Repealing Amdahl's Law

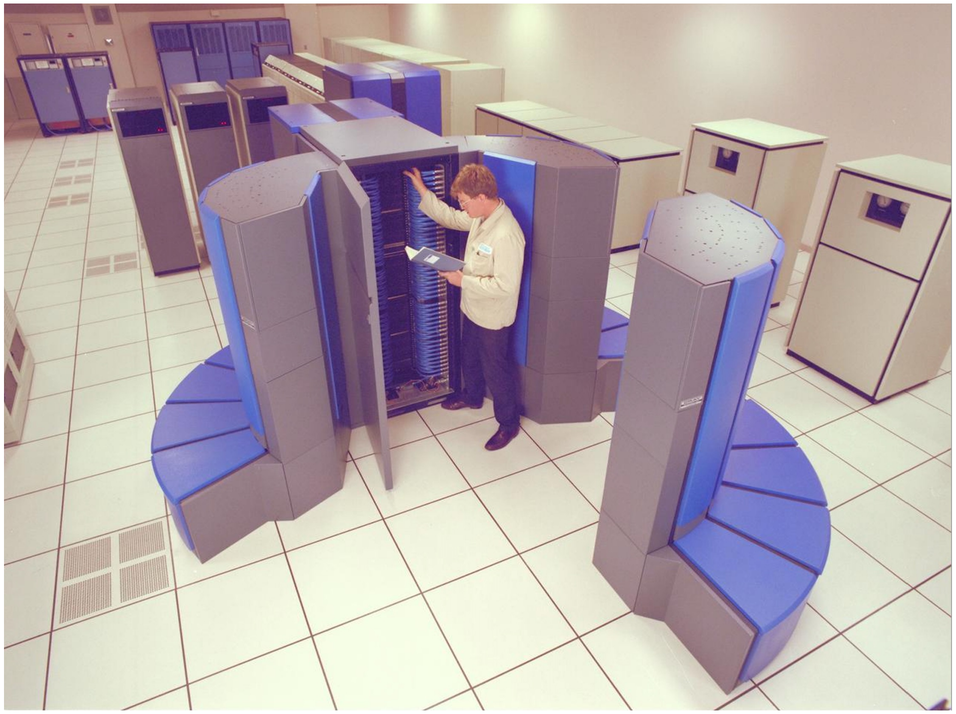
(And maybe replace it with something better?)

Danny Hillis

Applied Invention

40TH Anniversary Lecture Series

Back in the 1980's,
most supercomputers looked like this:



Supercomputers were super because they had such fast clock speeds

- Clock speed of Cray X-MP was 104 MHz
- Most supercomputers had 1 to 4 CPUs
- Most supercomputer users believed that more CPUs would lead to diminishing returns
- Massively parallel computing was not considered useful for most supercomputing tasks

I was trying to convince people to switch to supercomputers that looked like this:

Connection Machines

- Many processors connected by a fast communication network
  - CM-1 SIMD, 64,536 1-bit CPUs
  - CM-2 SIMD, 64,536 1-bit CPUs + 1024 64-bit FPUs
  - CM-5 MIMD, 1024 64-bit CPUs + 1024 64-bit FPUs
- Hardware for map/reduce to support ”data parallel” programming

Side comment: The distinction between SIMD and MIMD is not so important as it may seem, since each can emulate the other in proportional time

I had a tough time convincing the customers

"If you were plowing a field, which would you rather use: two strong oxen or 1024 chickens?"

