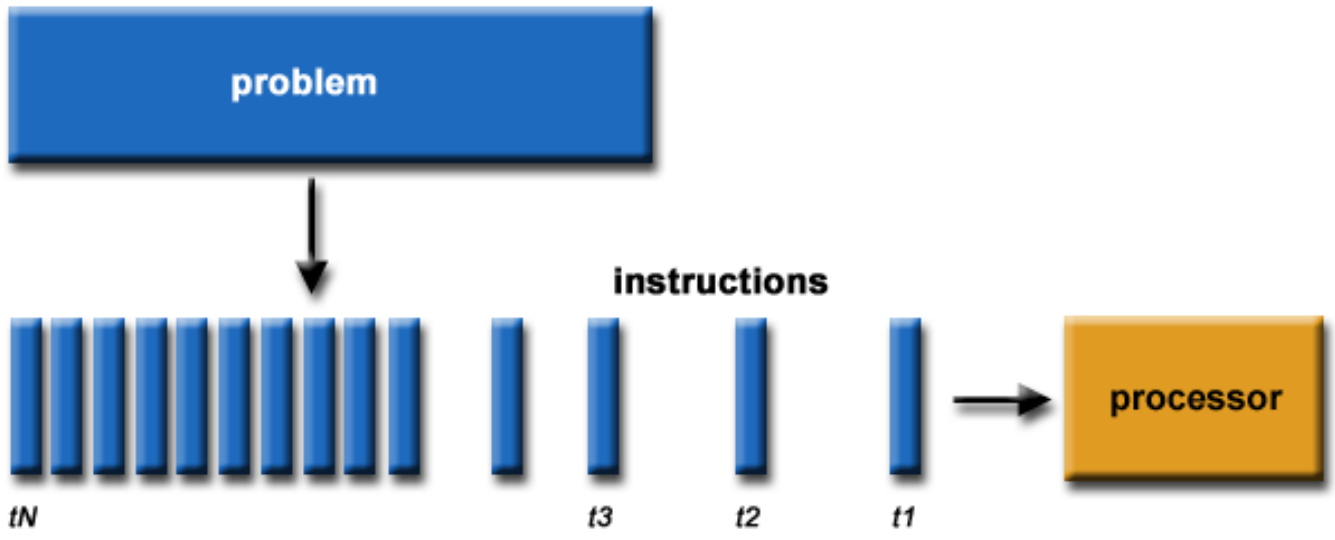
- Review basic parallel computing concepts
- Get you up and running on Expanse
- Embarrassingly parallel problems
- Thread-based parallel programming with OpenMP and Python

Parallel Computing Concepts (for Non-Programmers)

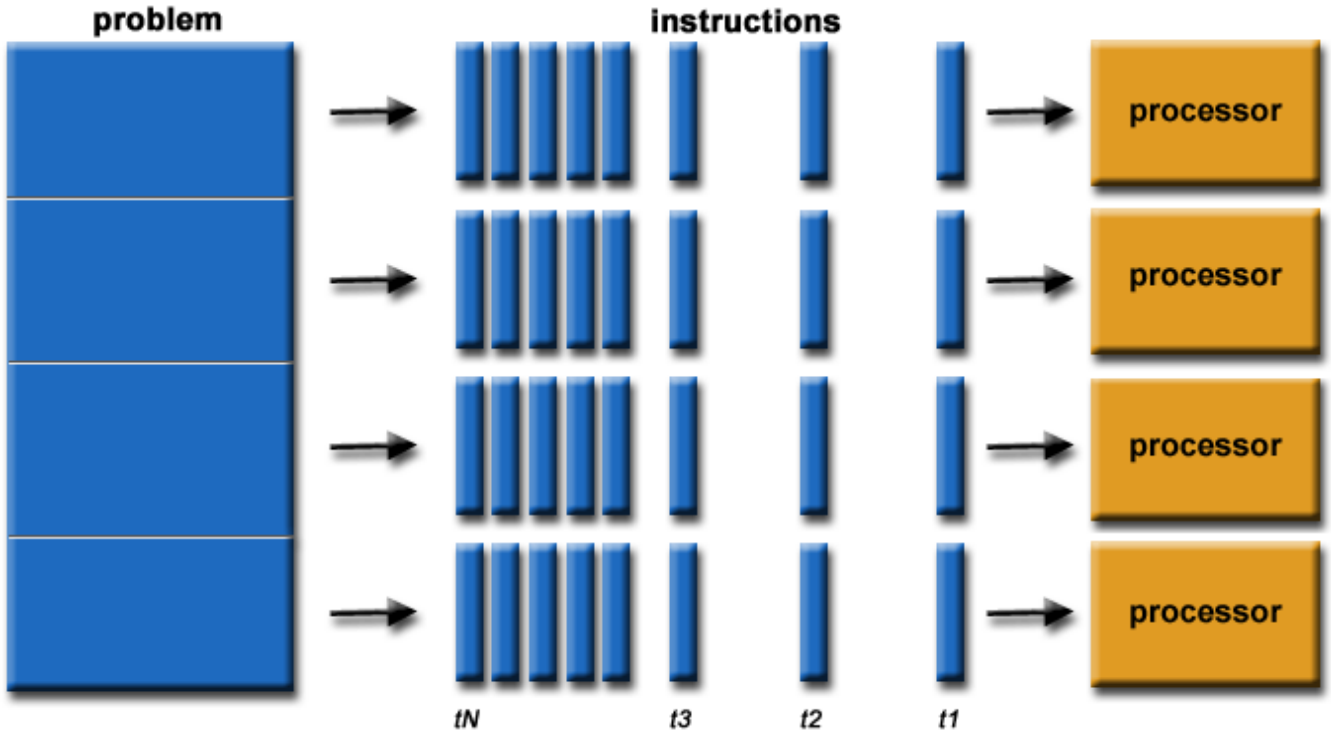
"The Parallel Computing Concepts talk presents a high-level overview of the technologies and software needed by HPC/CI users who need to access HPC systems, and understand the architecuture and ideas behind HPC/Parallel computing. The session includes code and figures for presenting scaling data - the right way and the wrong way - and illustrating the limits on scalability imposed by Amdahl's Law."

