Scheduling that allows context switches before a thread blocks, to prevent CPU "hogging".

An alternative to parallelism that subdivides tasks, saves their state, and returns to the top of the dispatch loop at every delay-inducing operation.

- Instruction-level
- Vector-level
- Thread-level

The first two are part of the implementation of a program, the third affects the program's high-level structure.

- Capturing the logical structure of a problem.
- Exploiting extra processors for speed.
- Coping with separate physical devices.

A section of code that multiple threads approach ("race to"), such that the behavior of the program depends on which gets there first.

To make a critical section appear atomic.

The difficulty in keeping copies of a memory location consistent with one another.
This problem affects shared-memory machines.

- Multiprocessor (shared-memory) systems.
- Message-based systems.

A well-defined unit of work that must be performed by some thread. In that sense a collection of threads share a "bag of tasks" to work on.

Lightweight threads have a shared memory space.