

# Glossary

- ABI** **A**pplication **B**inary **I**nterface. The compiled-code implementation of an API.
- Amdahl's law** A simple formula to compute the maximum performance of a parallel code depending on the degree of parallelism available in the hardware and the serial fraction of the code. In reality, it is an overly optimistic upper bound.
- API** **A**pplication **P**rogramming **I**nterface. A stable interface from a program to an underlying layer of software, be that a library or the OS itself.
- ASIC** **A**pplication-**S**pecific **I**ntegrated **C**ircuit. Die with an integrated circuit that is specifically geared towards a specific function or application, not intended for general-purpose usage.
- BasicLockable** A named requirement in the C++ standard which describes a class that can be used via `std::lock_guard`.
- cache coherence** The desirable property that all caches in a machine maintain the same value for any specific cache line.
- child task** A task that has been created by another task, the parent task.
- code** synonym for program or application.
- coherence fabric** The interconnect used to carry the messages used by cores to maintain cache coherence.
- compilation** The act of converting a human-readable representation of a program into a set of machine instructions (and data) which can be executed by a computer.
- core** see “physical core”.
- CPI** **C**ycles **p**er **I**nstruction. One way to measure the efficiency of execution of a code on a specific CPU implementation.
- CPU** **C**entral **P**rocessing **U**nit. See “core”. Though in some contexts people use CPU to mean the physical package that holds multiple cores, cache, etc.
- descendant task** A task that is a child task or a child of a descendant task.
- DMA** **D**irect **M**emory **A**ccess. A technique that allows devices such as network interfaces or disk controllers to access memory directly without requiring that the CPU moves every byte of data to or from the device.
- false sharing** The sharing which occurs because a single physical cache line is being used to hold two unrelated variables. False sharing causes cache-to-cache transfers, which can impact performance yet is hard to see in source code because the mapping from variables to locations in store is controlled by the compiler.
- futex** A Linux<sup>\*</sup> system call used to suspend threads or to wake them.
- hardware thread** See “logical core”.
- initial thread** See “main thread”.

**IP** **Internet Protocol.** The low-level protocol which allows different machines to communicate over the internet.

**IPC** **Inter-Process Communication.** Any method used to communicate between separate processes (which may be on the same machine, or remote). Sometimes also used for **Instructions Per Cycle**, i. e.,  $1/CPI$ , although we do not use that here.

**JIT** **Just In Time.** JIT compilation is an implementation in which compilation of a program occurs as it is executing. This avoids the need for any offline compilation and allows the compiler to know the exact machine for which it is compiling, as well as exploiting profile information to make good optimization choices. The cost is that the compilation has to happen every time the code is run.

**jitter** Noise injected into the system by the environment or other processes and threads executing on the same system.

**LILO time** **Last In to Last Out** time; for a collective operation, the elapsed wallclock time between the last thread arriving and the last thread leaving. For a tree barrier the LILO time will be the sum of the LIRO and RILO times of the join and broadcast operations.

**LIRO time** **Last In to Root Out** time; for a join operation where all threads arrive and a single, root thread leaves when it is aware that all have arrived, the elapsed wall-clock time between the last thread arriving and the root thread leaving.

**logical core** A single hardware thread of execution onto which the OS will schedule work. In an SMT machine a logical core is a single SMT thread, whereas a CPU or core would be able to execute multiple SMT threads simultaneously.

**logical CPU** Same as a “logical core”. The “logical CPU” terminology comes from the Linux OS.

**main thread** The thread which starts execution inside a process when the process is created.

**memory fence** In an out-of-order CPU, an instruction that enforces the externally visible ordering of memory instructions issued by the local core.

**OS** **Operating System.** The combination of an OS kernel and libraries which provide a way for user code to run on a machine while remaining abstracted from underlying hardware complexities.

**OS kernel** The part of an OS which executes in a privileged mode, controls devices, provides file systems, and allocates resources to user-level processes.

**package** See “processor package”.

**parallel fraction** The proportion of time in a serial execution of a code that could be executed in parallel.

**parent task** A task that has created another task.

**physical core** The hardware entity that executes instructions. In an SMT system a physical core may support more than one logical core. Sometimes also just called a core.

**process** The protection domain within which code executes. A process contains at least one thread.

**processor die** A slice of silicon that contains the processor plus supporting infrastructure like caches, memory controllers, etc. Plural “dice”.

**processor package** The physical unit that plugs into the socket of the system. Processor packages may contain multiple dice.

**recursion** See “recursion”.

**RILO time** **R**oot **I**n to **L**ast **O**ut time; for a broadcast operation, the elapsed wallclock time between the root thread arriving and the last thread leaving.

**root thread** In a centralizing collective operation (such as a centralizing barrier), the single thread which is the central thread, so that it knows all others have arrived.

**sched\_yield** A Linux system call that enters the kernel to request that it run the scheduler to choose a good thread to run. Do not use this in a polling loop; with modern Linux kernels calling `sched_yield()` here is futile.

**serial fraction** The proportion of time in a code that must be executed serially.

**sibling task** Tasks that have the same parent task.

**socket** The physical installation on the mainboard of the system to hold and connect the processor package and the mainboard of the computer. Sometimes the term is used as an alternative to “processor package”.

**task** A concurrent unit of work carrying its own data environment and code for execution in a thread.

**task pool** A container to hold tasks that are waiting for a thread to pick them up for execution.

**task queue** Misleading term for “task pool”.

**thread** A single entity of execution and its associated state. A thread executes inside a process.

**thread pool** A set of threads which are kept alive even when they have nothing to do so that they can be rapidly used when there is work for them.

**volatile** A C keyword that expresses that each load or store to a variable of this type must actually occur, preventing the compiler from optimizing them away if it thought they were unnecessary.

