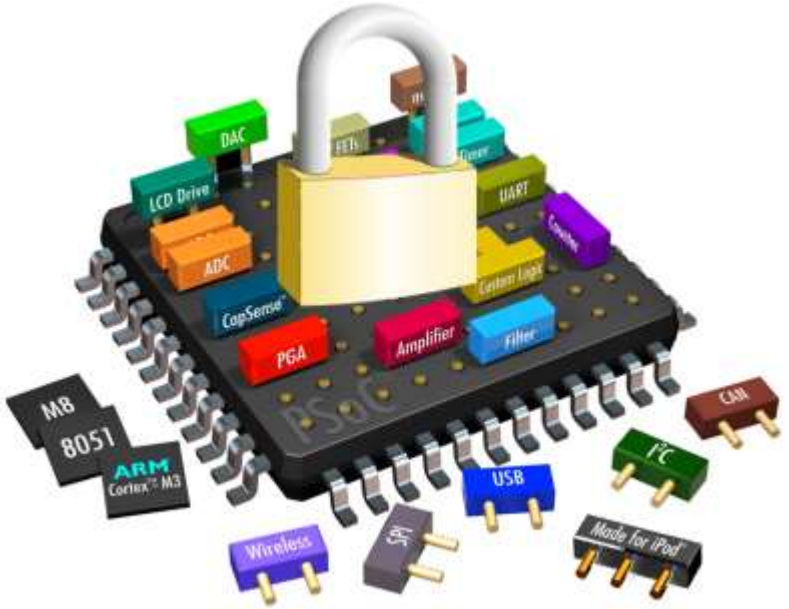
Advanced Computer Architecture

Data-Level Parallelism in Vector, SIMD, and GPU Architectures



Three Variations of SIMD Architecture

1

Vector Architectures

The oldest variation (by 30+ years), essentially pipelined execution of many data operations. Easier to understand and compile to than other SIMD variations, but historically considered too expensive for microprocessors until recently.

2

Multimedia SIMD Extensions

Found in most instruction set architectures today supporting multimedia applications. For x86, these include MMX (1996), followed by SSE versions, and continuing with AVX. Often necessary for achieving highest computation rates.

3

Graphics Processing Units (GPUs)

Offers higher potential performance than traditional multicore computers. While sharing features with vector architectures, GPUs have distinguishing characteristics due to their evolution in a heterogeneous ecosystem with system processors and separate memory.

For problems with significant data parallelism, all three SIMD variations are easier to program than classic parallel MIMD programming.

SIMD vs MIMD

- This figure assumes that two cores per chip for MIMD will be added every two years and the number of operations for SIMD will double every four years.
- For applications with both data-level parallelism and thread-level parallelism, the potential speedup in 2020 will be an order of magnitude higher than today, making understanding SIMD parallelism at least as important as MIMD parallelism.