



High Performance Computing Glossary

<http://contracosta.edu/hpc>

Amdahl's Law

If F is the fraction of a calculation that is sequential, and $(1-F)$ is the fraction that can be parallelized, then the maximum speedup that can be achieved by using P processors is $1/(F+(1-F)/P)$. The effect is that the sequential part of a program is the part that tends to dominate. For instance, we know a car traveling at 50 MPH takes one hour to go 50 miles. What is not immediately obvious, is if we could somehow have the car travel infinitely fast for the first 25 miles, the trip will take 1/2 hour.

Bandwidth

Describes the rate at which a network can transmit information, typically measured in either "megabits per second" (Mbit/sec, Mb/s, Mbs) or "megabytes per second" (Mbyte/sec, MB/s, MBs). Communication networks are often rated in terms of bandwidth and latency. Common bandwidths are Ethernet, Fast Ethernet, and Gigabit Ethernet.

Batch

A method of processing a series of commands on a computer with no human interaction. Typically, a list of commands is placed within a file, and then that file is executed.

Cache

Very fast, expensive memory, local to a given processor and "close to" the CPU. Cache memory contains copies of data recently accessed from main memory, allowing fast reuse of data in calculations. The fastest cache memory is designated "level 1" (L1); computers may also have slightly slower "level 2" (L2) cache (typically off-chip) and even "level 3" (L3).

CISC (Complex Instruction Set Computers)

A category of computer processors based around a set of fairly high-level, powerful instructions, as contrasted with RISC.

CPU (Central Processing Unit)

The main processing component in a computer.

Ethernet

Networking standards originally developed in 1973 by Xerox and formalized in 1980 by DEC, Intel, and Xerox which transmits data at a theoretical peak of 10 Mbps. It is the most popular Local Area Network (LAN) technology in use today.

Fast Ethernet

An Ethernet networking system which transmits data up to 100 Mbps; theoretically 10 times fast than Ethernet.

Flops

Acronym for floating point operations per second. The term "floating point" refers to that format of numbers that is most commonly used for scientific calculation. Flops is used as a measure of a computing system's speed of performing basic arithmetic operations such as adding, subtracting, multiplying, or dividing two numbers.

Gigabit Ethernet (GbE or gige)

Gigabit Ethernet, defined in 1998, is a version of Ethernet, which supports data transfer rates of up to 1 Gigabit (1,000 Mbs).

Gigaflops (Gflops) 1 billion flops.

IPC (Interprocess Communication)

A mechanism which allows one process to communicate with another process. The processes can be running on the same computer or on different computers connected through a network. IPC enables one application to control another application, and for several applications to share the same data without interfering with one another.

Local Area Network (LAN)

A network that is in one location and that allows users to share resources such as file storage and print services.

Latency

The time taken to start up an operation. Typically, message latency is the time delay incurred between one processor starting a message send operation, and the recipient processor completing the receive operation. Startup latency is the constant communication overhead incurred in sending a zero length message. Communication networks are often rated in terms of latency and bandwidth. Typical values of startup latency: 1 milliseconds between workstations connected by Ethernet, and 40 microseconds using MPI on the Cray class computer.

Linear Speedup

Speedup that is directly proportional to the number of parallel processors used. Linear speedup is the ideal for any parallel program, since Amdahl's Law restricts speedup based on the amount of the



program that is still serial. Superlinear speedup is sometimes achievable due to other factors related to the processor architecture (See superscalar below).

Load balance

A measure of how evenly work is distributed among a set of parallel processors. Parallel programs are most efficient when the load is perfectly balanced, i.e. each processor has exactly the same amount of work to do.

Locality (of data)

A measure of how much the calculation of one piece of data depends on other pieces “close to it”. Most easily thought about for grid-based computations, where to update the values at one point in the grid it might be necessary to know the values at each of its neighboring points. This would be a “local” calculation, of which the Game of Life is a classic example. If, to update a given point, it were necessary to know the values of the data at all the other points, this would be an extremely “non-local” update (in fact, maximally bad).

Megaflops (Mflops)- 1 million flops.

MPI (Message Passing Interface)

A standard software method for message passing in parallel applications. The MPI defines a library of routines through which message-passing processes communicate. [<http://www.mpi-forum.org>].

Multiprocessor

A computer with more than one processor but a single address space.

Multi-tasking

Executing more than one process on a single processor, usually achieved by timesharing under the control of the operating system.

Parallel Computer

A computer having multiple CPUs.

Pipelining

Overlapping the execution of two or more operations. Pipelining typically involves starting one operation, then immediately starting the other without having to wait for the first to complete. It applies the principle of the multi-stage factory production line to computer processors.

Process

An executing program. The term is used loosely as a synonym of task. Note that the user may be running an application which is actually comprised of many cooperating processes.

RISC (Reduced Instruction Set Computers)

A category of computer processors based around a small set of low-level, highly optimized instructions, as contrasted with CISC processors.

Scalable

Of parallel programs, delivering an increase in performance with a corresponding increase in size. Typically, this means “more processors equals better (faster) performance”. For a cautionary tale, see Amdahl’s Law above.

Scalar processors

Essentially the opposite of vector processor, one in which only one operation can be carried out at a time.

Supercomputer

Like a High Performance Computer, but often reserved for vector computers.

Superscalar

Of a computer processor, being able to schedule operations for side-by-side execution. Typically, a superscalar processor can schedule an integer operation, a floating point operation and a memory operation for simultaneous execution in different pieces of hardware. Superscalar chips also make heavy use of pipelining. Superscalar is similar to VLIW.

Teraflops (Tflops) - 1 trillion flops.

Vector Computer

A computer implementing instructions that work on a vector of data at once. This may be implemented as a pipeline (each operation is further decomposed into steps that each run on their own hardware). Vector supercomputers do not have a cache, but have very fast main memory instead, consequently outperforming commodity processors.

VLIW (Very Long Instruction Word)

A microprocessor architecture different from CISC and RISC. VLIW uses encoded instructions that are even more complex than the instructions used in CISC. While RISC took instruction sets in a simpler direction by making instructions smaller, VLIW moves them in a more complex direction by encoding four or more instructions into a single operation. Superscalar is similar to VLIW.

WAN (Wide Area Network)

A geographically dispersed telecommunications network. The term distinguishes a broader telecommunication structure from a local area network. A wide area network may be privately owned or rented, but the term usually connotes the inclusion of public (shared user) networks.

Workstation

A networked desktop computer running a modern Multi-tasking operating system such as Unix or MS Windows.