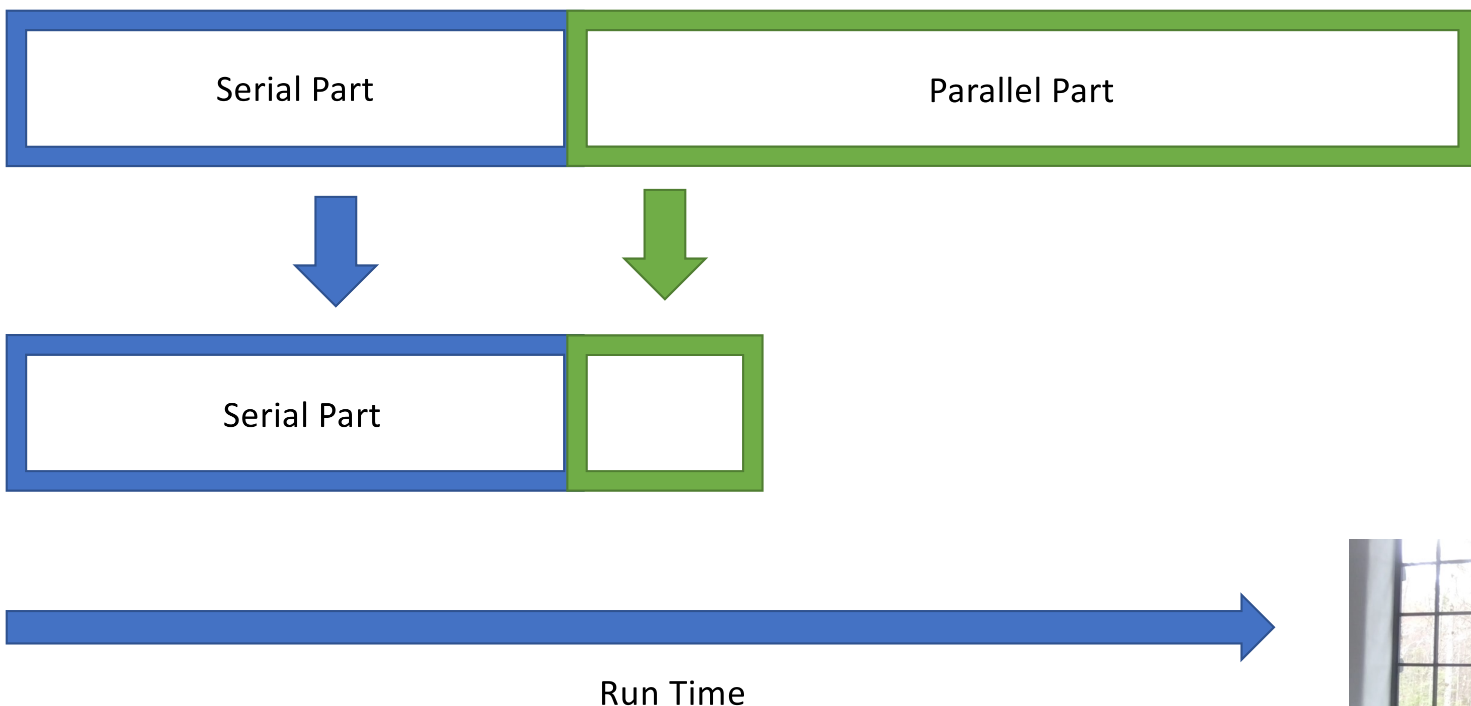
\- Seymour Cray

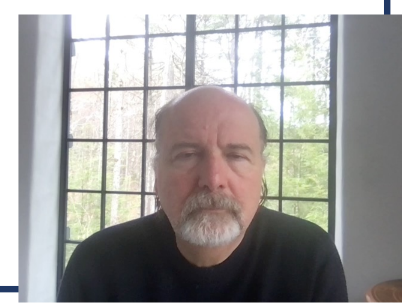
Customers believed in "Amdahl's Law"

