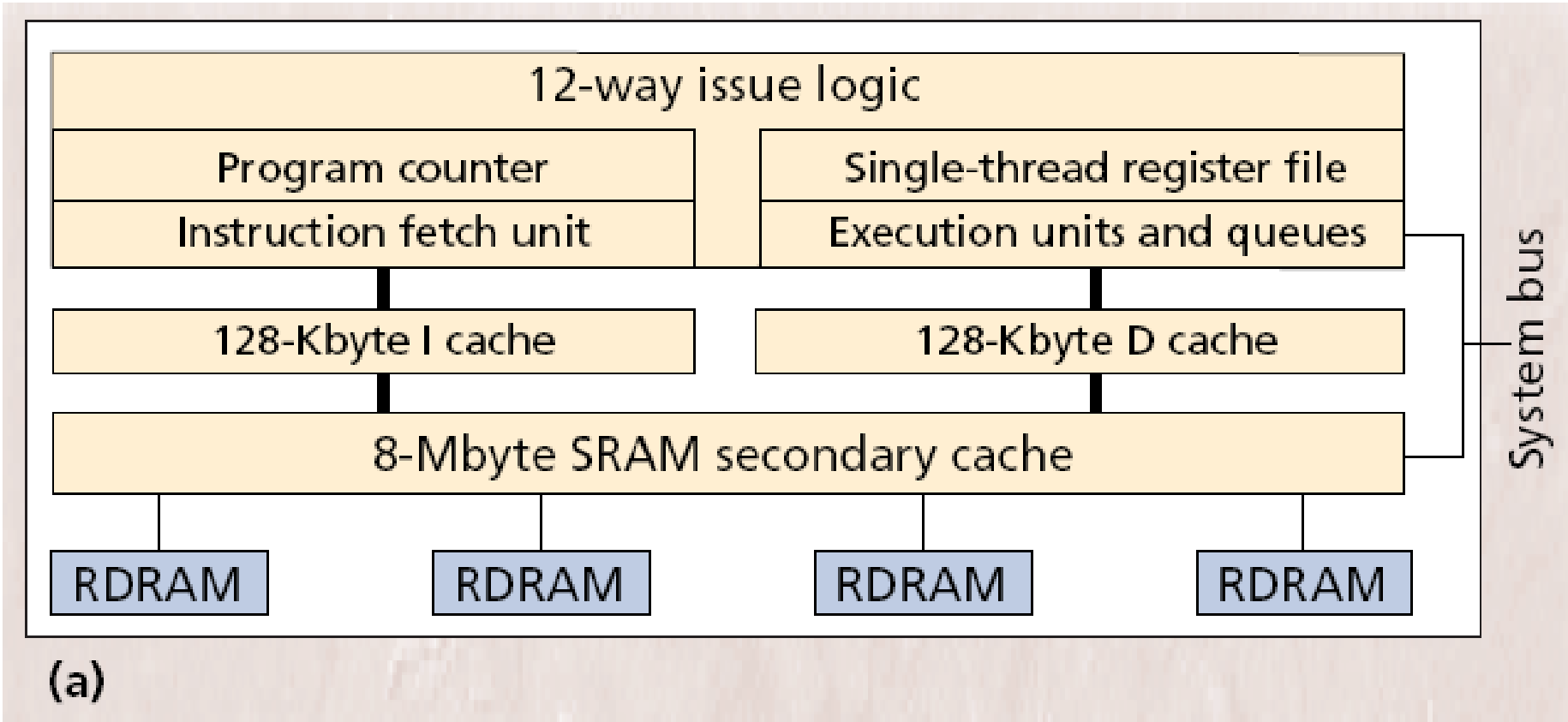
Chip Multiprocessors

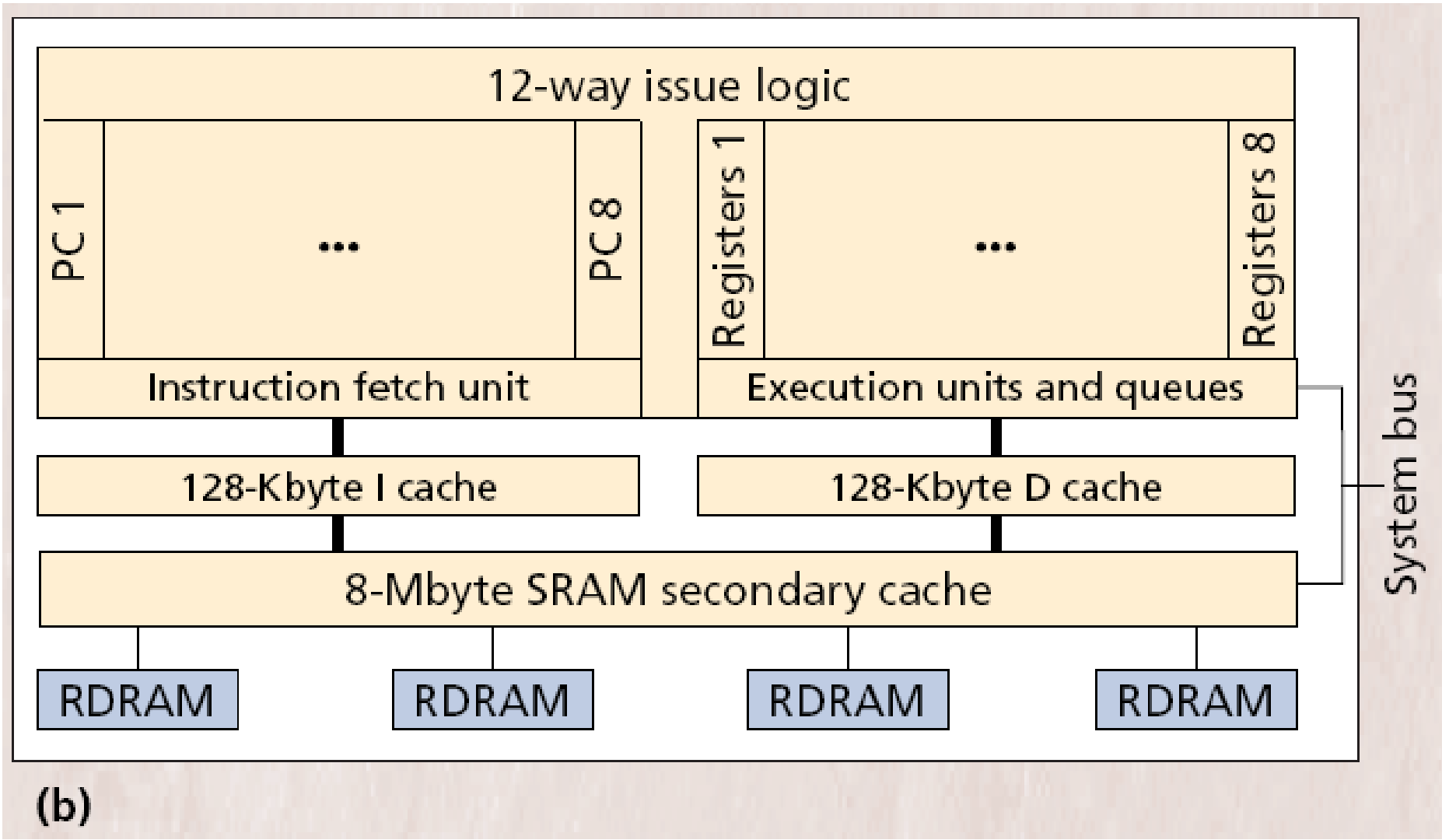
The best way to use
one billion transistors
?

Superscalar processor	Instruction-level parallelism
Simultaneous multithreading	Instruction-level and thread-level parallelism
Chip multiprocessor	Thread-level and process-level parallelism

