

Index

Page references in bold represent figures and tables.

Numbers

2:1 cache rule of thumb, C-28
3ASC Purple pSeries 575, **E-20, E-44, E-56**
80x86 processors. *See* Intel 80x86
99.999% (five nines) claims, 399

A

ABC (Atanasoff Berry Computer), K-5
ABI (application binary interface), B-20
absolute addressing mode, **B-9**
absolute value function, G-24
Accelerated Strategic Computing Initiative (ASCI), **E-20, E-44, E-56**
access bits, C-50
access time, 310, F-15 to F-16
access time gap, in disks and DRAM, 359, **359**
accumulator architecture, B-3, **B-4**
acknowledgments, 217, H-39 to H-41
ACS, K-20 to K-21
Ada, integer division and remainder in, **I-12**
adaptive routing, E-47, E-53, **E-54, E-73, E-93 to E-94**
adders
 carry-lookahead, 38, I-37 to I-41, **I-38, I-40, I-41, I-42, I-44**
 carry-propagate, I-48
 carry-save, I-47 to I-48, **I-48**
 carry-select, I-43 to I-44, **I-43, I-44**

 carry-skip, I-41 to I-43, **I-42, I-44**
 faster division with one, I-54 to I-58, **I-55, I-56, I-57**
 faster multiplication with many, I-50 to I-54, **I-50 to I-54**
 faster multiplication with single, I-47 to I-50, **I-48, I-49**
 ripple-carry, I-2 to I-3, **I-3, I-42, I-44**
addition. *See also* adders
 denormalized numbers, I-26 to I-27
 floating-point, I-21 to I-27, **I-24, I-36**
 multiple-precision, I-13
 speeding up, I-25 to I-26, I-37 to I-44
address aliasing prediction, 130
address faults, C-40
address mapping
 in AMD Opteron, C-12 to C-13, C-53 to C-54, **C-54**
 in multibanked caches, 299, **299**
 page coloring, C-37
 in trace caches, 296, **309**
address size, importance of, C-56
address space, shared, 202
address specifiers, B-21, J-68
address translations (memory mapping). *See also* translation lookaside buffers
 in AMD Opteron, C-44 to C-45, **C-45**
 during cache indexing, 291–292, C-36 to C-38, **C-37, C-39**
 caches dedicated to, C-43
 illegal, C-54 to C-55
 in Intel Pentium, C-50
 memory protection and, 317

 in virtual memory, C-40, C-44 to C-47, **C-45, C-47**
addressing modes
 in embedded architectures, J-5 to J-6, **J-6**
 in Intel 80x86, J-47 to J-49, **J-50, J-59 to J-62, J-59 to J-62**
 in MIPS data transfers, B-34 to B-35
 paged vs. segmented, C-3, **C-41, C-42**
 real, J-45, **J-50**
 in RISC desktop architectures, J-5 to J-6, **J-5**
 types of, B-9 to B-10, **B-9, B-11, B-12, B-13**
 in VAX, J-67, J-70 to J-71
 in vector computers, B-31 to B-32
advanced load address table (ALAT), F-27, G-40
advanced loads, G-40
advanced mobile phone service (AMPS), D-25
Advanced Research Project Agency (ARPA), E-97, E-99
Advanced RISC Machines. *See* ARM
Advanced RISC Machines Thumb. *See* ARM Thumb
Advanced Switching Interconnect (ASI), E-103
affine array indexes, G-6
age-based, E-49
aggregate bandwidth, E-13, E-17, E-24
Aiken, Howard, K-3 to K-4
ALAT table, F-27, G-40
algebraic right shift, J-33
alias analysis. *See* memory alias analysis

I-2 Index

- alias prediction. *See* address aliasing prediction
- aliases, C-36
- Alles, A., E-98
- Alliant computers, vector processors in, F-49
- Alpha
 - addressing modes in, J-5 to J-6, **J-5**
 - architecture overview, **J-4**
 - common MIPS extensions in, J-19 to J-24, **J-21 to J-23**
 - conditional branch options in, **B-19**
 - instructions unique to, J-27 to J-29
 - MIPS core subset in, J-6 to J-16, **J-7, J-9 to J-13, J-17**
 - page size changes in, C-56 to C-57
- Alpha 21064, A-43
- Alpha 21164, 220
- Alpha 21264, 88–89, 140
- Alpha MAX, **D-11**, J-16 to J-19, **J-18**
- AlphaServer 4100, 220–221, **221**
- AltaVista benchmark, 221, **221, 222**
- ALU instructions
 - in media extensions, D-10 to D-11
 - memory addressing and, **B-12**
 - in MIPS architecture, **B-37, B-37**
 - operand format in, **B-5, B-6**
 - overview of, A-4 to A-5
 - in unpipelined MIPS
 - implementation, A-27 to A-28, A-30
- AMAT. *See* average memory access time
- AMD 64, J-46
- AMD Athlon 64 processor
 - L1 cache size in, 294
 - performance of, **G-43**
 - SMT performance of, 179–181, **179, 180, 181**
- AMD ElanSC520, D-13, **D-13**
- AMD K6, 294, D-13, **D-13**
- AMD Opteron processor
 - 64-bit memory management, C-53 to C-55, **C-54, C-55**
 - antialiasing in, C-36
 - cache-coherent multiprocessors, 215
 - data cache organization, C-12 to C-14, **C-13, C-15**
 - dies, **22, 23**
 - interconnection networks in, 216–217, **216**
 - L1 cache size in, 294
 - memory hierarchy performance, 331–335, **332, 333, 334**
 - memory hierarchy structure, 326–331, **327, 328, 341**
 - multibanked caches in, 299, **309**
 - multilevel exclusion in, C-34
 - organization of, 326, **327**
 - Pentium 4 compared to, 136–138, **137, 138**
 - performance on SPEC
 - benchmarks, **35, 37, 255–257, 255, 256, 257**
 - translation lookaside buffer organization, C-44 to C-45, **C-45**
- AMD Pacifica, 320, 339
- Amdahl's Law
 - law of diminishing returns and, 40
 - limited available parallelism and, 202–203
 - in multiple-issue processor performance, 184
 - parallel computers and, 258–259
 - pitfalls in using, 48
 - resource allocation and, 40
 - speedup determination, 39–40
- America processor, K-21 to K-22
- AMPS (advanced mobile phone service), D-25
- Anderson, S. F., I-63
- Andreessen, Marc, E-98
- Andrew benchmark, 225
- annulling, G-26
- antialiasing, C-36 to C-37
- antidependences, 70, G-7 to G-8, K-23
- Apple Macintosh, memory addressing in, K-53
- application binary interface (ABI), **B-20**
- applied load, E-53
- arbitration
 - arbitration algorithm, E-49 to E-50, **E-49, E-52**
 - overview of, E-21 to E-22
 - in shared-media networks, E-23
 - in switch microarchitecture, E-57 to E-58, E-60 to E-61, E-62
 - in switched-media networks, E-24
 - techniques for, E-49, **E-49**
- arbitration algorithm, E-49, E-52
- areal density, 358
- arithmetic. *See* computer arithmetic arithmetic mean, 36
- ARM
 - addressing modes in, J-5 to J-6, **J-6**
 - architecture overview, **J-4**
 - common extensions in, J-19 to J-24, **J-23, J-24**
 - conditional branch options in, **B-19, J-17**
 - instructions unique to, J-36 to J-37
 - MIPS core subset in, J-6 to J-16, **J-8, J-9, J-14 to J-17**
 - multiply-accumulate in, J-19, **J-20**
- ARM Thumb
 - addressing modes in, J-5 to J-6, **J-6**
 - architecture overview, **J-4**
 - common extensions in, J-19 to J-24, **J-23, J-24**
 - instructions unique to, J-37 to J-38
 - MIPS core subset in, J-6 to J-16, **J-8, J-9, J-14 to J-17**
 - multiply-accumulate in, J-19, **J-20**
 - reduced code size in, B-23
- ARPA (Advanced Research Project Agency), E-97, E-99
- ARPANET, E-97 to E-98
- arrays, age of access to, 304, **304**
- array indexes, G-6
- array multipliers, I-50 to I-54, **I-50 to I-54**
- array processors, K-36
- ASCI (Accelerated Strategic Computing Initiative), **E-20, E-44, E-56**
- ASCI Red Paragon, **E-20, E-44, E-56**
- ASCI White SP Power3, **E-20, E-44, E-56**