https://github.com/sdsc-hpc-training-org/hpc-training-2022

Serial Computation

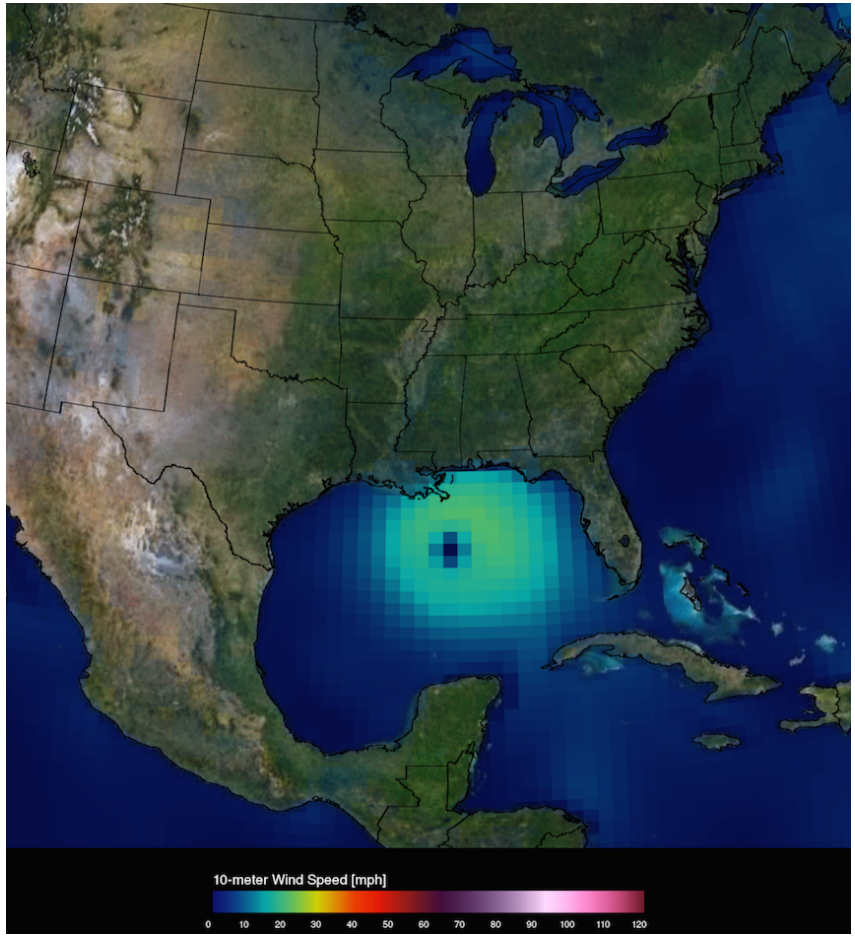


Parallel Computation



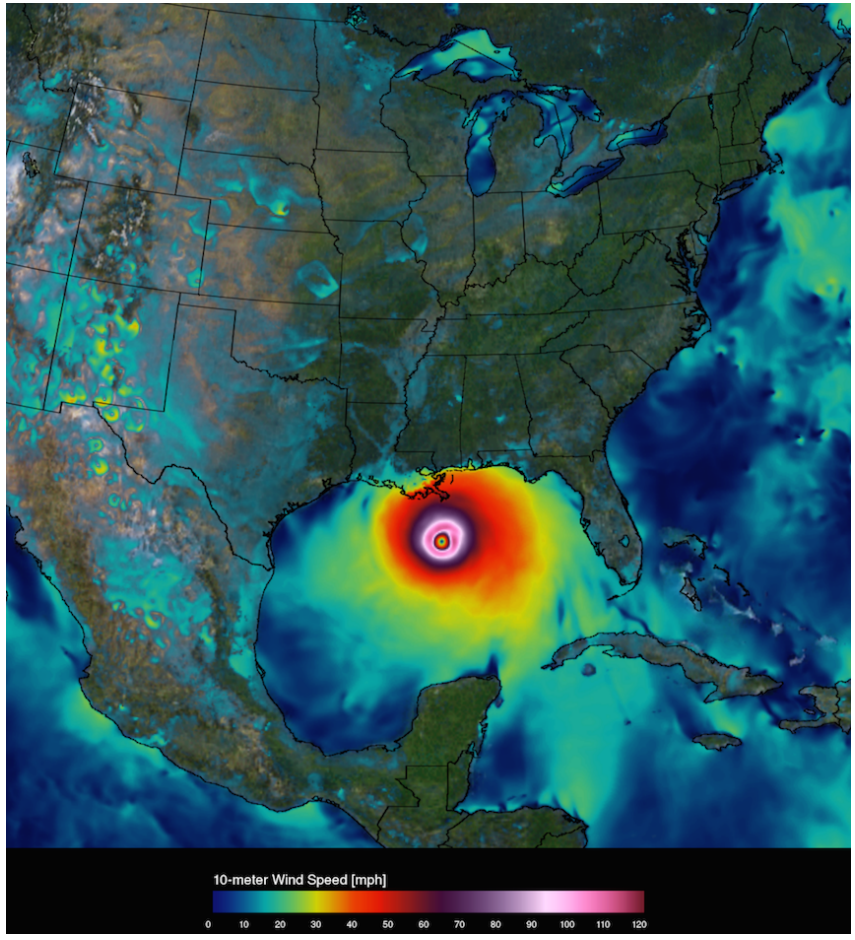
Why go parallel?

