Parallel Computing with GPUs

Introduction Part 2 - Supercomputing and Software

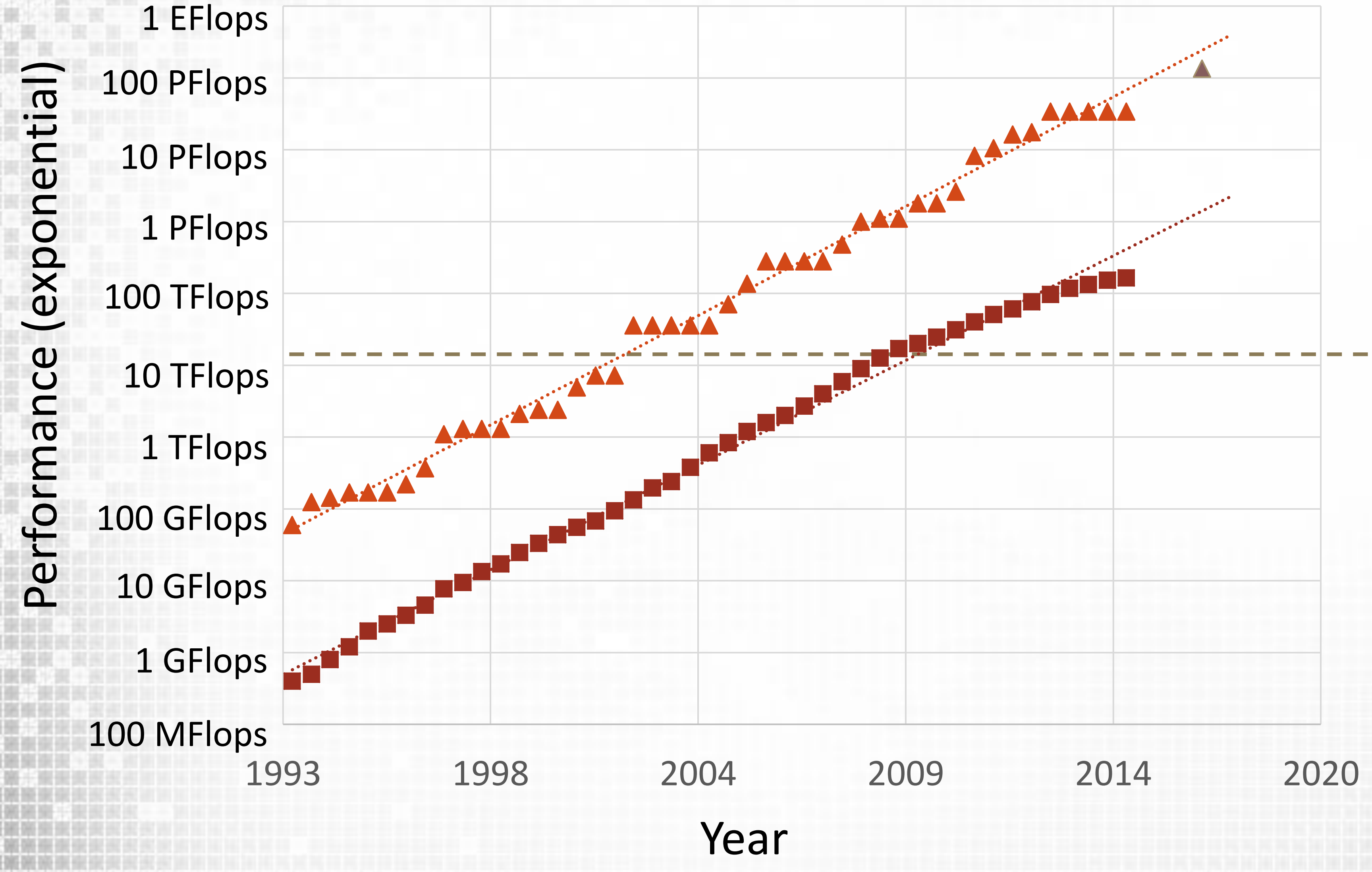
The University Of Sheffield.