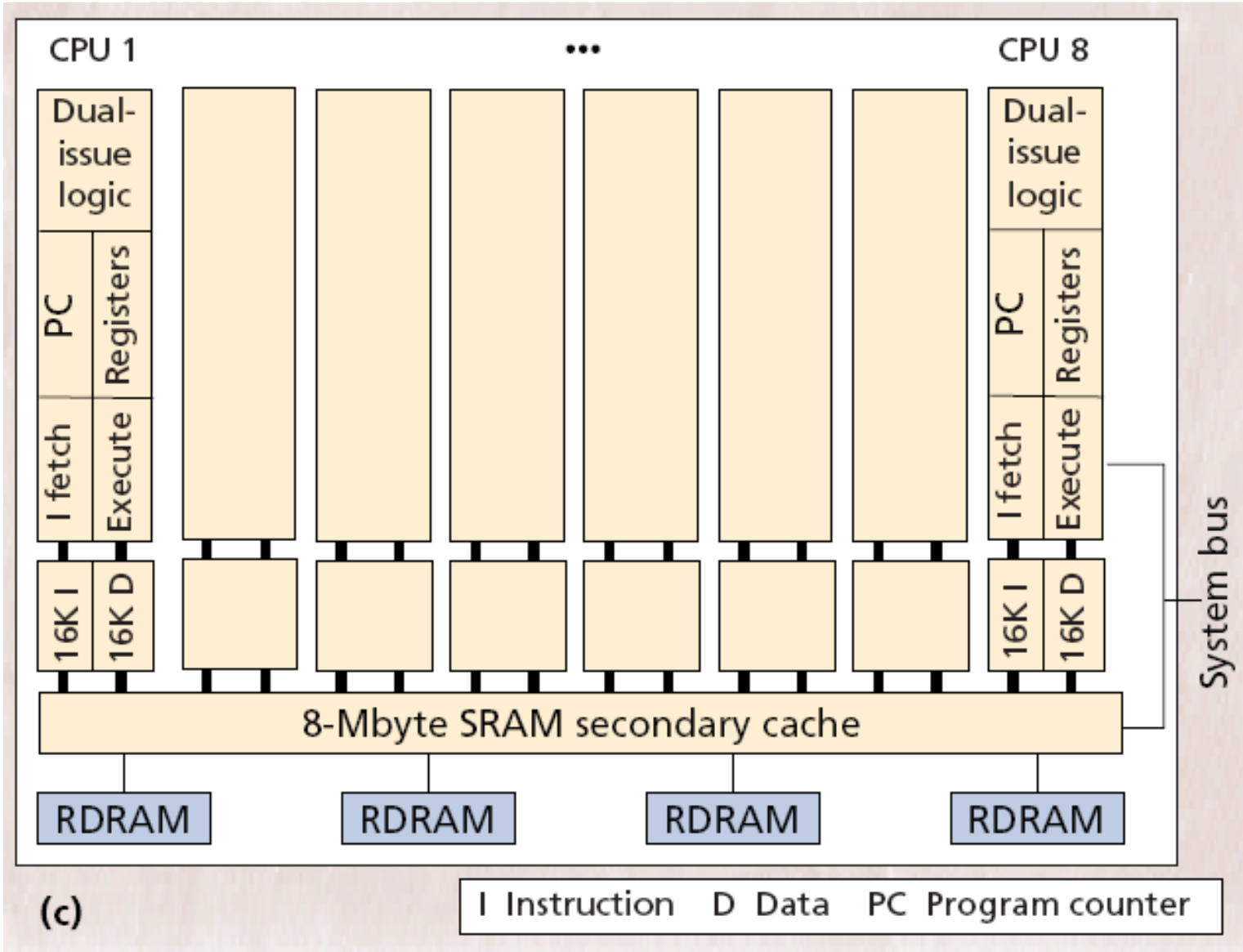
Superscalar processor



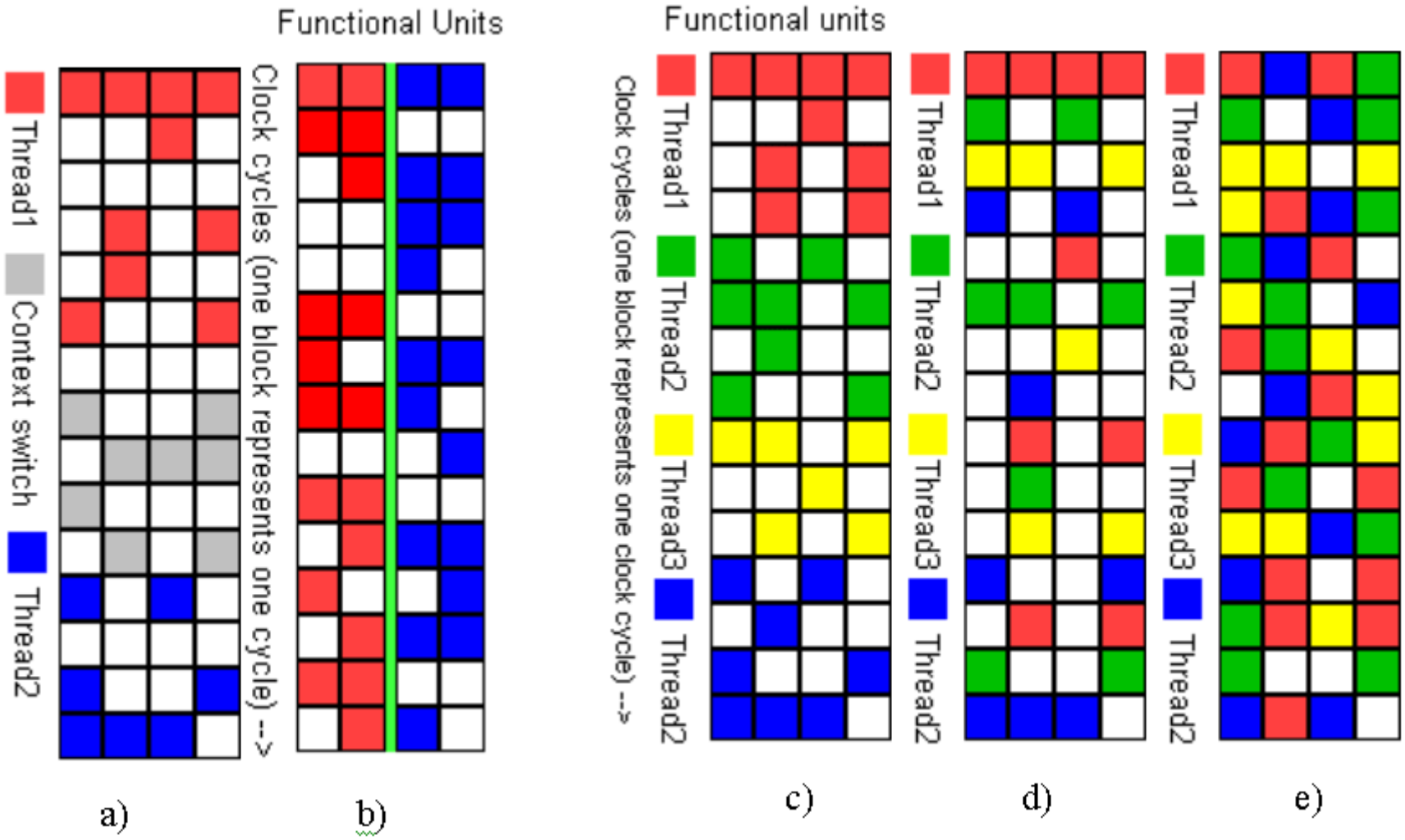
Simultaneous Multithreading



Chip Multiprocessor



A general comparison between various Multithreading Techniques



a) is a traditional superscalar

b) a 2-way CMP (On-chip multiprocessor)

c) a 4-way CMT (Coarse-grained Multithreading)

d) a 4-way FMT (Fine-grained Multithreading)