- ASI (Advanced Switching Interconnect), E-103
- associativity
- access times and, 294-295, **294**
 - cache indexes and, 38, 291, C-38
 - cache size and, 292
 - miss rates and, 291, C-28 to C-29, **C-29, C-39**
 - in multilevel caches, C-33 to C-34
 - in virtual memory, C-42
- Astronautics ZS-1, K-22
- asynchronous events, A-40, A-41, **A-42**
- asynchronous I/O, 319
- asynchronous transfer mode (ATM)
- development of, E-98, E-99
 - Ethernet compared with, E-89, **E-90**
 - packet format for, E-75
 - as telecommunications standard, E-79
 - virtual output queues and, E-60
 - as wide area network, E-4
- ATA disks, 360-361, **361**, 365
- Atanasoff, John, K-5
- Atanasoff Berry Computer (ABC), K-5
- Athlon 64 processor. *See* AMD Athlon 64 processor
- Atlas computer, K-52
- ATM. *See* asynchronous transfer mode
- atomic exchange synchronization
- primitives, 238-240
- atomic operations, 214
- "atomic swap," J-20, **J-21**
- attributes field, C-50 to C-51, **C-51**
- autodecrement addressing mode, **B-9**
- autoincrement addressing mode, **B-9**
- availability claims, 399-400, **400**. *See also* reliability
- average memory access time (AMAT)
- associativity and, C-28 to C-29, **C-29**
 - block size and, C-26 to C-28, **C-26, C-27**
 - cache size and, 295
 - formula for, 290, C-15, **C-21**
 - in multilevel caches, C-30 to C-31
 - in out-of-order computers, C-19 to C-21, **C-21**
 - processor performance and, C-17 to C-19
 - in split vs. unified caches, C-15 to C-17
 - for two-level caches, C-30
- average memory stall times, 298, C-56, **C-57**
- average reception factor, E-26, E-32
- average residual service time, 384
- B**
- back substitution, G-10
- backpressure, E-65
- Baer, J.-L., K-54
- bandwidth
- aggregate, E-13, E-17, E-24
 - bisection, E-39, E-41, E-42, E-55, E-89
 - communication, H-3
 - defined, 15, C-2, E-13
 - distributed memory and, 230
 - DRAM improvements, 313
 - in floating-point computations, 1-62
 - full bisection, E-39, E-41
 - high-memory, 337-338, **339**
 - improvements in, 15, **16**
 - injection, E-18, E-26, E-41, E-55, E-63
 - integrated instruction fetch units and, 126-127
 - I/O, 371
 - link injection, E-17
 - link reception, E-17
 - main memory, 310
 - multibanked caches and, 298-299, **299, 309**
 - in multiple processors, 205
 - network performance and, E-16 to E-19, **E-19**, E-25 to E-29, **E-28**, E-89, **E-90**
 - network switching and, E-50, E-52
 - network topology and, E-41
 - nonblocking caches and, 296-298, **297, 309**
 - overestimating, 336, **338**
 - packet discarding and, E-65
 - pipelined cache access and, 296, **309**, A-7
 - reception, E-18, E-26, E-41, E-55, E-63, E-89
 - return address predictors and, 125, **126**
 - vector performance and, F-45
- bandwidth gap, in disks and DRAM, 359
- bandwidth matching, E-112
- Banerjee, U., F-51, K-23
- Banerjee tests, K-23
- bank busy time, **F-15, F-16**
- Barnes-Hut algorithm
- characteristics of, H-8 to H-9, **H-11**
 - in distributed-memory multiprocessors, **H-28 to H-32**
 - in symmetric shared-memory multiprocessors, H-21 to H-26, **H-23 to H-26**
- barrier networks, H-42
- barrier synchronization, H-13 to H-16, **H-14, H-15, H-16**
- base fields, C-50
- base registers, A-4
- base station architectures, D-22
- based plus scaled index mode, **J-58**
- basic blocks, 67
- Baskett, F., F-48
- Bell, G., C-56, K-14, K-37 to K-39, K-52
- benchmark suites, 30
- benchmarks, 29-33. *See also* SPEC
- Alta Vista, 221, **221, 222**
 - Andrew, 225
 - changes over time, 50
 - compiler flags for, 29
 - of dependability, 377-379, **378**
 - desktop, 30-32
 - EEMBC, 30, D-12 to D-13, **D-12, D-13, D-14**
 - for embedded systems, D-12 to D-13, **D-12, D-13, D-14**
 - evolution of SPEC, 30-32, **31**
 - historical development of, K-6 to K-7
 - Linpack, F-8 to F-9, F-37 to F-38
 - NAS parallel, F-51
 - Perfect Club, F-51
 - reports of, 33
 - reproducibility of, 33
 - for servers, 32-33
 - SFS, 376
 - source code modifications and, 29-30
 - suites, 30

- benchmarks (*continued*)
 - summarizing results from, 33–37, **35**
 - synthetic, 29
 - Transaction Processing Council, 32, 374–375, **375**
 - Web server, 32–33
- Benes topology, E-33, **E-33**
- Beowulf project, K-42
- BER (bit error rate), D-21 to D-22
- Berkeley RISC processor, K-12 to K-13
- Berkeley's Tertiary Disk project, 368, **369**, 399, **399**
- Berners-Lee, Tim, E-98
- between vs. within instructions, A-41, **A-42**
- biased exponents, I-15 to I-16
- biased system, for signed numbers, I-7
- bidirectional multistage, E-33
- Big Endian byte order, B-7, B-34
- Bigelow, Julian, K-3
- BINAC, K-5
- binary tree multipliers, I-53 to I-54
- binary-coded decimal formats, B-14
- binary-to-decimal conversion, I-34
- Birman, M. A., I-58
- bisection bandwidth, E-39, E-41, E-42, E-55, E-89
- bisection traffic fraction, E-41 to E-42, E-55
- bit error rate (BER), D-21 to D-22
- bit selection, C-7
- bit vectors, 232
- bits
 - access, C-50
 - dirty, C-10, C-44
 - NaT (Not a Thing), G-38 G-40
 - poison, G-28, G-30 to G-32
 - present, C-50
 - sticky, I-18
 - use, C-43 to C-44
 - valid, C-8
- block addressing, 299, **299**, C-8 to C-9, **C-8**
- block multithreading, K-26
- block offsets, C-8, **C-8**
- block servers, 390–391
- block size
 - miss rates and, 252, **252**, 291, C-25 to C-28, **C-26**, **C-27**, **C-39**
 - multilevel inclusion and, 248–249
 - multiprogrammed cache misses and, 227–230, **228**, **229**
 - in shared-memory multiprocessors, H-27 to H-29, **H-29**, **H-31**
 - SMP cache misses and, 223–224
- block transfer engines (BLT), E-87, **E-87**
- blocking floating point, D-6
- blocking
 - in centralized switched networks, E-32
 - network topology and, E-41
 - to reduce miss rate, 303–305, **304**
 - in switching, E-51
- blocking factors, 303
- blocks
 - defined, C-2
 - destination, H-8
 - in directory-based protocols, 234–237, **235**
 - dirty vs. clean, C-10
 - exclusive, 210–211, **214**, **215**
 - invalid, 211–212, **213**, **214**, **215**
 - modified, 211, **213**, 231
 - owners of, 211, 231, **235**
 - placement in main memory, C-42
 - replacement, C-9, **C-10**, C-14, C-40, C-43 to C-44
 - set-associative placement of, 38, 289
 - shared, 211, **213**, **214**, **215**, 231
 - state transition diagrams, 234–236, **235**, **236**
 - uncached, 231
 - unmodified, **214**
 - victim, 301
 - write invalidate protocols and, 211
- Blue Gene/L. *See* IBM Blue Gene/L
- BOMB, K-4
- Booth recoding, I-8 to I-9, **I-9**, I-48 to I-49, **I-49**
- Bouknight, W. Jack, 195, K-36
- bounds checking, C-50, **C-51**
- branch costs, 80–89, **81**, **84**, **85**, **87**, **88**
- branch delay, A-35 to A-37, A-60, **A-60**, A-65. *See also* delayed branch schemes
- branch delay slot, A-23 to A-25, **A-23**, **A-24**
- branch displacements, J-39
- branch folding, 125
- branch hazards, A-21 to A-26. *See also* control hazards
 - instruction fetch cycle and, A-21, **A-21**
 - in MIPS pipeline, A-35 to A-37, **A-38**
 - in MIPS R4000 pipeline, A-64, **A-65**
 - performance of branch schemes, A-25 to A-26, **A-26**
 - reducing penalties, **A-21**, A-22 to A-25, **A-22**, **A-23**, **A-24**
 - restarting, A-42 to A-43
- branch history tables, 82–86, **83**, **84**, **85**
- branch prediction. *See* hardware branch prediction; static branch prediction
- branch registers, J-32 to J-33
- branches
 - branch target distances, B-18, **B-18**
 - conditional branch operations, B-19, **B-19**, **B-20**
 - in control flow instructions, B-16, **B-17**, B-18, B-37 to B-38, **B-38**
 - history tables, 82–86, **83**, **84**, **85**
 - in IBM 360, J-86 to J-87, **J-86**, **J-87**
 - penalties, A-36 to A-37, **A-39**
 - registers, J-32 to J-33
 - in RISC architecture, A-5
 - straightening, 302
 - vectored, J-35
- branch-prediction buffers, 82–86, **83**, **84**, **85**
- branch-target buffers (BTB), 122–125, **122**, **124**
- branch-target calculations, A-35, **A-38**, **A-39**
- breakpoints, **A-40**, **A-42**
- Brent, R. P., I-63
- bridges, E-78
- bristling, E-38, E-92
- broadcasting, E-24, H-35 to H-36
- Brooks, F. P., Jr., C-49
- BTB (branch-target buffers), 122–125, **122**, **124**
- bubble flow control, E-53, E-73

- bubbles, pipeline, A-13, A-20, E-47.
 See also pipeline stalls
 - buckets, in histograms, 382
 - buffered crossbars, E-62
 - buffered wormhole switching, E-51
 - buffers. *See also* translation lookaside buffers
 - branch-prediction, 82–86, **83**, **84**, **85**
 - branch-target, 122–125, **122**, **124**
 - buffered crossbars, E-62
 - central, E-57
 - development of, K-22
 - in disk storage, 360, **360**
 - instruction, A-15
 - limited, H-38 to H-40
 - load, 94, **94**, 95, 97, **101**, 102–103
 - reorder, 106–114, **107**, **110**, **111**, **113**, G-31 to G-32
 - streaming, K-54
 - victim, 301, 330, C-14
 - write, 210, 289, 291, 300–301, **301**, **309**
 - bundles, G-34 to G-36, **G-36**, **G-37**
 - Burks, A. W., 287, I-62, K-3
 - buses
 - in barrier synchronization, H-15 to H-16, **H-16**
 - bottlenecks in, 216, **216**
 - data misses and, H-26, **H-26**
 - development of, K-62 to K-63
 - fairness in, H-13
 - point-to-point links replacing, 390, **390**
 - in scoreboarding, A-70, A-73
 - in shared-media networks, E-22, **E-22**, **E-40**
 - single-bus systems, 217–218
 - in snooping coherence protocols, 211–212, **213**, **214**, **215**
 - in Tomasulo's approach, 93, 95, 96, 98, **101**
 - in write invalidates, 209–210, 212, **213**
 - bypassing. *See* forwarding
 - byte addressing, 9, **299**
 - byte order, B-7 to B-8, **B-8**
 - C**
 - C description language extensions, **B-9**, B-36 to B-37
 - C language, integer division and remainder in, **I-12**
 - caches
 - 2:1 cache rule of thumb, C-28
 - AMD Opteron data cache
 - example, C-12 to C-14, **C-13**, **C-15**
 - block addressing, 299, **299**, C-8 to C-9, **C-8**
 - block placement in, C-7 to C-8, **C-7**
 - block size, H-25 to H-26, **H-25**
 - data, C-9, C-13, **C-15**, F-46
 - defined, C-2
 - development of, K-53
 - in IBM Blue Gene/L, H-42
 - interprocessor communication and, H-5
 - L1 (*See* L1 cache)
 - L2 (*See* L2 cache)
 - multibanked, 298–299, **299**, **309**
 - multilevel, 291
 - nonblocking, 296–298, **297**, **309**, **K-54**
 - remote communication latency and, 205
 - in RISC pipelines, A-7
 - separated, C-14
 - SMT challenges to, 176–177
 - states of, 212
 - in superscalar processors, 155
 - tags in, 210–211, 289, C-36
 - trace, 131, **132**, **133**, 296, **309**
 - victim, 301, K-54
 - virtual, C-36 to C-38, **C-37**
 - virtual memory compared with, C-40, **C-41**
 - writes in, C-9 to C-12, **C-10**, C-13
 - cache access, pipelined, 296
 - cache associativity. *See* associativity
 - cache banks, 298–299, **299**
 - cache blocks. *See* blocks
 - cache coherence problem. *See also* cache coherence protocols and, 207–208
 - I/O, 325–326
 - overview of, 205–207, **206**
 - snooping protocols, 208–209, **209**
 - snooping protocol example, 211–215, **213**, **214**, **215**
 - state diagrams and, **214**, **215**
 - write invalidate protocol implementation, 209–211
- cache coherence protocols. *See also* cache coherence problem; directory-based cache coherence protocols
 - avoiding deadlock from limiting buffering, H-38 to H-40
- directory controller implementation, H-40 to H-41
- in distributed shared-memory multiprocessors, H-36 to H-37
- in distributed-memory multiprocessors, 232–233, **232**, **233**
- in large-scale multiprocessors, H-34 to H-41
- memory consistency in, 243–246
- snooping protocols and, 208–218, H-34, H-35
- spin lock scheme and, 241–242, **242**
- synchronization and, 240–242, **242**
- uniform memory access and, 217
- cache CPI equation, 168
- cache hierarchy. *See also* memory hierarchy
 - in AlphaServer 4100, 220
 - cache organization overview, 288–293, **292**
 - multilevel inclusion, 248–249
- cache hits, C-2
- cache indexing
 - address translation during, 291–292, C-36 to C-38, **C-37**, **C-39**
 - in AMD Opteron, 326, 329, C-12
 - equation for, 326, 329, **C-21**
 - index size and, C-38, C-46
- cache misses
 - block replacement in, **C-10**, C-14
 - categories of, C-22 to C-24, **C-23**, **C-24**
 - communication of, H-35 to H-36
 - defined, C-2
 - in in-order execution, C-2 to C-3
 - in invalidate protocols, 210–211