Dr Paul Richmond
http://paulrichmond.shef.ac.uk/teaching/COM4521/

This Lecture (learning objectives)

- ❑ Supercomputing
  - ❑ Analyse High Performance Computing (HPC) observations
  - ❑ Predict future hardware trends in HPC
  - ❑ Contrast types of super computing system
- ❑ Software
  - ❑ Explain the limitations of parallelism with respect to speedup
  - ❑ Classify programming models and types of parallelism

Top Supercomputers



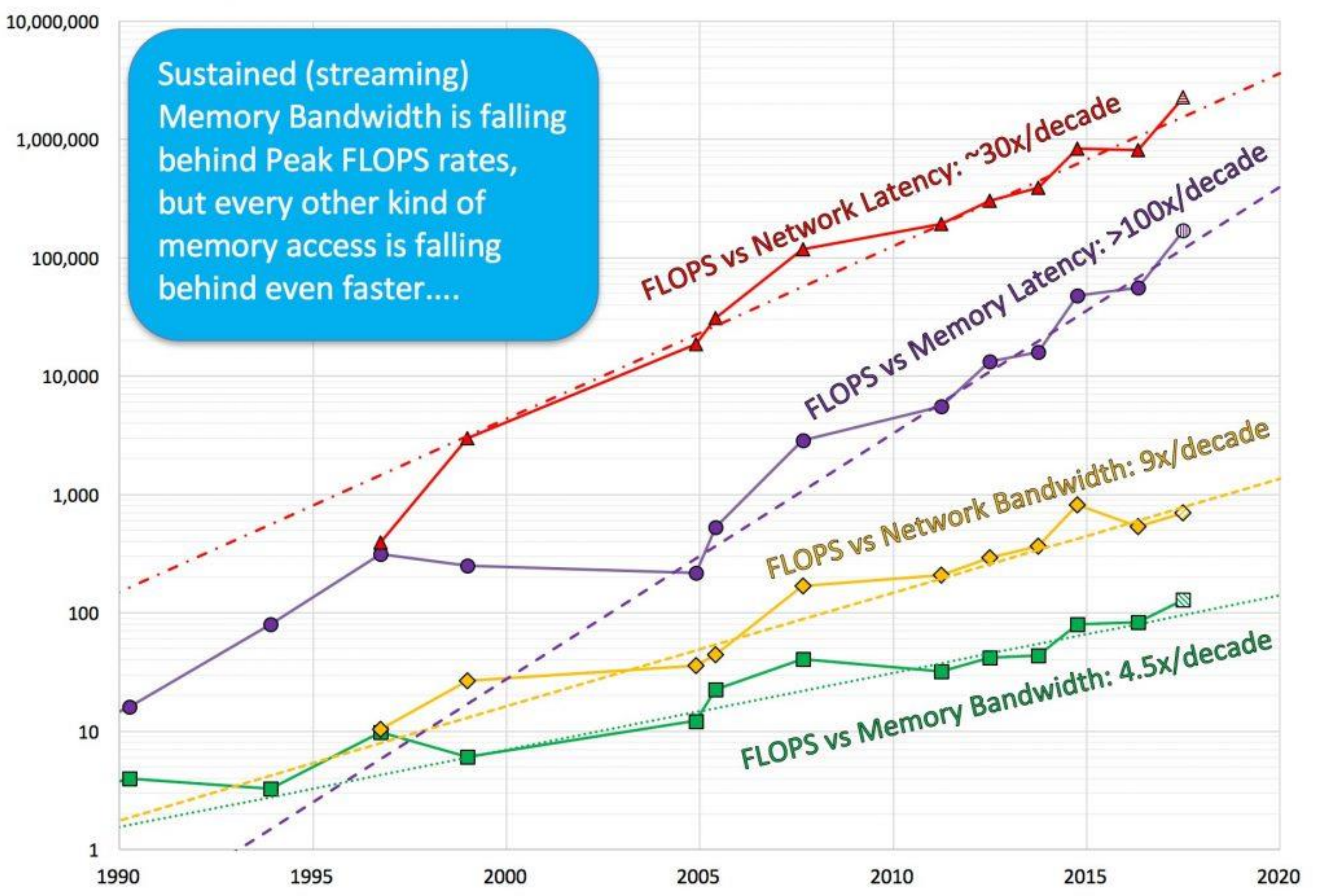
Volta V100 (15TFLOPS SP)

Supercomputing Observations

- Exascale computing
  - 1 Exaflop = 1M Gigaflops
  - Estimated for mid 2020s
- Pace of change
  - Desktop GPU top supercomputer in 2002
  - A desktop with a GPU would be in Top 500 in 2008
  - A Teraflop of performance took 1MW in 2000
- Extrapolating the trend
  - Current gen top500 on every desktop in < 10 years