For a calculation with parallel portion P and N processors

$$\text{Speedup} = \frac{1}{(1 - \text{P}) + \frac{\text{P}}{\text{N}}}$$

Gene Amdahl’s insight

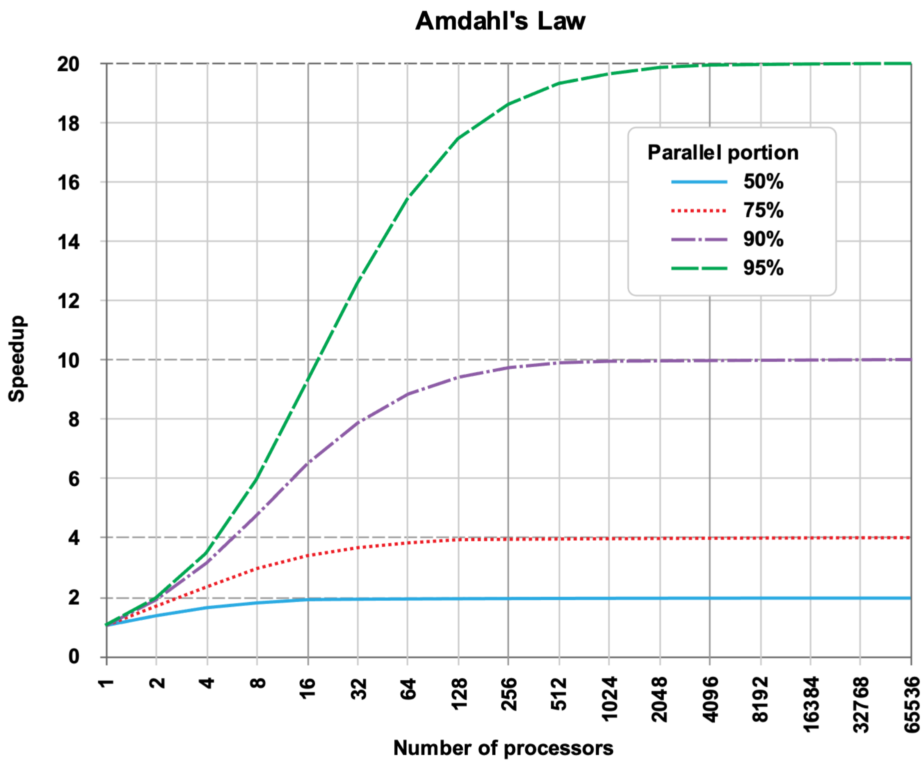




Amdahl's Example

- Solving a 3-dimensional finite difference model
- Assume a grid of N x N x N
- Potential parallelism in volume calculation is order N^3
- Potential parallelism in surface boundary calculation is order N^2
- Potential parallelism in corner case calculations is order 1
- I/O is serial

Seemed like parallelism wouldn't help much



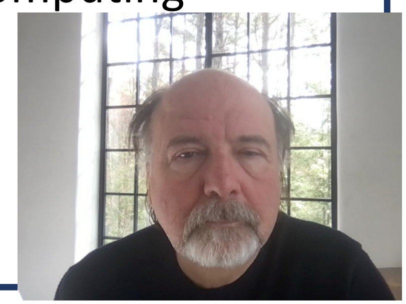
So, what was wrong with the reasoning?

- Amdahl assumed the portion of the calculation that is potential parallel is fixed
- In reality, faster computers are used to solve bigger problems

What determines the size of the problem?

- In data processing, the size of of the data that needs to be processed determines the size of the computation
  - The processors are pulled as needed from a cost-optimized pool (cloud)

- In supercomputing, the user wants the best possible result from the available computing resources
  - a bigger computation typically leads to better accuracy, a better optimization, or a better chance finding a solution, so the amount of available computing resources determines the size of the computation



In supercomputing, the run time is not determined by the size of the problem…

the size of the problem is determined by the run time

So, is there a more relevant law?

assumptions:

- The number of processor available N
- The time available is $T_{available}$
- The serial part requires time T_{serial}
- A calculation of size k can split into k parallel parts
- Each part can be complete in time $T_{parallel}$
- $k \geq N$

Is it realistic to assume k $\geq N$?

- The opportunity for parallelism grows with the amount data or the size of the search space
- Example: Amdahl's N×N×N 3D grid
  - scale N by 100
  - N^3 volume scales by 10^6
  - Even boundary scales by 10^4
- So, yes, the assumption is usually realistic

Proposed replacement for Amdahl's Law

$$Size \le \frac{(T_{available} - T_{serial})}{T_{parallel}} N$$

N is number of processors
Size is number of parallel parts
$T_{parallel}$ is max time for each of the parallel parts
T_{serial} is time for serial part

Equality can be achieved when *Size* is an exact multiple of *N*

Algorithms
Architecture
Applications

40TH Anniversary Lecture Series