Speed



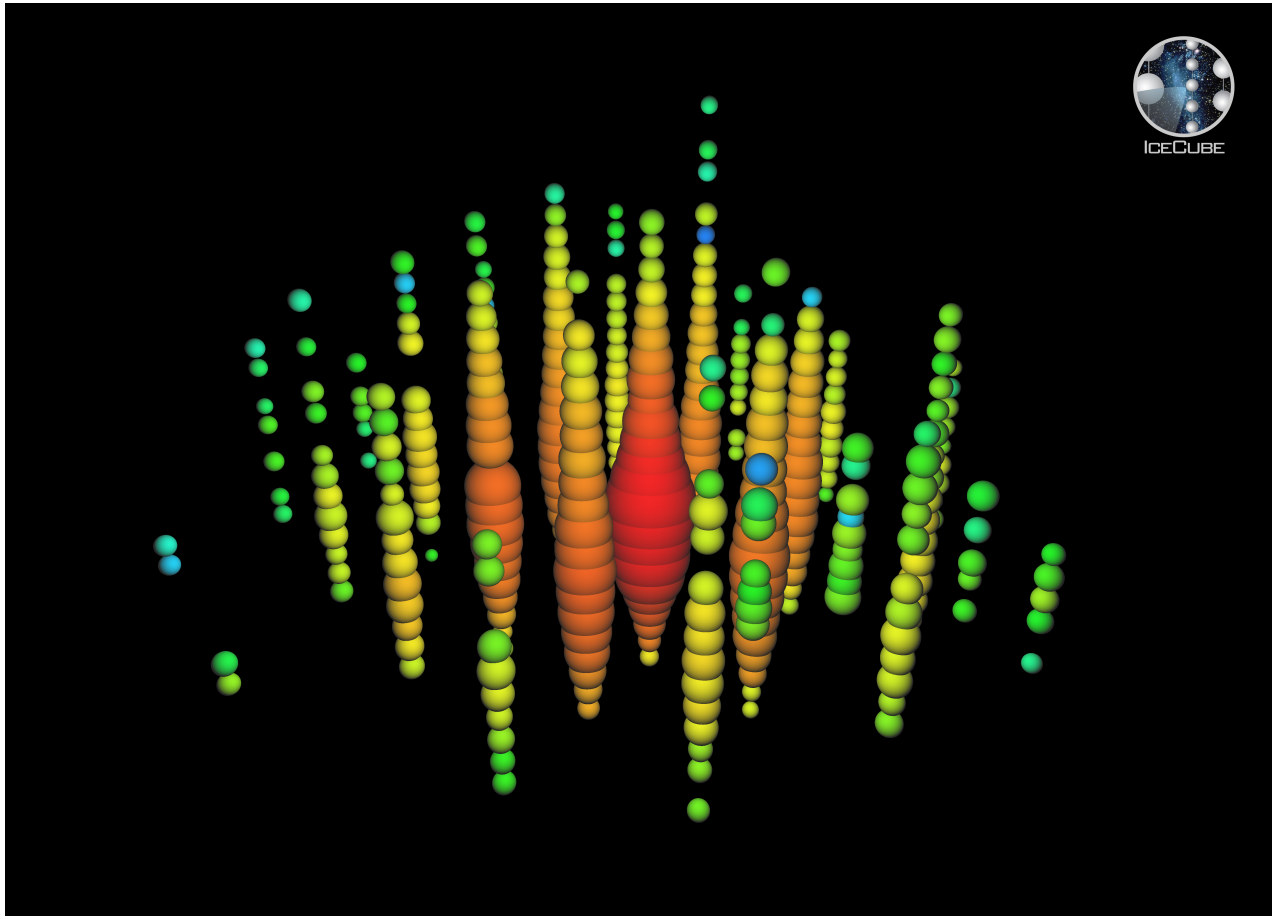
Solve a problem more quickly

Scale



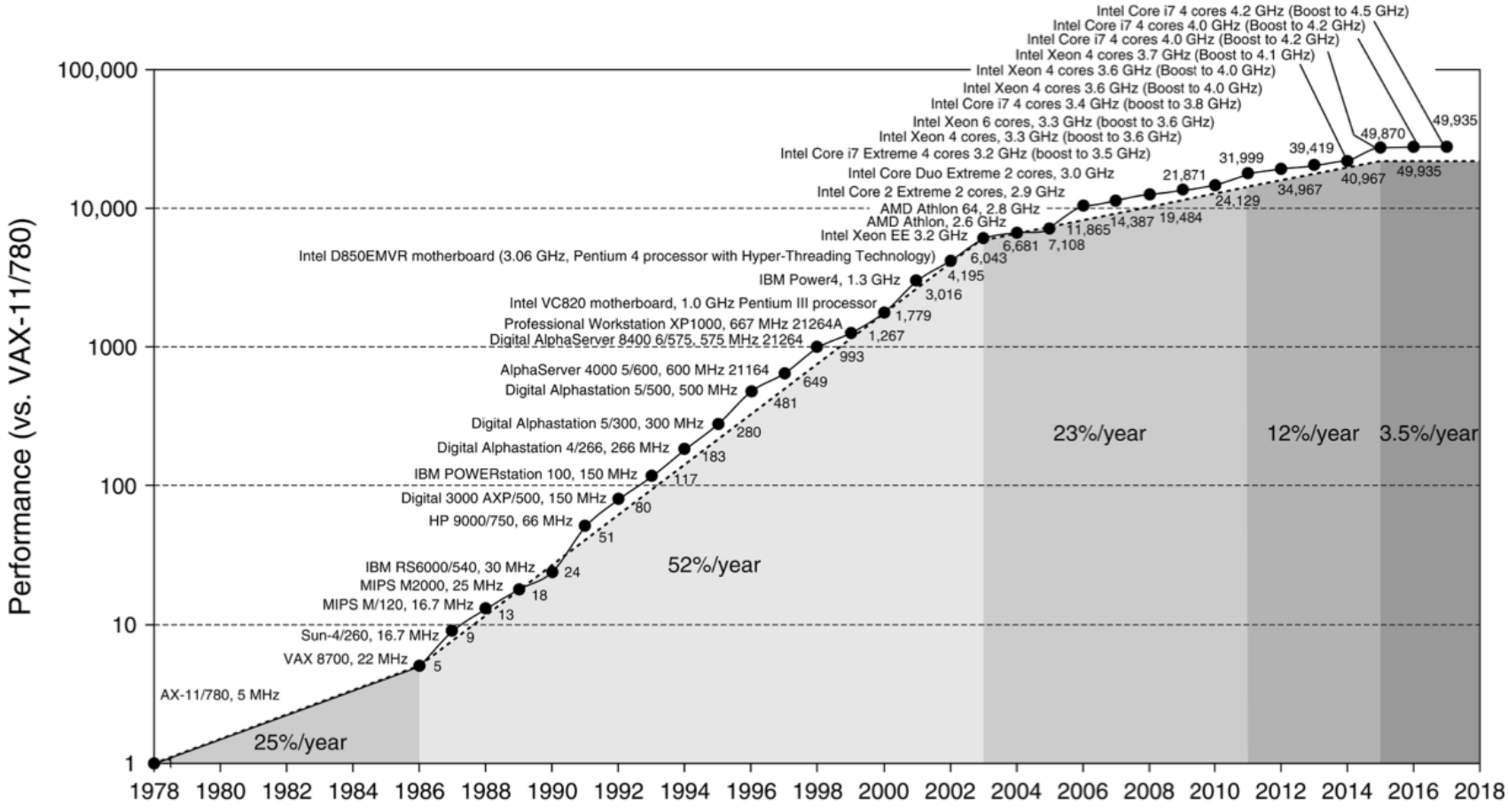