HPC Observations

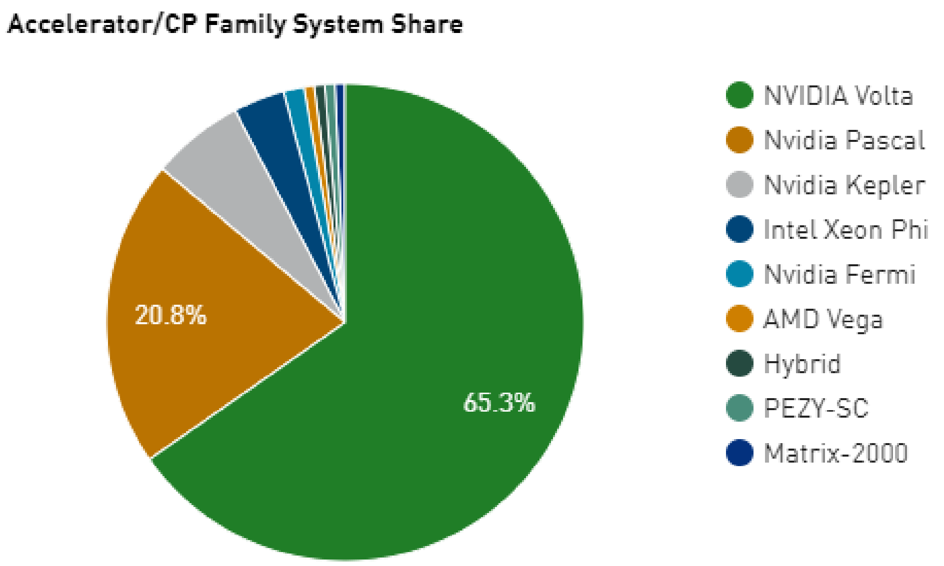
- ❑Improvements at individual computer node level are greatest
  - ❑Better parallelism
  - ❑Hybrid processing
  - ❑3D fabrication
- ❑Communication costs are increasing
  - ❑Memory per core is reducing
- ❑Throughput > Latency



http://sc16.supercomputing.org/2016/10/07/sc16-invited-talk-spotlight-dr-john-d-mccalpin-presents-memory-bandwidth-system-balance-hpc-systems/

Supercomputing Observations

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	2,414,592	148,600.0	200,794.9	10,096
2	DOE/NNSA/LLNL United States	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM / NVIDIA / Mellanox	1,572,480	94,640.0	125,712.0	7,438
3	National Supercomputing Center in Wuxi China	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
4	National Super Computer Center in Guangzhou China	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 NUDT	4,981,760	61,444.5	100,678.7	18,482
5	Texas Advanced Computing Center/Univ. of Texas United States	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR Dell EMC	448,448	23,516.4	38,745.9	
6	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 Cray/HPE	387,872	21,230.0	27,154.3	2,384
7	DOE/NNSA/LANL/SNL United States	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect Cray/HPE	979,072	20,158.7	41,461.2	7,578
8	National Institute of Advanced Industrial Science and Technology (AIST) Japan	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR Fujitsu	391,680	19,880.0	32,576.6	1,649



Green 500

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	159	**A64FX prototype** - Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, Tofu interconnect D , Fujitsu Fujitsu Numazu Plant Japan	36,864	1,999.5	118	16.876
2	420	**NA-1** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	1,271,040	1,303.2	80	16.256
3	24	**AiMOS** - IBM Power System AC922, IBM POWER9 20C 3.45GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100 , IBM Rensselaer Polytechnic Institute Center for Computational Innovations (CCI) United States	130,000	8,045.0	510	15.771
4	373	**Satori** - IBM Power System AC922, IBM POWER9 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100 SXM2 , IBM MIT/MGHPCC Holyoke, MA United States	23,040	1,464.0	94	15.574
5	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	10,096	14.719
6	8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	1,649	14.423
7	494	**MareNostrum P9 CTE** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM Barcelona Supercomputing Center Spain	18,360	1,145.0	81	14.131
8	23	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 , HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	792	13.704