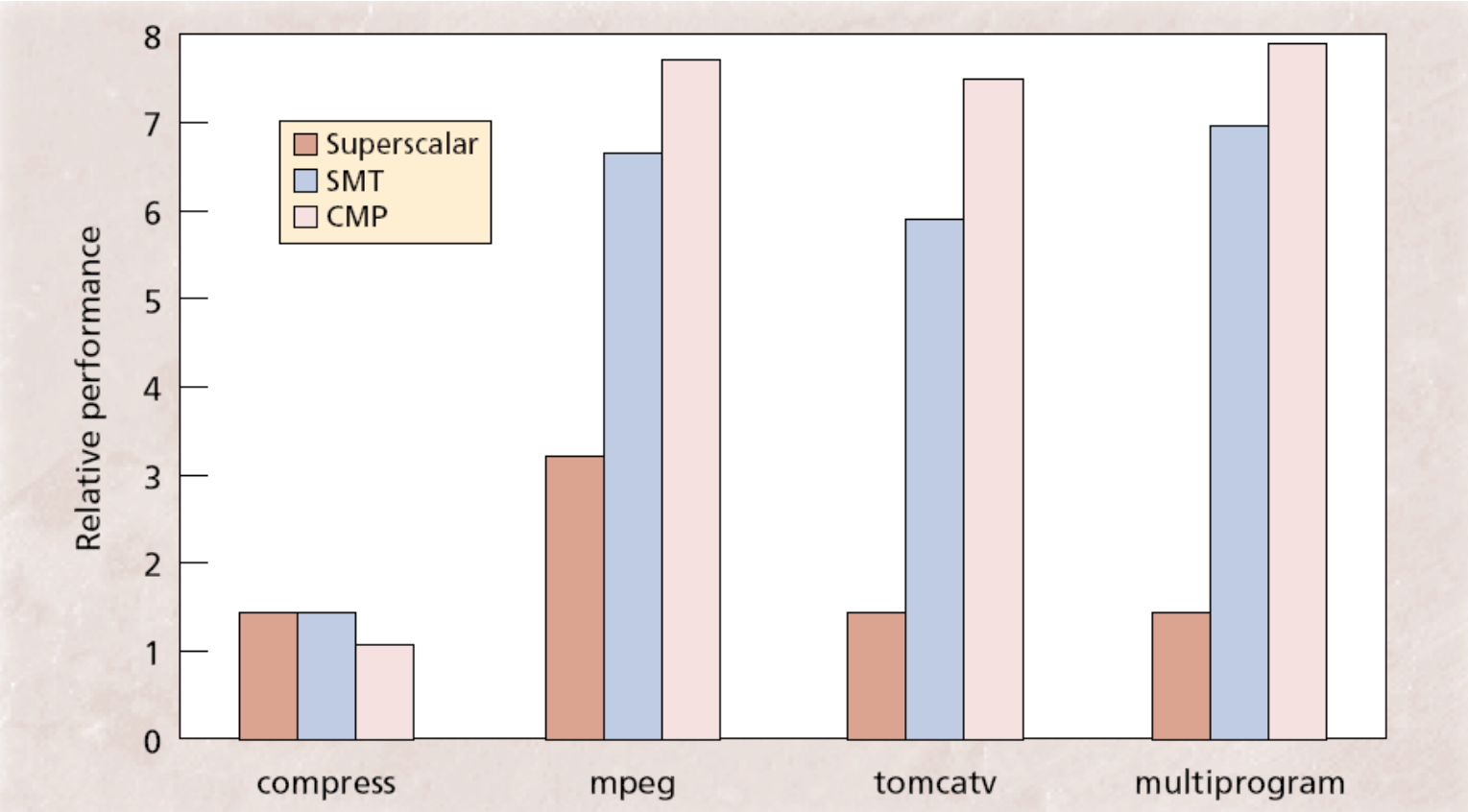
e) a 4-way SMT (Simultaneous Multithreading)

Characteristic Parameters

Table 1. Characteristics of superscalar, simultaneous multithreading, and chip multiprocessor architectures.

Characteristic	Superscalar	Simultaneous multithreading	Chip multiprocessor
Number of CPUs	1	1	8
CPU issue width	12	12	2 per CPU
Number of threads	1	8	1 per CPU
Architecture registers (for integer and floating point)	32	32 per thread	32 per CPU
Physical registers (for integer and floating point)	32 + 256	256 + 256	32 + 32 per CPU
Instruction window size	256	256	32 per CPU
Branch predictor table size (entries)	32,768	32,768	8 × 4,096
Return stack size	64 entries	64 entries	8 × 8 entries
Instruction (I) and data (D) cache organization	1 × 8 banks	1 × 8 banks	1 bank
I and D cache sizes	128 Kbytes	128 Kbytes	16 Kbytes per CPU
I and D cache associativities	4-way	4-way	4-way
I and D cache line sizes (bytes)	32	32	32
I and D cache access times (cycles)	2	2	1
Secondary cache organization (Mbytes)	1 × 8 banks	1 × 8 banks	1 × 8 banks
Secondary cache size (bytes)	8	8	8
Secondary cache associativity	4-way	4-way	4-way
Secondary cache line size (bytes)	32	32	32
Secondary cache access time (cycles)	5	5	7
Secondary cache occupancy per access (cycles)	1	1	1
Memory organization (no. of banks)	4	4	4
Memory access time (cycles)	50	50	50
Memory occupancy per access (cycles)	13	13	13

Relative Performance



compres: (from SPEC95) Little ILP, no TLP
mpeg-2: (multimedia) Much ILP and TLP. Computationally intensive
tomcatv: (from SPEC95) Scientific fl-point appl. Much loop-level parallelism, high memory bandwidth
multiprogram: (several different simulations). Process-level parallelism