I-6 Index

- cache misses (*continued*)
 - nonblocking caches and, 296–298, **297**, **309**
 - processor performance and, C-17 to C-19
 - in SMP commercial workloads, 222–223, **222**, **223**
- cache optimizations, C-22 to C-38
 - average memory access time formula, 290, C-15, **C-21**
 - avoiding address translation during cache indexing, C-36 to C-38, **C-37**, **C-39**
 - categories of, C-22
 - higher associativity and, C-28 to C-29, **C-29**, **C-39**
 - larger block sizes and, C-25 to C-28, **C-26**, **C-27**, **C-39**
 - larger cache sizes and, **C-23**, **C-24**, C-28, **C-39**
 - miss rate components and, C-22 to C-25, **C-23**, **C-24**
 - multilevel caches and, C-29 to C-34, **C-32**, **C-39**
 - read priorities over writes, C-34 to C-35, **C-39**
- cache performance, C-15 to C-21
 - average memory access time and, 290, 295, C-15 to C-17
 - cache size and, 293–295, **294**, **309**
 - compiler optimizations and, 302–305, **304**, **309**
 - compiler-controlled prefetching and, 305–309, **309**
 - critical word first and early restart, 299–300, **309**
 - hardware prefetching and, 305, **306**, **309**
 - high memory bandwidth and, 337–338, **339**
 - merging write buffers and, 300–301, **301**, **309**
 - miss penalties and out-of-order processors, C-19 to C-21, **C-21**
 - multibanked caches and, 298–299, **299**, **309**
 - nonblocking caches and, 296–298, **297**, **309**
 - optimization summary, 309, **309**
 - overemphasizing DRAM bandwidth, 336, **338**
 - overview of, C-3 to C-6
 - pipelined cache access and, 296, **309**
 - predicting from other programs, 335, **335**
 - sufficient simulations for, 336, **337**
 - trace caches and, 296, **309**
 - way prediction and, 295, **309**
- cache prefetching, 126–127, 305–306, **306**, **309**
- cache replacement miss, **214**
- cache size
 - 2:1 cache rule of thumb, C-28
 - hit time and, 293–295, **294**, **309**
 - miss rates and, 252, **252**, 291, C-25 to C-28, **C-26**, **C-27**, **C-39**
 - multiprogrammed workload misses and, 227–230, **228**, **229**
 - performance and, H-22, H-24, **H-24**, H-27, **H-28**
 - SMP workload misses and, 223–224, **223**, **224**, **226**
- cache-only memory architecture (COMA), K-41
- CACTI, 294, **294**
- call gates, C-52
- callee saving, B-19 to B-20, B-41 to B-43
- caller saving, B-19 to B-20
- canceling branches, A-24 to A-25
- canonical form, C-53
- capabilities, in protection, C-48 to C-49, K-52
- capacitive load, 18
- capacity misses
 - defined, 290, C-22
 - relative frequency of, C-22, **C-23**, **C-24**
 - in shared-memory multiprocessors, H-22 to H-26, **H-23 to H-26**
- carrier sensing, E-23
- carrier signals, D-21
- carry-lookahead adders (CLA), 38, I-37 to I-41, **I-38**, **I-40 to I-42**, **I-44**, I-63
- carry-propagate adders (CPAs), I-48
- carry-save adders (CSAs), I-47 to I-48, **I-48**, **I-50**, I-55
- carry-select adders, I-43 to I-44, **I-43**, **I-44**
- carry-skip adders, I-41 to I-43, **I-42**, **I-44**
- CAS (column access strobe), 311–313, **313**
- case statements, register indirect jumps for, B-18
- CCD (charged-couple device), D-19
- CDB (common data bus), 93, 95, 96, 98, **101**
- CDC 6600 processor
 - data trunks in, A-70
 - dynamic scheduling in, 95, A-67 to A-68, **A-68**, K-19
 - multithreading in, K-26
 - pipelining in, K-10
- CDC STAR-100, F-44, F-47
- CDC vector computers, F-4, **F-34**
- CDMA (code division multiple access), D-25
- Cell Broadband Engines (Cell BE), E-70 to E-72, **E-71**
- cell phones, D-20 to D-25, **D-21**, **D-23**, **D-24**
- cells, in octrees, H-9
- centralized shared-memory architectures, 199–200, **200**. *See also* symmetric shared-memory multiprocessors
- centralized switched networks, E-30 to E-34, **E-31**, **E-33**, E-48
- centrally buffered, E-57
- CFM (current frame pointer), G-33 to G-34
- Chai, L., E-77
- chaining, F-35, **F-35**
- channel adapters, E-7
- channels, D-24
- character operands, B-13
- character strings, B-14
- charged-couple devices (CCD), D-19
- checksum, E-8, E-12
- chimes, F-10 to F-12, F-20, F-40
- choke packets, E-65
- Cholesky factorization method, H-8
- CIFS (Common Internet File System), 391
- circuit switching, E-50, E-64
- circular queues, E-56

- CISC (complex instruction set computer), J-65
- CLA (carry-lookahead adders), 38, I-37 to I-41, **I-38, I-40 to I-42, I-44, I-63**
- clock cycles (clock rate)
- associativity and, C-28 to C-29, **C-29**
 - CPI and, 140–141
 - memory stall cycles and, C-4 to C-5, C-20
 - processor speed and, 138–139, **139**
 - SMT challenges and, 176, **179, 181, 183**
- clock cycles per instruction (CPI)
- in AMD Opteron, 331–335, **332, 333, 334**
 - cache, 168
 - cache misses and, C-18
 - computation of, 41–44, 203–204
 - ideal pipeline, 66–67, **67**
 - in Pentium 4, 134, 136, **136**
 - pipelining and, A-3, **A-7, A-8, A-12**
 - processor speed and, 138–139
 - in symmetric shared-memory multiprocessors, 221, **222**
- clock rate. *See* clock cycles
- clock skew, A-10
- Clos topology, E-33, **E-33**
- clusters
- commodity vs. custom, 198
 - development of, K-41 to K-44
 - in IBM Blue Gene/L, H-41 to H-44, **H-43, H-44**
 - Internet Archive, 392–397, **394**
 - in large-scale multiprocessors, H-44 to H-46, **H-45**
- Cm* multiprocessor, K-36
- C.mmp project, K-36
- CMOS chips, 18–19, 294, **294, F-46**
- coarse-grained multithreading, 173–174, **174, K-26. *See also* multithreading**
- Cocke, John, K-12, K-20, K-21 to K-22
- code division multiple access (CDMA), D-25
- code rearrangement, miss rate reduction from, 302
- code scheduling. *See also* dynamic scheduling
- for control dependences, 73–74
 - global, 116, G-15 to G-23, **G-16, G-20, G-22**
 - local, 116
 - loop unrolling and, 79–80, 117–118
 - static scheduling, A-66
- code size, 80, 117, D-3, D-9
- CodePack, B-23
- coefficient of variance, 383
- coerced exceptions, A-40 to A-41, **A-42**
- coherence, 206–208. *See also* cache coherence problem; cache coherence protocols
- coherence misses
- defined, 218, C-22
 - in multiprogramming example, **229**
 - in symmetric shared-memory multiprocessors, H-21 to H-26, **H-23 to H-26**
 - true vs. false sharing, 218–219
- cold-start misses, C-22. *See also* compulsory misses
- collision detection, E-23
- collision misses, C-22
- collocation sites, E-85
- COLOSSUS, K-4
- column access strobe (CAS), 311–313, **313**
- column major order, 303
- COMA (cache-only memory architecture), K-41
- combining trees, H-18
- commercial workloads
- Decision Support System, 220
 - multiprogramming and OS performance, 225–230, **227, 228, 229**
 - online transaction processing, 220
 - SMP performance in, 220–224, **221 to 226**
- committed instructions, A-45
- commodities, computers as, 21
- commodity clusters, 198, H-45 to H-46, **H-45**
- common case, focusing on, 38
- common data bus (CDB), 93, 95, 96, 98, **101**
- Common Internet File System (CIFS), 391
- communication
- bandwidth, H-3
 - cache misses and, H-35 to H-36
 - global system for mobile communication, D-25
 - interprocessor, H-3 to H-6
 - latency, H-3 to H-4
 - message-passing vs. shared-memory, H-4 to H-6
 - multiprocessing models, 201–202
 - NEWS, E-41 to E-42
 - peer-to-peer, E-81 to E-82
 - remote access, 203–204
 - user-level, E-8
- compare, select, and store units (CSSU), D-8
- compare and branch instruction, B-19, **B-19**
- compare instructions, B-37
- compiler optimization, 302–305
- branch straightening, 302
 - compared with other techniques, **309**
 - compiler structure and, B-24 to B-26, **B-25**
 - examples of, B-27, **B-28**
 - graph coloring, B-26 to B-27
 - impact on performance, B-27, **B-29**
 - instruction set guidelines for, B-29 to B-30
 - loop interchange, 302–303
 - phase-ordering problem in, B-26
 - reducing code size and, B-43, **B-44**
 - technique classification, B-26, **B-28**
 - in vectorization, F-32 to F-34, **F-33, F-34**
- compilers
- compiler-controlled prefetching, 305–309, **309**
 - development of, K-23 to K-24
 - eliminating dependent computations, G-10 to G-12
 - finding dependences, G-6 to G-10
 - global code scheduling, 116, G-15 to G-23, **G-16, G-20, G-22**
 - Java, K-10
- compilers