Solve a larger, more complex problem with higher fidelity

Throughput



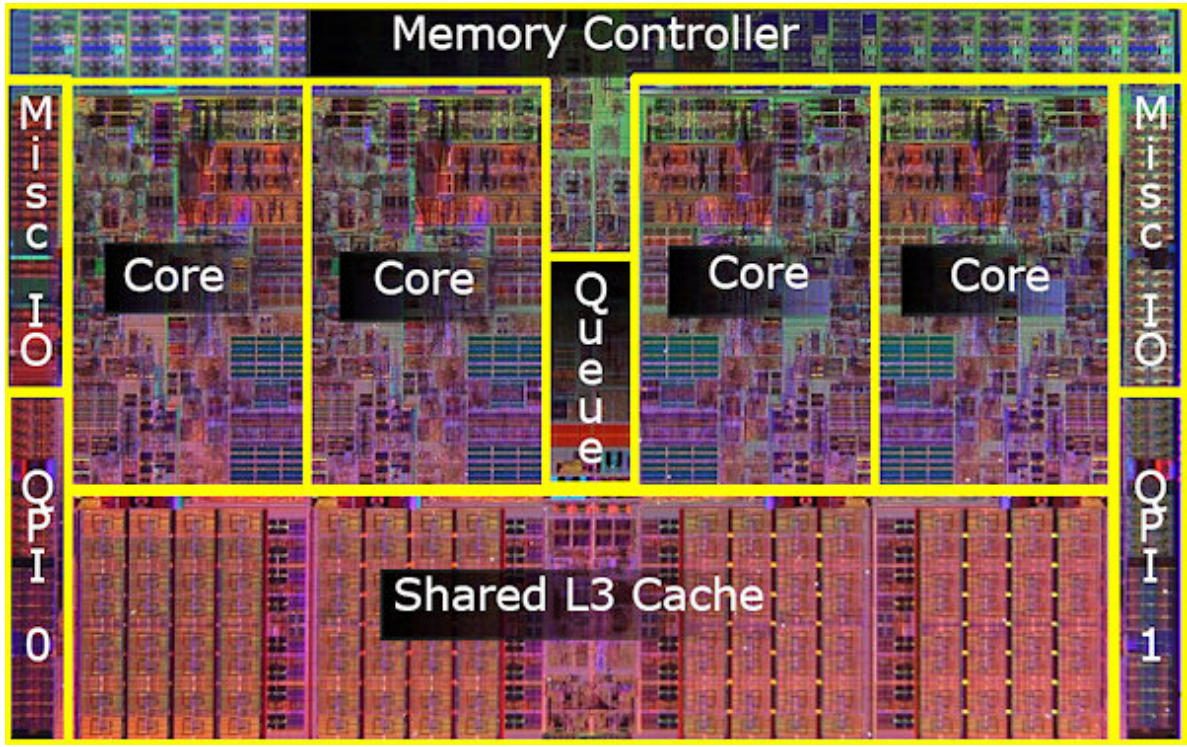
Solve many (simple) problems more quickly