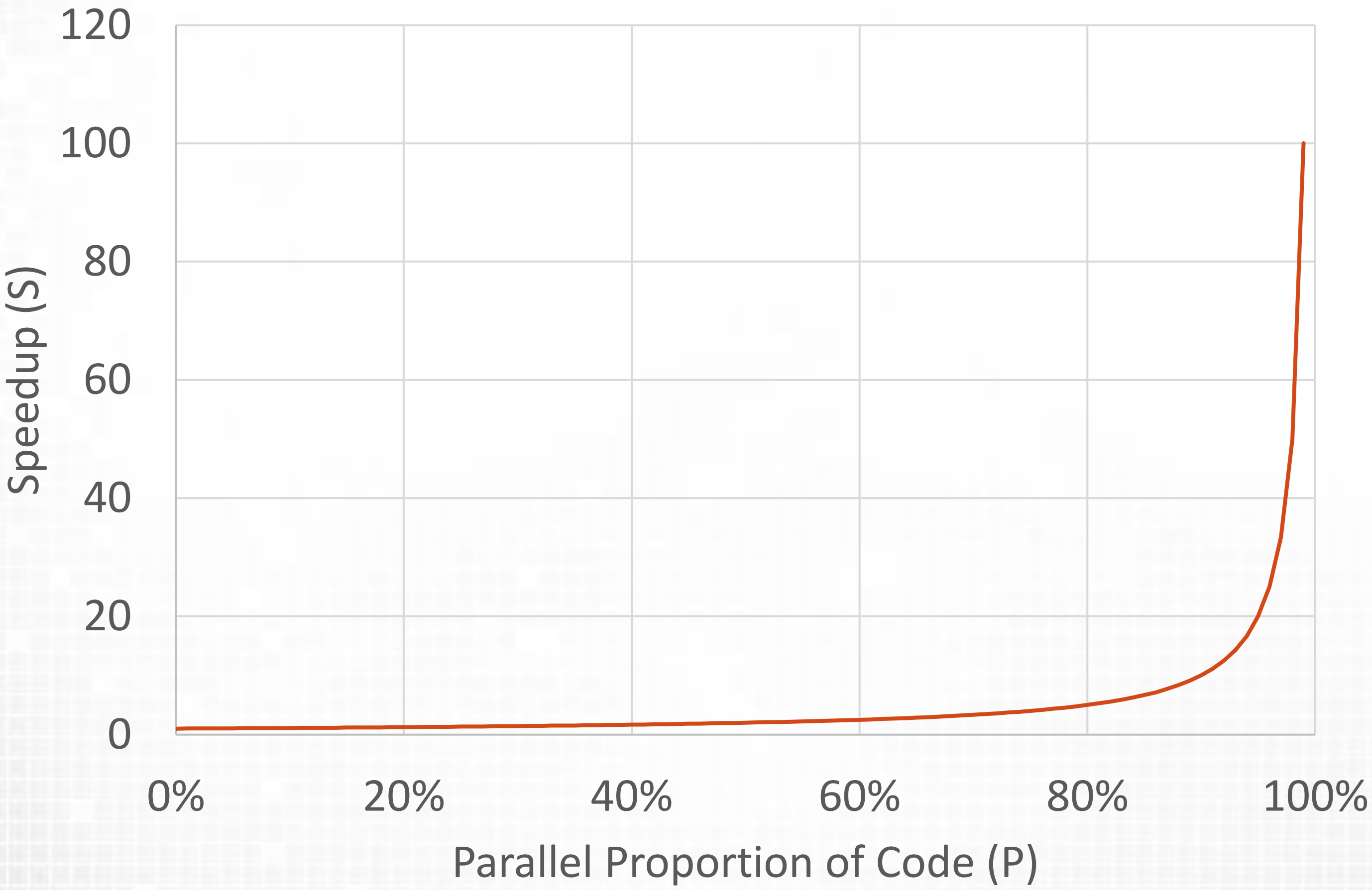
- ❑ Top energy efficient supercomputers

Software Challenge

- How to use this hardware efficiently?
- Software approaches
  - Parallel languages: some limited impact but not as flexible as sequential programming
  - Automatic parallelisation of serial code: >30 years of research hasn't solved this yet
  - **Design software with many core parallelisation in mind**