- compilers (*continued*)
 multimedia instruction support,
 B-31 to B-32
 performance of, B-27, **B-29**
 recent structures of, B-24 to B-26,
B-25
 register allocation in, B-26 to
 B-27
 scheduling, A-66
 software pipelining in, G-12 to
 G-15, **G-13, G-15**
 speculation, G-28 to G-32
- complex instruction set computer
 (CISC), J-65
- component failures, 367
- compulsory misses
 defined, 290, C-22
 in multiprogramming example,
 228, **229**
 relative frequency of, C-22, **C-23,**
C-24
 in SMT commercial workloads,
 222, **224, 225**
- computation-to-communication ratios,
 H-10 to H-12, **H-11**
- computer architecture
 defined, 8, 12, J-84, K-10
 designing, 12–13, **13**
 flawless design fallacy, J-81
 functional requirements in, **13**
 historical perspectives on, J-83 to
 J-84, K-10 to K-11
 instruction set architecture, 8–12,
9, 11, 12
 organization and hardware,
 12–15, **13**
 quantitative design principles,
 37–44
 signed numbers in, I-7 to I-10
 trends in, 14–16, **15, 16**
- computer arithmetic, I-1 to I-65
 carry-lookahead adders, I-37 to
 I-41, **I-38, I-40, I-41, I-42,**
I-44
 carry-propagate adders, I-48
 carry-save adders, I-47 to I-48,
I-48
 carry-select adders, I-43 to I-44,
I-43, I-44
 carry-skip adders, I-41 to I-43,
I-42, I-44
- chip design and, I-58 to I-61, **I-58,**
I-59, I-60
- denormalized numbers, I-15, I-20
 to I-21, I-26 to I-27, I-36
- exceptions, I-34 to I-35
- faster division with one adder,
 I-54 to I-58, **I-55, I-56, I-57**
- faster multiplication with many
 adders, I-50 to I-54, **I-50 to**
I-54
- faster multiplication with single
 adders, I-47 to I-50, **I-48,**
I-49
- floating-point addition, I-21 to
 I-27, **I-24, I-36**
- floating-point arithmetic, I-13 to
 I-16, I-21 to I-27, **I-24**
- floating-point multiplication, I-17
 to I-21, **I-18, I-19, I-20**
- floating-point number
 representation, I-15 to I-16,
I-16
- floating-point remainder, I-31 to
 I-32
- fused multiply-add, I-32 to I-33
- historical perspectives on, I-62 to
 I-65
- instructions in RISC architectures,
 J-22, **J-22, J-23, J-24**
- iterative division, I-27 to I-31,
I-28
- overflow, I-8, I-10 to I-12, **I-11,**
 I-20
- in PA-RISC architecture, J-34 to
 J-35, J-36
- pipelining in, I-15
- precision in, **I-16, I-21, I-33 to**
 I-34
- radix-2 multiplication and
 division, I-4 to I-7, **I-4, I-6,**
 I-55 to I-58, **I-56, I-57**
- ripple-carry adders, I-2 to I-3, **I-3,**
I-42, I-44
- shifting over zeros technique, I-45
 to I-47, **I-46**
- signed numbers, I-7 to I-10, I-23,
I-24, I-26
- special values in, I-14 to I-15
- subtraction, I-22 to I-23, I-45
- systems issues, I-10 to I-13, **I-11,**
I-12
- underflow, I-36 to I-37, I-62
- computers, classes of, 4–8
- condition codes, A-5, A-46, **B-19, J-9**
 to J-16, J-71
- condition registers, **B-19**
- conditional branch operations
 in control flow, B-19, **B-19, B-20**
 in RISC architecture, J-11 to J-12,
J-17, J-34, J-34
- conditional instructions. *See*
 predicated instructions
- conditional moves, G-23 to G-24
- conflict misses
 defined, 290, C-22
 four divisions of, C-24 to C-25
 relative frequency of, C-22, **C-23,**
C-24
- congestion management, E-11, E-12,
 E-54, E-65
- connectedness, E-29
- Connection Multiprocessor 2, K-35
- connectivity, E-62 to E-63
- consistency. *See* cache coherence
 problem; cache coherence
 protocols; memory
 consistency models
- constant extension, in RISC
 architecture, J-6, **J-9**
- constellation, **H-45**
- contention
 in centralized switched networks,
 E-32
 congestion from, E-89
 in network performance, E-25,
 E-53
 network topologies and, E-38
 in routing, E-45, E-47
 in shared-memory
 multiprocessors, H-29
- contention delay, E-25, E-52
- context switch, 316, C-48
- control dependences, 72–74, 104–105,
 G-16
- control flow instructions, B-16 to B-21
 addressing modes for, B-17 to
 B-18, **B-18**
 conditional branch operations,
 B-19, **B-19, B-20**
 in Intel 80x86, J-51
 in MIPS architecture, B-37 to
 B-38, **B-38**

- procedure invocation options,
 - B-19 to B-20
 - types of, B-16 to B-17, **B-17**
 - control hazards, A-11, A-21 to A-26, **A-21 to A-26**, F-3. *See also*
 - branch hazards; pipeline hazards
 - control stalls, 74
 - Convex C-1, **F-7, F-34, F-49**
 - Convex Exemplar, K-41
 - convoys, F-10 to F-12, **F-13, F-18, F-35, F-39**
 - Conway, L., I-63
 - cooling, 19
 - Coonen, J., I-34
 - copy propagation, G-10 to G-11
 - core plus ASIC (system on a chip),
 - D-3, D-19, **D-20**
 - correlating predictors, 83–86, **84, 85, 87, 88**
 - Cosmic Cube, K-40
 - costs, 19–25
 - in benchmarks, 375
 - of branches, 80–89, **81, 84, 85, 87, 88**
 - commodities and, 21
 - disk power and, **361**
 - of integrated circuits, 21–25, **22, 23**
 - in interconnection networks, **E-40**, E-89, E-92
 - of Internet Archive clusters, 394–396
 - learning curve and, 19
 - linear speedups in multiprocessors and, 259–260, **261**
 - prices vs., 25–28
 - of RDRAM, 336, **338**
 - of transaction-processing servers, 49–50, **49**
 - trends in, 19–25
 - of various computing classes, **D-4**
 - volume and, 20–21
 - yield and, 19–20, **20, 22–24**
 - count registers, J-32 to J-33
 - CPAs (carry-propagate adders), I-48
 - CPI. *See* clock cycles per instruction
 - CPU time, 28–29, 41–45, C-17 to C-18, **C-21**
 - Cray, Seymour, F-1, F-48, F-50
 - Cray arithmetic algorithms, I-64
 - Cray C90, **F-7, F-32, F-50**
 - Cray J90, F-50
 - Cray SV1, **F-7**
 - Cray T3D, E-86 to E-87, **E-87**, F-50, K-40
 - Cray T3E, 260, K-40
 - Cray T90, **F-7, F-14, F-50**
 - Cray T932, F-14
 - Cray X1
 - characteristics of, **F-7**
 - memory in, F-46
 - multi-streaming processors in, F-43
 - processor architecture in, F-40 to F-43, **F-41, F-42, F-51**
 - Cray X1E, **E-20, E-44, E-56, F-44, F-51**
 - Cray X-MP
 - characteristics of, **F-7**
 - innovations in, F-48
 - memory pipelines on, F-38
 - multiple processors in, F-49
 - peak performance in, F-44
 - vectorizing compilers in, **F-34**
 - Cray XT3, **E-20, E-44, E-56**
 - Cray Y-MP, **F-7, F-32 to F-33, F-33, F-49 to F-50**
 - Cray-1
 - chaining in, F-23
 - characteristics of, **F-7**
 - development of, K-12
 - innovations in, F-48
 - memory bandwidth in, F-45
 - peak performance on, F-44
 - register file in, F-5
 - Cray-2, **F-34, F-46, F-48**
 - Cray-3, F-50
 - credit-based flow control, E-10, E-65, E-71, E-74
 - critical path, G-16, G-19
 - critical word first strategy, 299–300, **309**
 - crossbars, 216, E-30, **E-31, E-60**
 - cryptanalysis machines, K-4
 - CSAs (carry-save adders), I-47 to I-48, **I-48, I-50, I-55**
 - CSSU (compare, select, and store units), D-8
 - current frame pointers (CFM), G-33 to G-34
 - custom clusters, 198, **H-45**
 - cut-through switching, E-50, E-60, E-74
 - CYBER 180/990, A-55
 - CYBER 205, F-44, F-48
 - cycle time, 310–311, **313**
 - Cydrome Cydra 5, K-22 to K-23
- ## D
- Dally, Bill, E-1
 - DAMQ (dynamically allocatable multi-queues), E-56 to E-57
 - Darley, H. M., I-58
 - DARPA (Defense Advanced Research Projects Agency), F-51
 - data alignment, B-7 to B-8, **B-8**
 - data caches, C-9, C-13, **C-15**, F-46
 - data dependences, 68–70, G-16
 - data flow
 - control dependences and, 73–74
 - double data rate, 314–315, **314**
 - executions, 105
 - hardware-based speculation and, 105
 - as ILP limitation, 170
 - value prediction and, 170
 - data hazards. *See also* RAW hazards; WAR hazards
 - 2-cycle stalls, A-59, **A-59**
 - minimizing stalls by forwarding, A-17 to A-18, **A-18**, A-35, **A-36, A-37**
 - in MIPS pipelines, A-35 to A-37, **A-38, A-39**
 - in pipelining, A-11, A-15 to A-21, **A-16, A-18 to A-21**
 - requiring stalls, A-19 to A-20, **A-20, A-21**
 - in Tomasulo's approach, 96
 - in vector processors, F-2 to F-3, F-10
 - data miss rates
 - on distributed-memory multiprocessors, H-26 to H-32, **H-28 to H-32**
 - hardware-controlled prefetch and, 307–309
 - in multiprogramming and OS workloads, 228, **228, 229**
 - on symmetric shared-memory multiprocessors, H-21 to H-26, **H-23 to H-26**