The End is Near: Single-Core (Uniprocessor) Performance



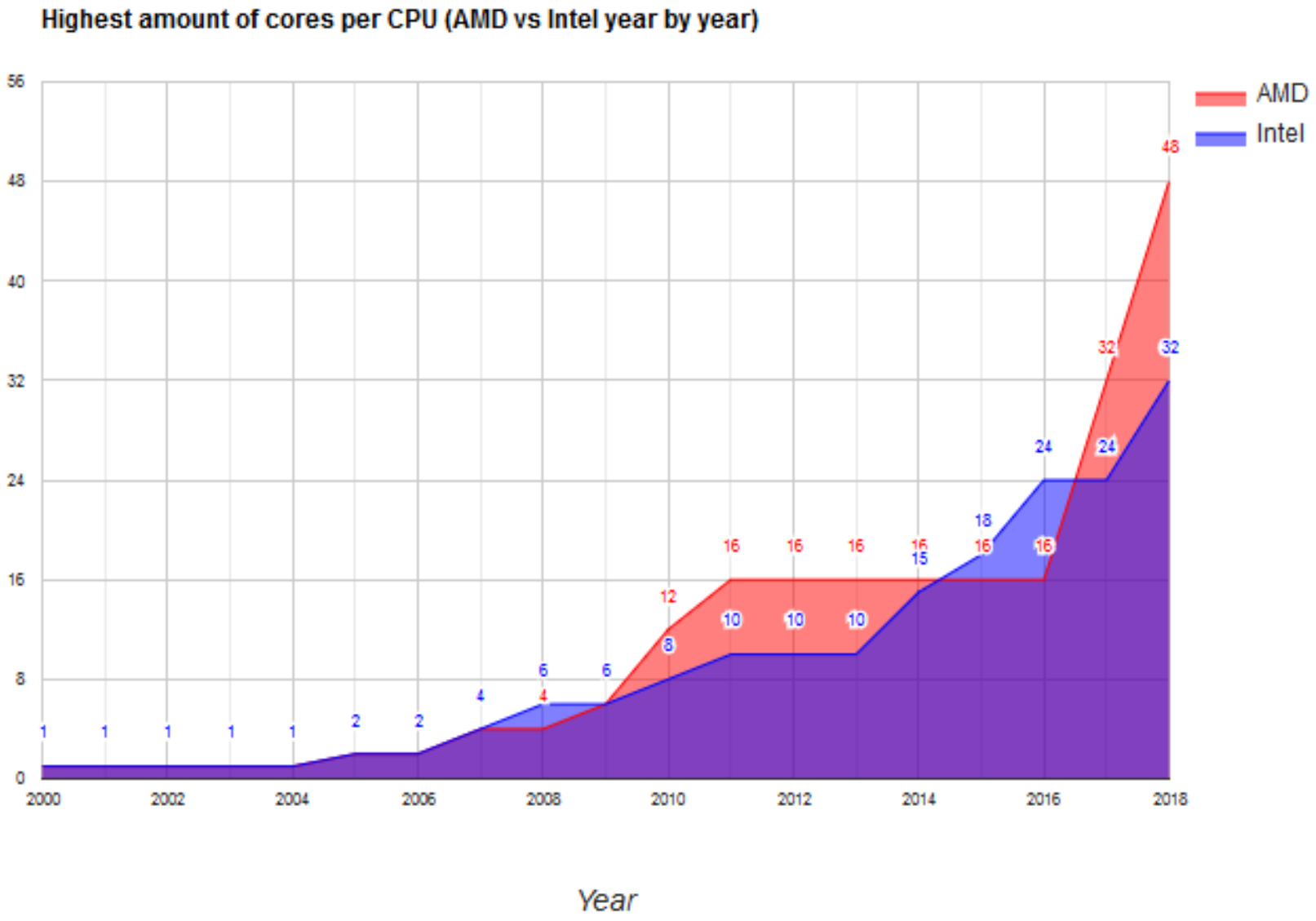
What's the solution?

