

Pipelining's goal is a throughput of 1 instruction per cycle
data hazards & control hazards make this difficult

Scalar systems handle 1 instruction per stage
each stage requires one clock cycle
only one instruction is started each cycle
all instructions go through all stages
some instructions may take longer than without the pipeline
however the instruction throughput is increased

Superscalar systems aim for 2 or more instructions per cycle
extra hardware executes multiple instruction per cycle
multiple pipelines can operate in parallel
or each stage can process multiple instructions at one time

Multiple instructions are fetched at one time
requires a wider CPU-to-memory bus

Multiple instructions are decoded together
dependencies between the instructions are detected
independent instructions are issued to their execute units

Multiple execute units process instructions at the same time

- multiple ALUs for integer operations
- multiple floating point units
- load/store unit (to compute memory addresses)
- branch unit to analyze branch conditions & make predictions

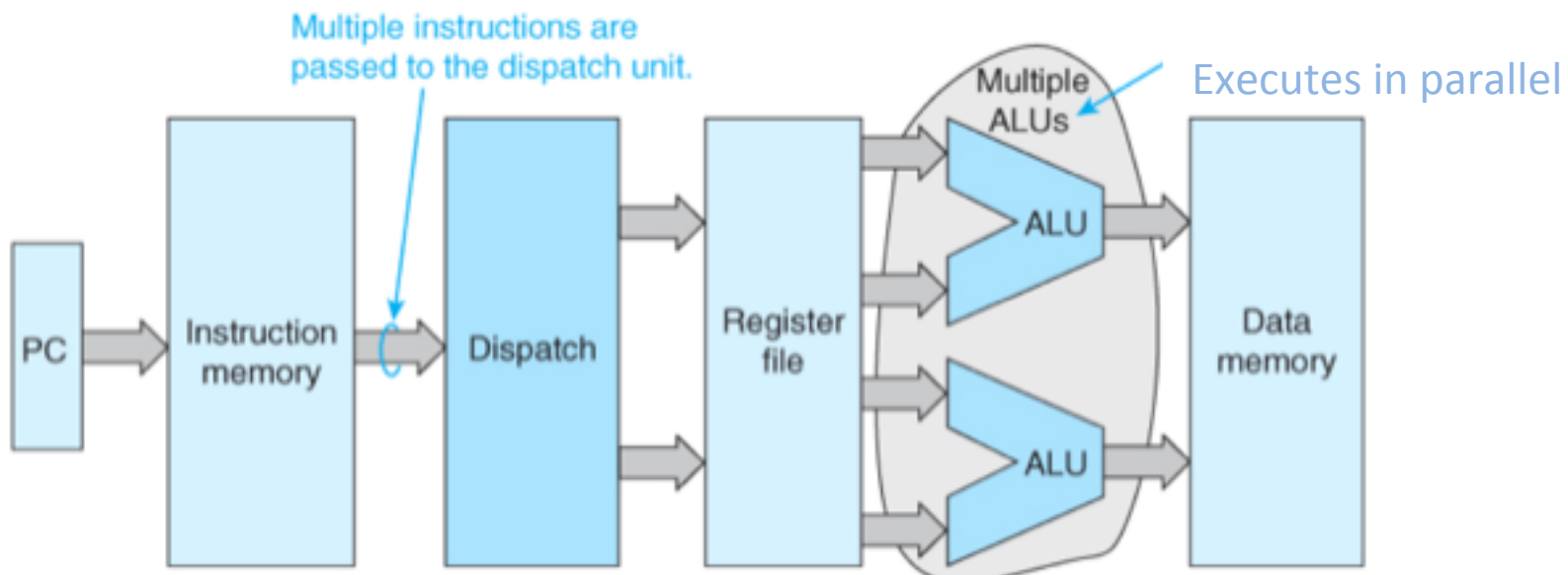
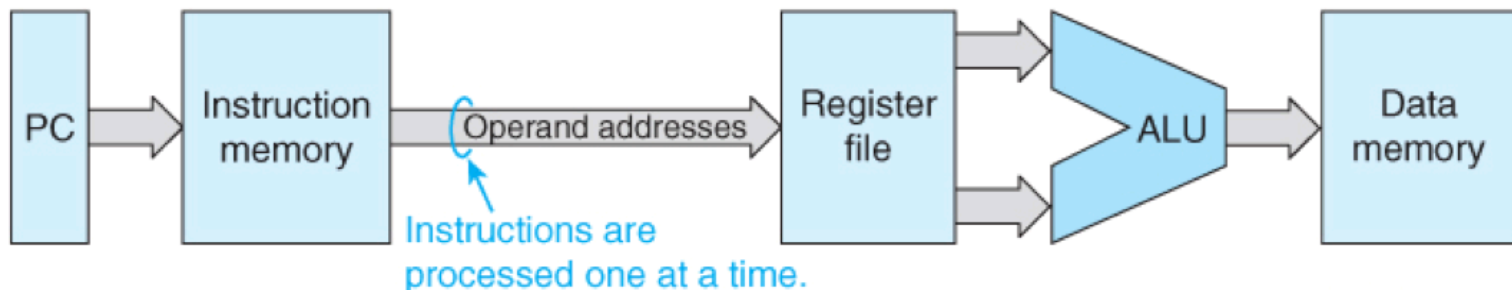
Resource Hazards are possible with superscalar systems

- the required unit may not be available for next instruction
- required registers may already be in use

Out-of-order execution may be needed for good performance

- instructions may start in a different order than in program
- later instructions may complete before earlier ones

Scalar Pipeline



Superscalar Pipeline

An m-way superscalar system has m parallel pipelines
m-fold increases in performance are seldom achieved

The Pentium P5 had 2 integer pipelines
the 2nd pipeline could only be used about 30% of the time

The original Pentium used superscalar operation
two 5-stage integer pipelines U and V
U pipe included a shifter not in the V pipe
a 6-stage floating point unit

Compilers generate machine instructions in program order

Superscalar processors *dispatch* multiple instructions in parallel
operands are obtained from multiplexed register files
Instructions that have operands can execute in parallel
The execute units needed must be available

The program semantics (meaning must be preserved)

Sequence 1	Sequence 2
add \$11,\$12,\$13	add \$11,\$12,\$13
add \$14,\$11,\$13	add \$15,\$16,\$17
add \$15,\$16,\$17	add \$14,\$11,\$13

Different instruction order but same meaning or net effect