- data parallelism, K-35
- data paths
 - for eight-stage pipelines, A-57 to A-59, **A-58**, **A-59**
 - in MIPS implementation, **A-29**
 - in MIPS pipelines, A-30 to A-31, **A-31**, A-35, **A-37**
 - in RISC pipelines, A-7, **A-8**, A-9
- data races, 245
- data rearrangement, miss rate
 - reduction from, 302
- data transfer time, 311–313, **313**
- data trunks, A-70
- datagrams, E-8, E-83
- data-level parallelism, 68, 197, 199
- data-race-free programs, 245, K-44
- DDR (double data rate), 314–315, **314**
- dead time, F-31 to F-32, **F-31**
- dead values, 74
- deadlock avoidance, E-45
- deadlock recovery, E-46
- deadlocked protocols, 214
- deadlocks
 - adaptive routing and, E-93
 - bubble flow control and, E-53
 - characteristics of, H-38
 - in dynamic network
 - reconfiguration, E-67
 - from limited buffering, H-38 to H-40
 - in network routing, E-45, E-47, E-48
- DeCegama, Angel, K-37
- decimal operations, J-35
- decision support systems (DSS), 220–221, **221**, **222**
- decoding
 - forward error correction codes, D-6
 - in RISC instruction set
 - implementation, A-5 to A-6
 - in unpipelined MIPS
 - implementation, A-27
- dedicated link networks, E-5, E-6, **E-6**
- Defense Advanced Research Projects Agency (DARPA), F-51
- delayed branch schemes
 - development of, K-24
 - in MIPS R4000 pipeline, A-60, **A-60**
- in pipeline hazard prevention, A-23 to A-25, **A-23**
- in restarting execution, A-43
- in RISC architectures, J-22, **J-22**
- Dell 2650, **322**
- Dell PowerEdge 1600SC, **323**
- Dell PowerEdge 2800, **47**, **48**, **49**
- Dell PowerEdge 2850, **47**, **48**, **49**
- Dell Precision Workstation 380, **45**
- denormalized numbers, I-15, I-20 to I-21, I-26 to I-27, I-36
- density-optimized processors, E-85
- dependability. *See* reliability
- dependence analysis, G-6 to G-10
- dependence distance, G-6
- dependences, 68–75. *See also* pipeline hazards
 - control, 72–74, 104–105, G-16
 - data, 68–70, G-16
 - eliminating dependent
 - computations, G-10 to G-12
 - finding, G-6 to G-10
 - greatest common divisor test, G-7
 - interprocedural analysis, G-10
 - loop unrolling and, G-8 to G-9
 - loop-carried, G-3 to G-5
 - name, 70–71
 - number of registers to analyze, 157
 - recurrences, G-5, G-11 to G-12
 - types of, G-7 to G-8
 - unnecessary, as ILP limitations, 169–170
- depth of pipeline, A-12
- descriptor privilege level (DPL), **C-51**
- descriptor tables, C-50 to C-51
- design faults, 367, **370**
- desktop computers
 - benchmarks for, 30–32
 - characteristics of, **D-4**
 - disk storage on, K-61 to K-62
 - instruction set principles in, B-2
 - memory hierarchy in, **341**
 - multimedia support for, **D-11**
 - operand type and size in, B-13 to B-14
 - performance and
 - price-performance of, 44–46, **45**, **46**
 - rise of, 4
 - system characteristics, 5, **5**
- Dest field, 109
- destination blocks, H-8
- deterministic routing, E-46, E-53, **E-54**, E-93
- devices, E-2
- Dhrystone performance, 30, D-12
- die yield, 22–24
- dies, costs of, 21–25, **22**, **23**
- Digital Alpha. *See* Alpha
- digital cameras, D-19, **D-20**
- Digital Equipment Vax, 2
- Digital Linear Tape, K-59
- digital signal processors (DSP), D-5 to D-11
 - in cell phones, D-23, **D-23**
 - defined, D-3
 - media extensions, D-10 to D-11, **D-11**
 - multiply-accumulate in, J-19, **J-20**
 - overview, D-5 to D-7, **D-6**
 - saturating arithmetic in, D-11
 - TI 320C6x, D-8 to D-10, **D-9**, **D-10**
 - TI TMS320C55, D-6 to D-8, **D-6**, **D-7**
- dimension-order routing, E-46, E-53
- DIMMs (dual inline memory modules), 312, 314, **314**
- direct addressing mode, **B-9**
- direct attached disks, 391
- direct networks, E-34, **E-37**, E-48, E-67, E-92
- Direct RDRAM, 336, **338**
- direct-mapped caches
 - block addresses in, C-8, **C-8**
 - block replacement with cache misses, C-9, **C-10**
 - defined, 289, C-7 to C-8, **C-7**
 - development of, K-53
 - size of, 291, **292**
- directory controllers, H-40 to H-41
- directory-based cache coherence
 - protocols
 - defined, 208, 231
 - development of, K-40 to K-41
 - distributed shared-memory and, 230–237, **232**, **233**, **235**, **236**
 - example, 234–237, **235**, **236**
 - overview of, 231–234, **232**, **233**

- directory-based multiprocessors, H-29, **H-31**
 - dirty bits, C-10, C-44
 - Discrete Cosine Transform, D-5
 - Discrete Fourier Transform, D-5
 - disk arrays, 362–366, **363, 365**, K-61 to K-62. *See also* RAID
 - disk storage
 - areal density in, 358
 - buffers in, 360, **360**
 - development of, K-59 to K-61
 - disk arrays, 362–366, **363, 365**
 - DRAM compared with, 359
 - failure rate of, 50–51
 - intelligent interfaces in, 360, **360, 361**
 - power in, 361
 - RAID, K-61 to K-62 (*See also* RAID)
 - Tandem disks, 368–369, **370**
 - technology growth in, 14, 358–359, **358**
 - Tertiary Disk project, 368, **369, 399, 399**
 - dispatch stage, 95
 - displacement addressing mode
 - in Intel 80x86, J-47
 - overview, **B-9**, B-10 to B-11, **B-11, B-12**
 - display lists, D-17 to D-18
 - distributed routing, E-48
 - distributed shared-memory (DSM)
 - multiprocessors. *See also* multiprocessing
 - cache coherence in, H-36 to H-37
 - defined, 202
 - development of, K-40
 - directory-based coherence and, 230–237, **232, 233, 235, 236**
 - in large-scale multiprocessors, **H-45**
 - latency of memory references in, **H-32**
 - distributed switched networks, E-34 to E-39, **E-36, E-37, E-40**, E-46
 - distributed-memory multiprocessors
 - advantages and disadvantages of, 201
 - architecture of, 200–201, **201**
 - scientific applications on, H-26 to H-32, **H-28 to H-32**
 - division
 - faster, with one adder, I-54 to I-58, **I-55, I-56, I-57**
 - floating-point remainder, I-31 to I-32
 - fused multiply-add, I-32 to I-33
 - iterative, I-27 to I-31, **I-28**
 - radix-2 integer, I-4 to I-7, **I-4, I-6, I-55** to I-56, **I-55**
 - shifting over zeros technique, I-45 to I-47, **I-46**
 - speed of, I-30 to I-31
 - SRT, I-45 to I-47, **I-46, I-55** to I-58, **I-57**
 - do loops, dependences in, 169
 - Dongarra, J. J., F-48
 - double data rate (DDR), 314–315, **314**
 - double extended precision, **I-16**, I-33
 - double precision, **A-64, I-16**, I-33, J-46
 - double words, J-50
 - double-precision floating-point operands, **A-64**
 - downtime, cost of, **6**
 - DPL (descriptor privilege level), **C-51**
 - DRAM (dynamic RAM)
 - costs of, 19–20, **359**
 - DRDRAM, 336, **338**
 - embedded, D-16 to D-17, **D-16**
 - historical performance of, 312, **313**
 - memory performance
 - improvement in, 312–315, **314**
 - optimization of, 310
 - organization of, 311–312, **311**
 - overestimating bandwidth in, 336, **338**
 - redundant memory cells in, 24
 - refresh time in, 312
 - synchronous, 313–314
 - technology growth, 14
 - in vector processors, F-46, F-48
 - DRDRAM (direct RDRAM), 336, **338**
 - driver domains, 321–322, **323**
 - DSM. *See* distributed shared-memory (DSM) multiprocessors
 - DSP. *See* digital signal processors
 - DSS (decision support system), 220–221, **221, 222**
 - dual inline memory modules (DIMMs), 312, 314, **314**
 - Duato's Protocol, E-47
 - dynamic branch frequency, 67
 - dynamic branch prediction, 82–86, D-4. *See also* hardware branch prediction
 - dynamic memory disambiguation. *See* memory alias analysis
 - dynamic network reconfiguration, E-67
 - dynamic power, 18–19
 - dynamic RAM. *See* DRAM
 - dynamic scheduling, 89–104. *See also* Tomasulo's approach
 - advantages of, 89
 - defined, 89
 - development of, K-19, K-22
 - evaluation pitfalls, A-76
 - examples of, 97–99, **99, 100**
 - loop-based example, 102–104
 - multiple issue and speculation in, 118–121, **120, 121**
 - overview, 90–92
 - scoreboarding technique, A-66 to A-75, **A-68, A-71** to **A-75**
 - Tomasulo's algorithm and, 92–97, 100–104, **101, 103**
 - dynamically allocatable multi-queues (DAMQs), E-56 to E-57
 - dynamically shared libraries, B-18
- ## E
- early restart strategy, 299–300
 - Earth Simulator, F-3 to F-4, F-51
 - Ecache, F-41 to F-43
 - Eckert, J. Presper, K-2 to K-3, K-5
 - e-cube routing, E-46
 - EDN Embedded Microprocessor Benchmark Consortium (EEMBC), 30, D-12 to D-13, **D-12, D-13, D-14**
 - EDSAC, K-3
 - EDVAC, K-2 to K-3
 - EEMBC benchmarks, 30, D-12 to D-13, **D-12, D-13, D-14**
 - effective address, A-4, B-9
 - effective bandwidth
 - defined, E-13
 - in Element Interconnect Bus, E-72
 - latency and, E-25 to E-29, **E-27, E-28**