Multi-Core Processor

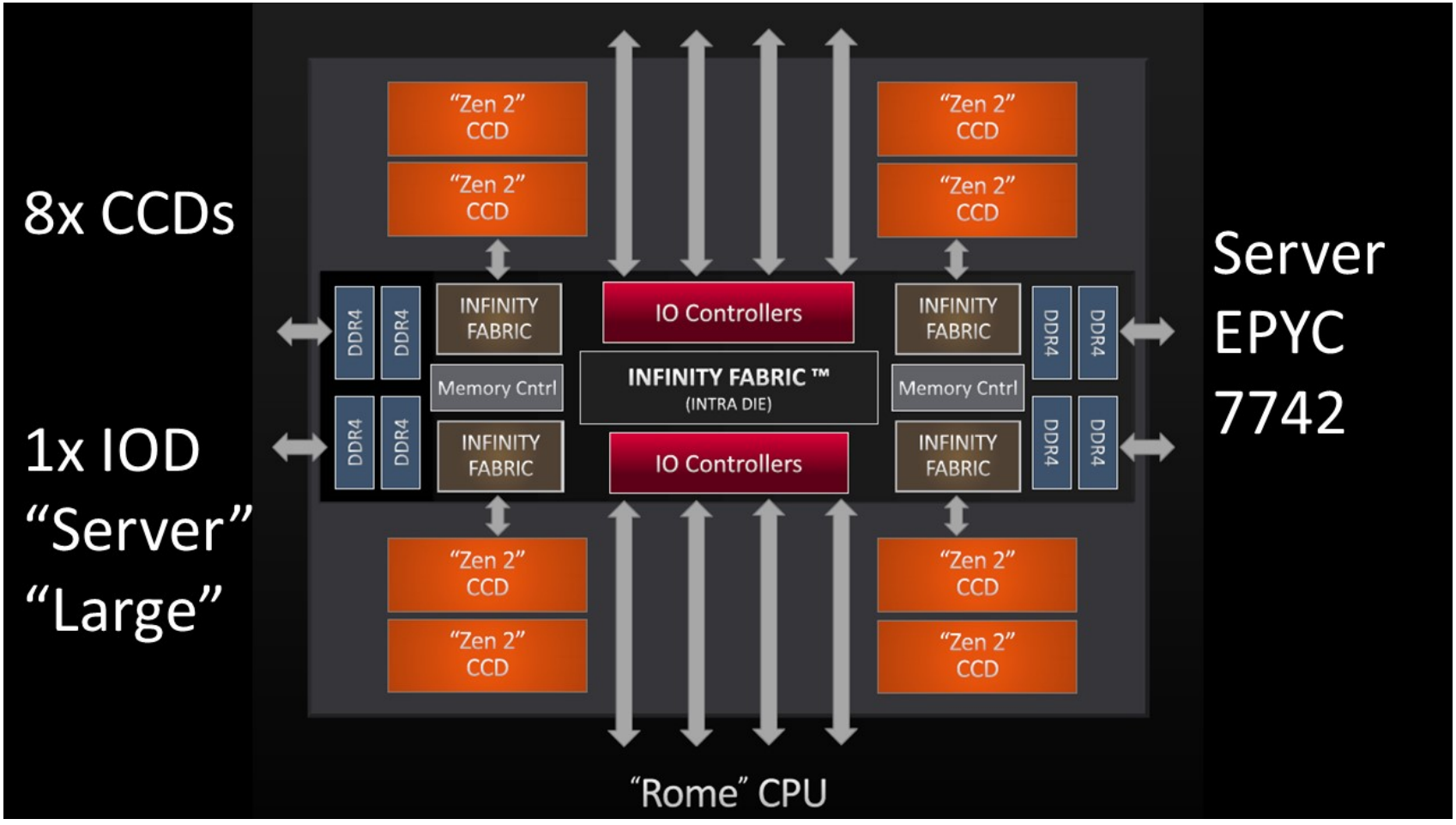


A multi-core processor is a computer processor with two or more separate central processing units or cores, each of which reads and executes program instructions, as if the computer had several single-core uniprocessors.