Amdahl's Law

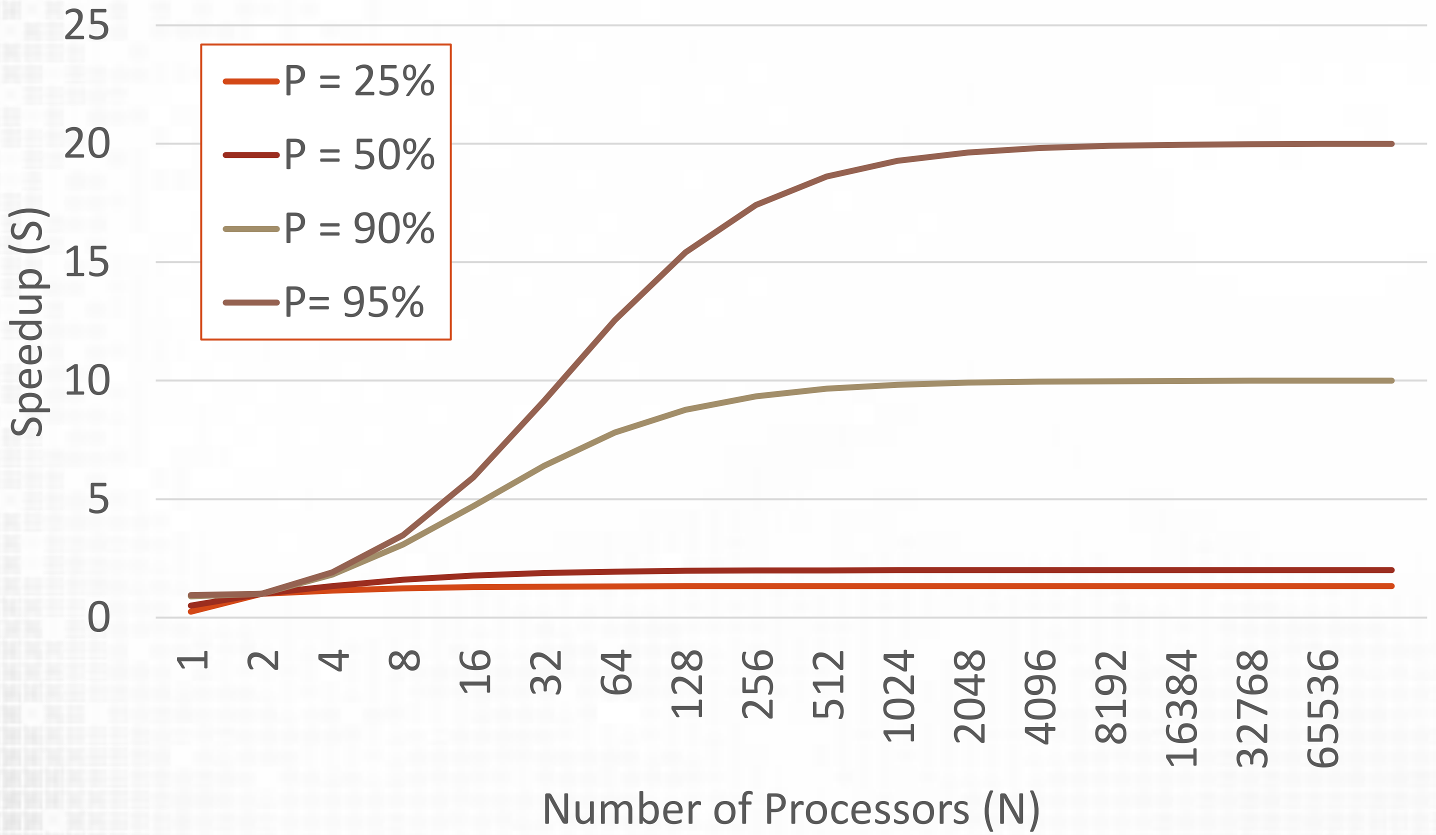
- Speedup of a program is limited by the proportion than can be parallelised



$$Speedup\ (S) = \frac{1}{1 - P}$$

Amdahl's Law cont.

- ❑Addition of processing cores gives diminishing returns



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

Parallel Programming Models

- Distributed Memory
  - Geographically distributed processors (clusters)
  - Information exchanged via messages
- Shared Memory
  - Independent tasks share memory space
  - Asynchronous memory access
  - Serialisation and synchronisation to ensure correctness
  - No clear ownership of data
  - Not necessarily performance oriented

Types of Parallelism

- Bit-level
  - Parallelism over size of word, 8, 16, 32, or 64 bit.
- Instruction Level (ILP)
  - Pipelining
- Task Parallel
  - Program consists of many independent tasks
  - Tasks execute on asynchronous cores
- Data Parallel
  - Program has many similar threads of execution
  - Each thread performs the same behaviour on different data

The University Of Sheffield.

Summary

- Supercomputing
  - Analyse High Performance Computing (HPC) observations
  - Predict future hardware trends in HPC
  - Contrast types of super computing system
- Software
  - Explain the limitations of parallelism with respect to speedup
  - Classify programming models and types of parallelism

- Next Lecture: Module Details