- effective bandwidth (*continued*)
 network performance and, E-16 to E-19, **E-19**, E-25 to E-29, **E-28**, E-89, **E-90**
 network switching and, E-50, E-52
 network topology and, E-41
 packet size and, E-18, **E-19**
- effective errors, 367
- efficiency, EEMBC benchmarks for, D-13, **D-13**, **D-14**
- efficiency factor, E-52, E-55
- EIB (Element Interconnect Bus), E-3, E-70, **E-71**
- eigenvalue method, H-8
- eight-way conflict misses, C-24
- 80x86 processors. *See* Intel 80x86
- ElanSC520, D-13, **D-13**
- elapsed time, 28. *See also* latency
- Element Interconnect Bus (EIB), E-3, E-70, **E-71**
- embedded systems, D-1 to D-26
 benchmarks in, D-12 to D-13, **D-12**, **D-13**, **D-14**
 cell phones, D-20 to D-25, **D-21**, **D-23**, **D-24**
 characteristics of, **D-4**
 costs of, **5**
 data addressing modes in, J-5 to J-6, **J-6**
 defined, **5**
 digital signal processors in, J-19
 instruction set principles in, 4, **B-2**
 media extensions in, D-10 to D-11, **D-11**
 MIPS extensions in, J-19 to J-24, **J-23**, **J-24**
 multiprocessors, D-3, D-14 to D-15
 overview, 7–8
 power consumption and efficiency in, D-13, **D-13**
 real-time constraints in, D-2
 real-time processing in, D-3 to D-5
 reduced code size in RISCs, B-23 to B-24
 in Sanyo VPC-SX500 digital camera, D-19, **D-20**
 in Sony Playstation 2, D-15 to D-18, **D-16**, **D-18**
- in TI 320C6x, D-8 to D-10, **D-9**, **D-10**
 in TI TMS320C55, D-6 to D-8, **D-6**, **D-7**
 vector instructions in, F-47
- Emer, Joel, K-7
- Emotion Engine, SP2, D-15 to D-18, **D-16**, **D-18**
- encoding, B-21 to B-24
 fixed-length, 10, B-22, **B-22**
 hybrid, **B-22**, B-23
 in packet transport, E-9
 reduced code size in RISCs, B-23 to B-24
 variable-length, 10, B-22 to B-23, **B-22**
 in VAX, J-68 to J-70, **J-69**
- end-to-end flow control, E-65, E-94 to E-95
- energy efficiency, 182
- EnergyBench, D-13, **D-13**
- Engineering Research Associates (ERA), K-4
- ENIAC (Electronic Numerical Integrator and Calculator), K-2, K-59
- environmental faults, 367, 369, **370**
- EPIC (Explicitly Parallel Instruction Computer), 114, **115**, 118, G-33, K-24
- ERA (Engineering Research Associates), K-4
- error latency, 366–367
- errors
 bit error rate, D-21 to D-22
 effective, 367
 forward error correction codes, D-6
 latent, 366–367
 meaning of, 366–367
 round-off, D-6, **D-6**
- escape path, E-46
- escape resource set, E-47
- eServer p5 595, **47**, **48**, **49**
- Eshraghian, K., I-65
- ETA-10, **F-34**, F-49
- Ethernet
 as local area network, E-4
 overview of, E-77 to E-79, **E-78**
 packet format in, **E-75**
 performance, E-89, **E-90**
 as shared-media network, E-23
- Ethernet switches, 368, **369**
- even/odd multipliers, I-52, **I-52**
- EVEN-ODD scheme, 366
- EX. *See* execution/effective address cycle
- exceptions
 coerced, A-40 to A-41, **A-42**
 in computer arithmetic, I-34 to I-35
 dynamic scheduling and, 91, 95
 floating-point, A-43
 inexact, I-35
 instruction set complications, A-45 to A-47
 invalid, I-35
 in MIPS pipelining, A-38 to A-41, **A-40**, **A-42**, A-43 to A-45, **A-44**
 order of instruction, A-38 to A-41, **A-40**, **A-42**
 precise exceptions, A-43, A-54 to A-56
 preserving, in compiler speculation, G-27 to G-31
 program order and, 73–74
 restarting execution, A-41 to A-43
 underflow, I-36 to I-37, I-62
- exclusion policy, in AMD Opteron, 329, 330
- exclusive cache blocks, 210–211
- execution time, 28, 257–258, C-3 to C-4. *See also* response time
- execution trace cache, 131, **132**, **133**
- execution/effective address cycle (EX)
 in floating-point MIPS pipelining, A-47 to A-49, **A-48**
 in RISC instruction set, A-6
 in unpipelined MIPS implementation, A-27 to A-28, **A-29**
- expand-down field, C-51
- explicit parallelism, G-34 to G-37, **G-35**, **G-36**, **G-37**
- Explicitly Parallel Instruction Computer (EPIC), 114, **115**, 118, G-33, K-24
- exponential back-off, H-17 to H-18, **H-17**
- exponential distributions, 383–384, 386. *See also* Poisson distribution
- exponents, I-15 to I-16, **I-16**

- extended accumulator architecture, B-3, J-45
 - extended precision, I-33 to I-34
 - extended stack architecture, J-45
- F**
- failure, defined, 366–367
 - failure rates, 26–28, 41, 50–51
 - failures in time (FIT), 26–27
 - fairness, E-23, E-49, H-13
 - false sharing misses
 - in SMT commercial workloads, 222, **224**, **225**
 - in symmetric shared-memory multiprocessors, 218–219, 224
 - fast page mode, 313
 - fat trees, **E-33**, E-34, E-36, E-38, **E-40**, E-48
 - fault detection, 51–52
 - fault tolerance, IEEE on, 366
 - faulting prefetches, 306
 - faults. *See also* exceptions
 - address, C-40
 - categories of, 367, **370**
 - design, 367, **370**
 - environmental, 367, 369, **370**
 - hardware, 367, **370**
 - intermittent, 367
 - meaning of, 366–367
 - page, C-3, C-40
 - permanent, 367
 - transient, 367, 378–379
 - fault-tolerant routing, E-66 to E-68, **E-69**, E-74, E-94
 - FCC (Federal Communications Commission), 371
 - feature size, 17
 - Federal Communications Commission (FCC), 371
 - Feng, Tse-Yun, E-1
 - fetch-and-increment synchronization primitive, 239–240, H-20 to H-21, **H-21**
 - FFT kernels
 - characteristics of, H-7, **H-11**
 - on distributed-memory multiprocessors, H-27 to H-29, **H-28** to **H-32**
 - on symmetric shared-memory multiprocessors, H-21 to H-26, **H-23** to **H-26**
 - FIFO (first in, first out), 382, C-9, **C-10**
 - file server benchmarks, 32
 - filers, 391, 397–398
 - fine-grained multithreading, 173–175, **174**. *See also* multithreading
 - finite-state controllers, 211
 - first in, first out (FIFO), 382, C-9, **C-10**
 - first-reference misses, C-22
 - Fisher, J., 153, K-21
 - FIT (failures in time), 26–27
 - five nines (99.999%) claim, 399
 - fixed point computations, I-13
 - fixed-field decoding, A-6
 - fixed-length encoding, 10, B-22, **B-22**
 - fixed-point arithmetic, D-5 to D-6
 - flash memory, 359–360
 - flexible chaining, F-24 to F-25
 - flit, E-51, E-58, E-61
 - Floating Point Systems AP-120B, K-21
 - floating-point arithmetic. *See also* floating-point operations
 - addition in, I-21 to I-27, **I-24**, I-36 in Alpha, J-29
 - chip design and, I-58 to I-61, **I-58**, **I-59**, **I-60**
 - conversions to integer arithmetic, I-62
 - denormalized numbers, I-15, I-20 to I-21, I-26 to I-27, I-36
 - development of, K-4 to K-5
 - exceptions in, A-43, I-34 to I-35
 - fused multiply-add, I-32 to I-33
 - historical perspectives on, I-62 to I-65
 - in IBM 360, J-85 to J-86, **J-85**, **J-86**, **J-87**
 - IEEE standard for, I-13 to I-14, **I-16**
 - instructions in RISC architectures, **J-23**
 - in Intel 80x86, J-52 to J-55, **J-54**, **J-61**
 - iterative division, I-27 to I-31, **I-28**
 - in MIPS 64, J-27
 - multiplication, I-17 to I-21, **I-18**, **I-19**, **I-20**
 - pipelining in, I-15
 - precision in, I-21, I-33 to I-34
 - remainder, I-31 to I-32
 - representation of floating-point numbers, I-15 to I-16, **I-16**
 - in SPARC architecture, J-31 to J-32
 - special values in, I-14 to I-15
 - subtraction, I-22 to I-23
 - underflow, I-36 to I-37, I-62
 - floating-point operations. *See also* floating-point arithmetic
 - blocked floating point, D-6
 - conditional branch options, **B-20**
 - instruction operators in, **B-15**
 - latencies of, 75, **75**
 - maintaining precise exceptions, A-54 to A-56
 - in media extensions, D-10
 - memory addressing in, **B-12**, **B-13**
 - in MIPS architecture, B-38 to B-39, **B-40**
 - MIPS pipelining in, A-47 to A-56, **A-48** to **A-51**, **A-57**, **A-58**
 - MIPS R4000 pipeline example, A-60 to A-65, **A-61** to **A-65**
 - multicore processor comparisons, 255
 - nonblocking caches and, 297–298
 - operand types and sizes, B-13 to B-14, **B-15**
 - paired single operations and, D-10 to D-11
 - parallelism and, 161–162, **162**, **166**, 167
 - performance growth since mid-1980s, **3**
 - scoreboarding, A-66 to A-75, **A-68**, **A-71** to **A-75**
 - in Tomasulo's approach, 94, **94**, **107**
 - in vector processors, F-4, F-6, **F-8**, F-11
 - floating-point registers (FPRs), B-34, B-36
 - floating-point status register, B-34
 - floppy disks, K-60
 - flow control
 - bubble, E-53, E-73
 - in buffer overflow prevention, E-22
 - in congestion management, E-65