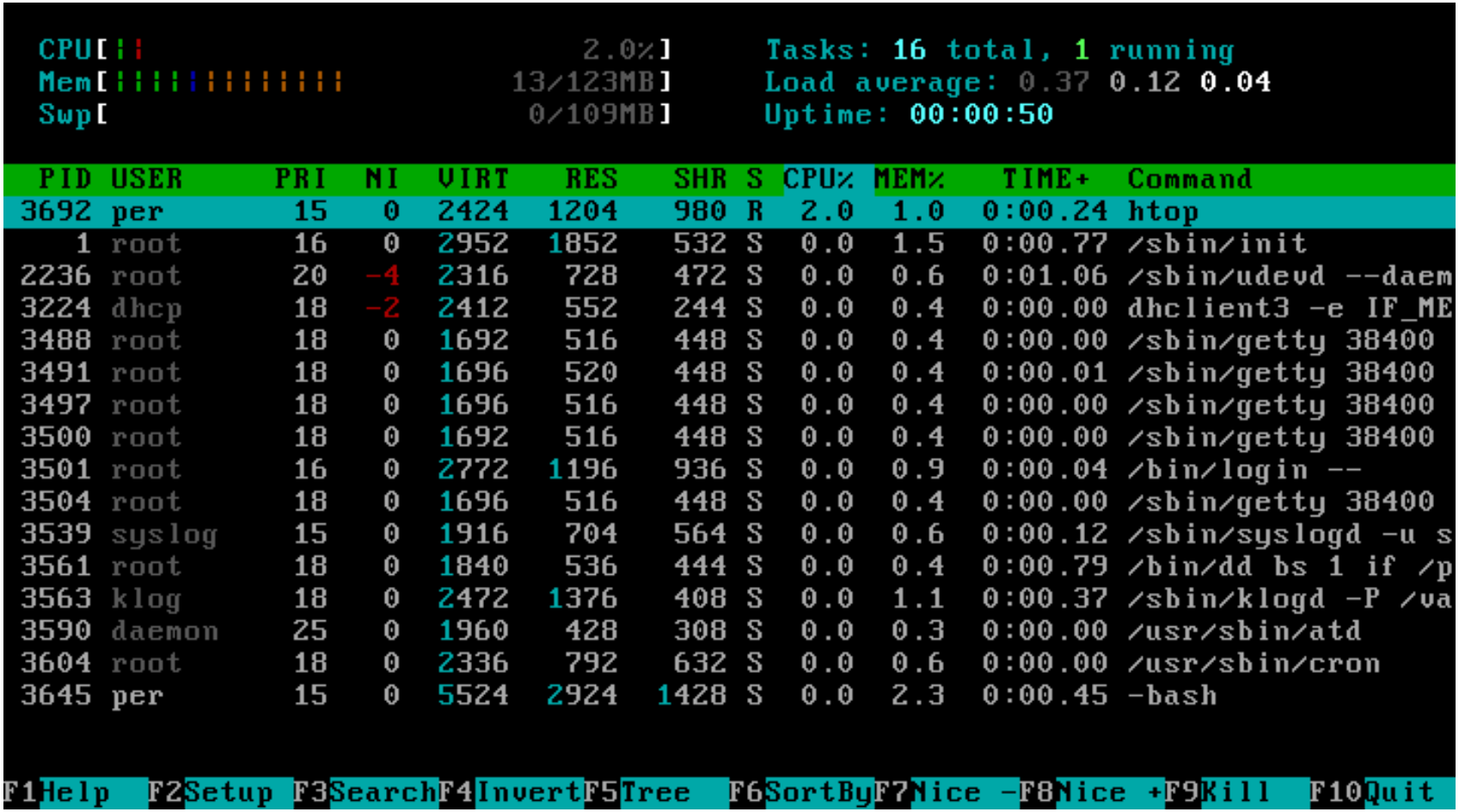
How many cores do you want with that?



E X P A N S E - AMD EPYC (ROME) 7742

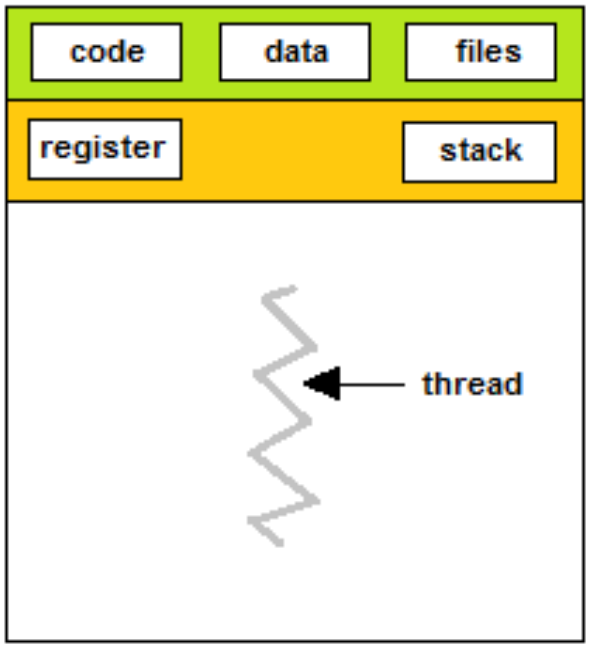


What is a Process?

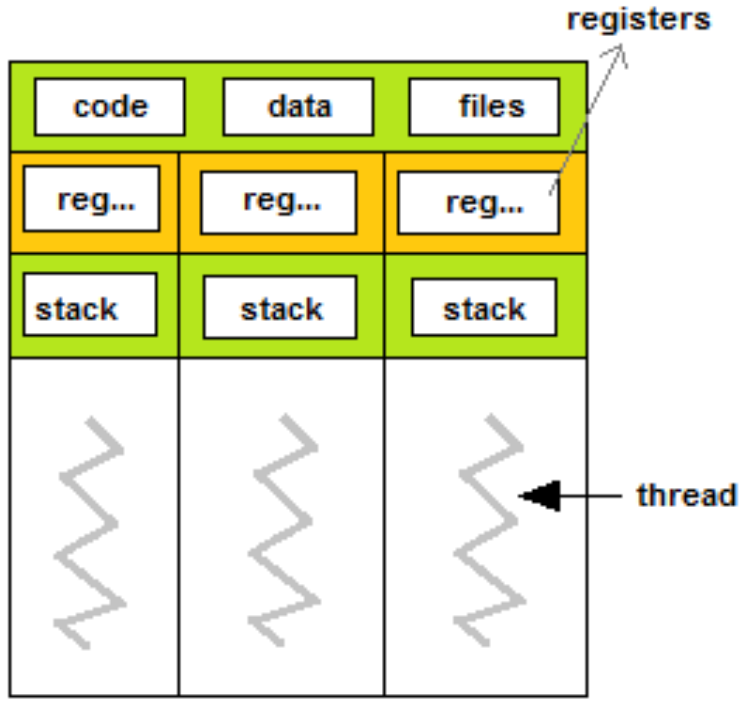


A process is the instance of a computer program that is being executed by one or more *threads*.