- flow control (*continued*)
 credit-based, E-10, E-65, E-71, E-74
 defined, E-10
 in distributed switched networks, E-38
 end-to-end, E-65
 link-level, E-58, E-62, E-65, E-72, E-74
 in lossless networks, E-11
 network performance and, E-17
 Stop & Go, E-10
 switching and, E-51
 Xon/Xoff, E-10
- flow-balanced state, **379**
- flush pipeline scheme, A-22, **A-25**
- FM (frequency modulations), D-21
- form factor, E-9
- FORTRAN
 integer division and remainder in, **I-12**
 vector processors in, F-17, F-21, F-33, **F-34**, F-44 to F-45, **F-45**
- forward error correction codes, D-6
- forward path, in cell phone base stations, D-24
- forwarding
 chaining, F-23 to F-25, **F-24**
 in longer latency pipelines, A-49 to A-54, **A-50**, **A-51**
 minimizing data hazard stalls by, A-17 to A-18, **A-18**
 in MIPS pipelines, A-35, **A-36**, **A-37**, A-59, **A-59**
- forwarding logic, 89
- forwarding tables, E-48, E-57, E-60, E-67, E-74
- Fourier-Motzkin test, K-23
- four-way conflict misses, C-24
- FP. *See* floating-point arithmetic; floating-point operations
- FPRs (floating-point registers), B-34, B-36
- fragment field, E-84
- Frank, S. J., K-39
- freeze pipeline scheme, A-22
- Freiman, C. V., I-63
- frequency modulations (FM), D-21
- Fujitsu VP100/VP200, **F-7**, F-49, F-50
- Fujitsu VPP5000, **F-7**
- full access, E-29, E-45, E-47
- full adders, I-2 to I-3, **I-3**
- full bisection bandwidth, E-39, E-41
- full-duplex mode, E-22
- fully associative caches, 289, C-7, **C-7**, C-25
- fully connected, E-34, **E-40**
- function pointers, register indirect jumps for, B-18
- fused multiply-add, I-32 to I-33
- future file, A-55
- G**
- galaxy evolution, H-8 to H-9
- gallium arsenide, F-46, F-50
- gateways, E-79
- gather operations, F-27
- gather/scatter addressing, B-31
- GCD (greatest common divisor) test, G-7
- general-purpose register (GPR)
 computers, B-3 to B-6, **B-4**, **B-6**
- general-purpose registers (GPRs), B-34, G-38
- GENI (Global Environment for Network Innovation), E-98
- geometric mean, 34–37
- geometric standard deviation, 36–37
- Gibson instruction mix, K-6
- Gilder, George, 357
- global address space, C-50
- global code motion, G-16 to G-19, **G-17**
- global code scheduling, G-15 to G-23
 control and data dependences in, G-16
- global code motion, G-16 to G-19, **G-17**
 overview of, G-16, **G-16**
 predication with, G-24
 superblocks, G-21 to G-23, **G-22**
 trace scheduling, G-19 to G-21, **G-20**
 in VLIW, 116
- global collective networks, H-42, **H-43**
- global common subexpression elimination, B-26, **B-28**
- global data area, in compilers, B-27
- Global Environment for Network Innovation (GENI), E-98
- global miss rate, C-30 to C-33, **C-32**
- global optimizations, B-26, **B-28**
- global scheduling, 116
- global system for mobile communication (GSM), D-25
- global/stack analysis, 164–165, **164**
- Goldberg, D., I-34
- Goldberg, I. B., I-64
- Goldberg, Robert, 315
- Goldschmidt's algorithm, I-29, **I-30**, I-61
- Goldstine, H. H., 287, I-62, K-2 to K-3
- Google, E-85
- GPR (general-purpose registers), B-34, G-38
- GPR computers, B-3 to B-6, **B-4**, **B-6**
- gradual underflow, I-15, I-36
- grain size, defined, 199
- graph coloring, B-26 to B-27
- greatest common divisor (GCD) test, G-7
- grid, E-36
- GSM (global system for mobile communication), D-25
- guest domains, 321–322, **323**
- guests, in virtual machines, 319–320, 321
- H**
- hackers, J-65
- half adders, I-2 to I-3
- half-duplex mode, E-22
- half-words, B-13, B-34
- handshaking, E-10
- hard real-time systems, D-3 to D-4
- hardware, defined, 12
- hardware branch prediction, 80–89
 branch-prediction buffers and, 82–86, **83**, **84**, **85**
 branch-target buffers, 122–125, **122**, **124**
 correlating predictors, 83–86, **84**, **85**, **87**, **88**
 development of, K-20
 effects of branch prediction schemes, 160–162, **160**, **162**
 in ideal processor, 155, 160–162, **160**, **162**
 integrated instruction fetch units and, 126–127
 in Pentium 4, 132–134, **134**
 speculating through multiple branches, 130

- tournament predictors, 86–89, **160**, 161, **162**, K-20
 - trace caches and, 296
 - hardware description notation, **J-25**
 - hardware faults, 367, **370**
 - hardware prefetching. *See* prefetching
 - hardware-based speculation, 104–114.
 - See also* speculation
 - data flow value prediction, 170
 - limitations of, 170–171, 182–183
 - reorder buffer in, 106–114, **107**, **110**, **111**, **113**
 - Tomasulo's approach and, 105–109
 - Harvard architecture, K-4
 - hazards. *See* pipeline hazards; RAW hazards; WAR hazards; WAW hazards
 - head-of-line (HOL) blocking
 - adaptive routing and, E-54
 - buffer organizations and, E-58, **E-59**
 - buffered crossbar switches and, E-62
 - congestion management and, E-64
 - virtual channels and, E-93
 - heaps, in compilers, B-27 to B-28
 - heat dissipation, 19
 - helical scan, K-59
 - Hennessy, J., K-12 to K-13
 - HEP processor, K-26
 - Hewlett-Packard PA-RISC. *See* PA-RISC
 - High Productivity Computing Systems (HPCS). F-51
 - higher-radix division, I-55 to I-58, **I-55**, **I-56**, **I-57**
 - higher-radix multiplication, I-48 to I-49, **I-49**
 - high-level language computer architecture (HLLCA), B-26, **B-28**, B-39 to B-43, B-45, K-11
 - high-level optimizations, B-26, **B-28**
 - high-order functions, register indirect jumps for, B-18
 - Hill, M. D., 247, K-54
 - Hillis, Danny, K-38
 - histograms, 382–383
 - historical perspectives, K-1 to K-67
 - on clusters, K-41 to K-44
 - on computer arithmetic, I-62 to I-65
 - cryptanalysis machines, K-4
 - on DRAM, 312, **313**
 - early development of computers, K-2 to K-7
 - on floating point, K-4 to K-5
 - on IBM 360/370, J-83 to J-84
 - on instruction set architecture, B-3, B-45 to B-46, K-9 to K-15
 - on Intel 80x86 development, J-45 to J-46, **J-48**, J-64 to J-65
 - on Intel IA-64 and Itanium 2, G-44
 - on interconnection networks, E-97 to E-104
 - on magnetic storage, K-59 to K-61
 - on memory hierarchy and protection, K-52 to K-54
 - on multiprocessors and parallel processing, K-34 to K-45
 - on pipelining and ILP, K-18 to K-27
 - on quantitative performance measures, K-6 to K-7
 - on vector processors, F-47 to F-51
 - history file, A-55
 - hit time
 - associativity and, C-28 to C-29
 - average memory access time and, C-15
 - during cache indexing, 291–292, C-36 to C-38, **C-37**
 - cache size and, 293–295, **294**
 - defined, 290
 - trace caches and, 296, **309**
 - hit under miss optimization, 296–298, **297**
 - Hitachi S810/S820, **F-7**, **F-34**, F-49
 - Hitachi SuperH. *See* SuperH
 - HLLCA (high-level language computer architecture), B-26, **B-28**, B-39 to B-43, B-45, K-11
 - Hoagland, Al, 357
 - HOL. *See* head-of-line (HOL) blocking
 - home nodes, 232, **233**
 - Honeywell Bull, K-61
 - hop count, E-30, E-38, E-40, E-50
 - Hopkins, M., G-1
 - hosts, in virtual machines, 318
 - hot-swapping, E-67
 - HP Alpha Server, **339**
 - HP Integrity rx2620-2, **45**, **46**
 - HP Integrity rx5670 Cluster, 47, **47**, **48**, **49**
 - HP Integrity Superdome, 47, **47**, **48**, **49**
 - HP PA-RISC MAX2, **D-11**
 - HP Precision Architecture, I-12
 - HP ProLiant BL25p, **45**, **46**
 - HP ProLiant ML350, **45**, **46**, **47**, **48**, **49**
 - HPCD (High Productivity Computing Systems), F-51
 - HPSm, K-22
 - hubs, E-79
 - Hwu, Wen-Mei, K-24
 - hybrid encoding, **B-22**, B-23
 - hypercubes, E-36, **E-37**, **E-40**, E-92, K-41
 - HyperTransport, E-63
- I**
- IA-32 microprocessors, A-46 to A-47
 - IA-64. *See* Intel IA-64
 - IAS computer, K-3
 - IBM 3ASC Purple pSeries 575, **E-20**, **E-44**, **E-56**
 - IBM 360, J-83 to J-89
 - architecture in, J-2, **J-42**, K-10
 - development of, J-83 to J-84
 - dynamic scheduling in, 92
 - frequency of instruction usage in programs, **J-89**
 - hardware-based speculation in, 171
 - instruction sets in, J-85 to J-88, **J-85**, **J-86**, **J-87**, **J-88**
 - memory addressing in, B-8, K-9, K-53
 - memory caches in, K-53
 - memory protection in, K-52
 - virtual memory in, K-53
 - IBM 360/91
 - dynamic scheduling in, 92
 - hardware-based speculation in, 171
 - innovations in, K-20
 - memory caches in, K-53