What is a Thread?



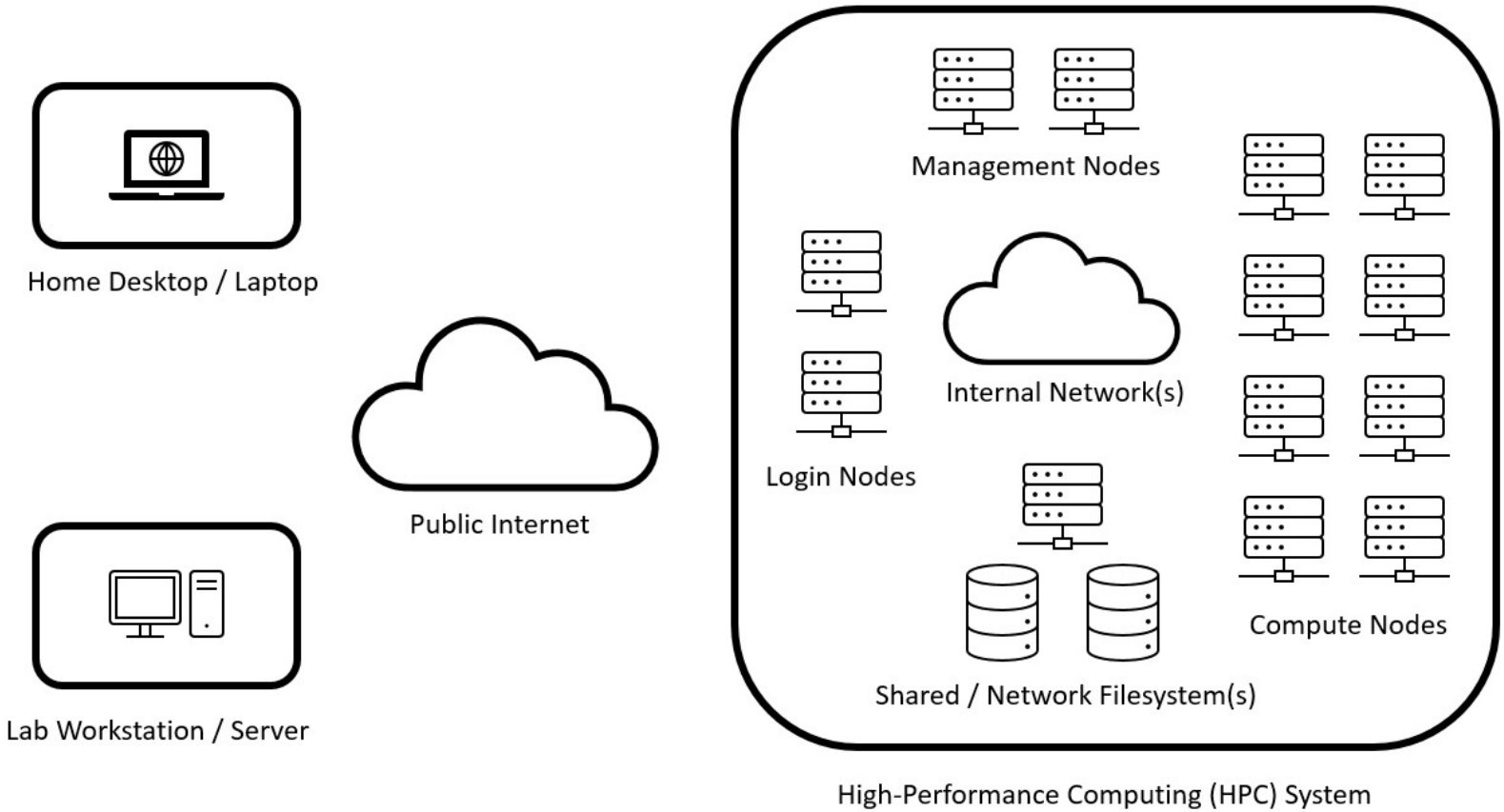
single-threaded process



multithreaded process

A thread is a subset of program instructions within a process that can be managed and run independently by an operating system's scheduler.

Your Standard HPC Ecosystem



Hilbert Matrix

A Hilbert matrix is a square matrix with elements that are unit fractions given by

$$H_{ij} = \frac{1}{i + j - 1}$$

For example, the Hilbert matrix of dimension 4 is

$$\mathrm{H} = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{bmatrix}$$

Exercise: Matrix-Vector Multiplication

If H is a Hilbert matrix of dimension $n = 122880$ and $\mathrm{x} = [1 \cdots 1]^T$ is an all-ones vector of the same dimension, compute the resultant vector $\mathrm{y} = \mathrm{Hx}$.

Check your result by computing the sum of the elements of y. For H and x of dimension n, the sum of the elements of y is given analytically by

$$\sum_{i=1}^{n} y_i = n + \sum_{j=1}^{n-1} \frac{n-j}{n+j}.$$

Extra credit: Optimize your code and recompute for both $n = 122880$ and $n = 1048576$. Plot the parallel speedup and efficiency of your code.

References

- Introduction to Parallel Computing
  https://hpc.llnl.gov/documentation/tutorials/introduction-parallel-computing-tutorial

- OpenMP Tutorial
  https://hpc-tutorials.llnl.gov/openmp

- An Intro to Threading in Python
  https://realpython.com/intro-to-python-threading

Questions?