I-16 Index

- IBM 370
 - development of, J-84
 - floating point formats in, I-63 to I-64
 - guest page tables in, 320
 - memory addressing in, B-8
 - vector length control in, F-18
 - virtualization in, 319
- IBM 701, K-3, K-5
- IBM 801, K-12 to K-13
- IBM 3033 cache, C-38
- IBM 3090/VF, **F-34**
- IBM 3480 cartridge, K-59
- IBM 7030, K-18
- IBM ASCI White SP Power3, **E-20**, **E-44**, **E-56**
- IBM Blue Gene/L
 - architecture of, **E-20**, H-41 to H-42
 - computing node in, H-42 to H-44, **H-43**, **H-44**
 - custom clusters in, 198
 - development of, K-43
 - interconnection networks in, E-72 to E-74
 - routing, arbitration, and switching characteristics in, **E-56**
 - topological characteristics of, **E-44**
- IBM CodePack, B-23
- IBM eServer p5 595, **47**, **48**, **49**
- IBM eServer pSeries 690, **47**, **48**, **49**
- IBM mainframes, virtualization in, 324–325
- IBM Power2 processor, 130
- IBM Power4 processor, **52**, K-27
- IBM Power5 processor
 - clock rate on, 139, **139**
 - eServer p5 575, **178**
 - instruction-level parallelism in, 156
 - memory hierarchy of, **341**
 - multicore, 198
 - on-chip networks, **E-73**
 - performance on SPEC benchmarks, 255–257, **255**, **256**, **257**, **G-43**
 - simultaneous multithreading in, K-27
 - SMT performance on, 176–181, **178 to 181**
 - tournament predictors in, 88
 - virtual registers in, 162
- IBM PowerPC. *See* PowerPC
- IBM RP3, K-40
- IBM RS 6000, K-13
- IBM SAGE, K-63
- IBM zSeries, F-49
- IBM/Motorola PowerPC. *See* PowerPC
- IC (instruction count), 42–44, C-4
- ID. *See* instruction decode/register fetch cycle
- ideal pipeline CPI, 66, **67**
- IDE/ATA disks, 368, **369**
- identifier fields, E-84
- Idle Control Register (ICR), D-8
- IEEE Computer Society Technical Committee on Fault Tolerance, 366
- IEEE standard for floating-point arithmetic
 - advantages of, I-13 to I-14
 - exceptions in, I-34 to I-35
 - format parameters, **I-16**
 - remainders, I-31 to I-32
 - rounding modes, **I-20**
 - standardization process, I-64
 - underflow in, I-36
- if conversion, G-24
- Illiac IV, K-35
- ILP. *See* instruction-level parallelism
- immediate addressing mode, **B-9**, **B-10** to **B-13**, **B-11**, **B-12**, **B-13**
- IMPACT, K-24
- implementation, in design, 8
- implicit parallelism, G-34
- in flight instructions, 156
- inclusion property, 211
- index field, C-8 to C-9, **C-8**
- index vectors, F-27, F-28
- indexed addressing mode, **B-9**, J-67
- indexing. *See* cache indexing
- indirect networks, E-31, E-48, E-67
- inexact exceptions, I-35
- InfiniBand, E-4, E-64, E-74 to E-77, **E-75**, **E-76**, E-102
- infinite population model, 386
- initiation intervals, A-48 to A-49, **A-49**, **A-62**
- initiation rate, F-10
- injection bandwidth, E-18, E-26, E-41, E-55, E-63
- Inktomi search engine, K-42
- input-buffered switch, E-57, E-59, **E-59**, E-62, E-73
- input-output-buffered switch, E-57, **E-57**, E-60, **E-61**, E-62
- instruction buffers, A-15
- instruction cache misses, 227, **227**, 329, **C-15**
- instruction commit, 105–106, 108–110, **110**
- instruction count (IC), 42–44, C-4
- instruction decode, 90
- instruction decode/register fetch cycle (ID)
 - branch hazards and, A-21
 - in RISC instruction set, A-5 to A-6
 - in unpipelined MIPS implementation, A-26 to A-27, **A-29**
- instruction delivery, 121–127, **122**, **124**, **126**
- instruction fetch cycle (IF)
 - branch hazards and, A-21
 - in RISC instruction set, A-5, **A-15**
 - in unpipelined MIPS implementation, A-26 to A-27, **A-29**
- instruction fetch units, 126–127
- instruction formats, **J-7**, **J-8**
- instruction groups, G-34 to G-35, **G-35**
- instruction issue, A-33, A-67 to A-70, **A-68**, **A-71**
- instruction length, J-60, **J-60**. *See also* instruction count
- instruction packets, in embedded systems, D-9, **D-10**
- instruction path length, 42–44, C-4
- instruction set architecture (ISA), 8–12, B-1 to B-47. *See also* RISC architectures
 - addressing modes and, B-9 to B-10, **B-9**, **B-11** to **B-13**
 - classification of, B-3 to B-7
 - compiler role in, B-24 to B-32
 - compiler technology and, B-27 to B-29

- conditional branch operations,
 - B-19, **B-19**, **B-20**
- in Cray X1, F-41
- defined, 8–9
- encoding, B-21 to B-24, **B-22**
- flaws in, B-44 to B-45
- hardware, 9–10, **9**
- high-level language structure and,
 - B-39 to B-43, B-45
- historical perspectives on, B-3,
 - B-45 to B-46, K-9 to K-15
- in IBM 360/370 mainframe, J-83 to J-89
- instructions for control flow, B-16 to B-21
- integer arithmetic issues in, I-10 to I-13, **I-11**, **I-12**, **I-13**
- in Intel 80x86, J-45 to J-65
- in Intel IA-64, G-32 to G-40,
 - G-35**, **G-36**, **G-37**, **G-39**
- measurements, B-2
- memory addressing in, B-7 to
 - B-13, **B-8**, **B-9**, **B-11**, **B-12**, **B-13**
- memory protection and, 324–325
- multimedia instructions, B-31 to B-32
- operand type and size, B-13 to B-14, **B-15**
- operations in the instruction set,
 - B-14 to B-16, **B-15**, **B-16**
- organization of, 12
- orthogonal, B-30, J-83, K-11
- procedure invocation options,
 - B-19 to B-20
- reducing code size and, B-43,
 - B-44**
- variations in, B-43, **B-44**
- in VAX, J-65 to J-83
- vectorized, F-3, F-4 to F-6
- in virtual machines, 317, 319–320
- instruction set complications, A-45 to A-47
- instruction-level parallelism (ILP),
 - 66–141. *See also*
 - instruction-level parallelism
 - limitations; pipelining
 - branch prediction buffers and,
 - 82–86, **83**, **84**, **85**
 - compiler techniques for exposing,
 - 74–80, **75**
 - data dependences and, 68–70
 - data hazards and, 71–72
 - defined, 66
 - development of, K-24 to K-25
 - in embedded systems, D-8
 - hardware *vs.* software approach to, 66
 - instruction fetch bandwidth
 - increases in, 121–127, **122**, **124**, **126**
 - limitations in ideal processors,
 - 154–165, **159**, **160**, **162**, **163**, **164**
 - limitations in realizable
 - processors, 165–170, **166**, 181–184
 - loop unrolling and, 75–80, **75**, 117–118
 - loop-level parallelism, 67–68
 - multiple issue and speculation
 - examples, 118–121, **120**, **121**
 - overview of, 67–68
 - pipeline scheduling and, 75–79
 - processor comparisons for,
 - 179–181, **179**, **180**, **181**
 - in RISC-based machines, 2
 - in simultaneous multithreading,
 - 173, 175
 - SPEC92 benchmarks and, 156, **157**
 - switch from, to TLP and DLP, 4
 - thread-level parallelism *vs.*, 172
 - value prediction in, 130
- instruction-level parallelism
 - limitations, 154–165
 - data flow limit, 170
 - finite registers and, 162–164, **163**
 - hardware model, 154–156
 - imperfect alias analysis and,
 - 164–165, **164**
 - realistic branch and jump
 - prediction effects, 160–162
 - for realizable processors,
 - 165–169, **166**
 - unnecessary dependences,
 - 169–170
 - WAR and WAW hazards through
 - memory, 169
 - on window size and maximum
 - issue count, 156–160, 166–167, **166**
- instructions per clock (IPC), 42, 253
- integer arithmetic
 - carry-lookahead adders, I-37 to
 - I-41, **I-38**, **I-40**, **I-41**, **I-42**, **I-44**
 - carry-select adders, I-43 to I-44,
 - I-43**, **I-44**
 - carry-skip adders, I-41 to I-43,
 - I-42**, **I-44**
 - conversions to floating-point, I-62
 - faster division with one adder,
 - I-54 to I-58, **I-55**, **I-56**, **I-57**
 - faster multiplication with many
 - adders, I-50 to I-54 **I-50** to **I-54**
 - faster multiplication with single
 - adders, I-47 to I-50, **I-48**, **I-49**
 - in Intel 80x86, J-50 to J-52, **J-52**, **J-53**
 - radix-2 multiplication and
 - division, I-4 to I-7, **I-4**, **I-6**, I-55 to I-56, **I-55**
 - ripple-carry addition, I-2 to I-3,
 - I-3**, **I-42**, **I-44**
 - shifting over zeros technique, I-45 to I-47, **I-46**
 - signed numbers, I-7 to I-10
 - speeding up addition, I-37 to I-44
 - speeding up multiplication and
 - division, I-44 to I-58
 - systems issues, I-10 to I-13, **I-11**, **I-12**
- integer registers, B-34
- integrated circuits
 - costs of, 21–25, **22**, **23**
 - dependability of, 25–28
 - feature size in, 17
 - prices of, 25–29
 - technology growth in, 14
 - trends in power in, 17–19
 - wire delay in, 17
- integrated instruction fetch units,
 - 126–127
- Intel 80x86, J-45 to J-65
 - comparative operation
 - measurements, J-62 to J-64, **J-63**, **J-64**
 - conditional branch options in,
 - B-19**
 - development of, J-45 to J-46, J-64 to J-65
 - exceptions in, **A-40**

- Intel 80x86, J-45 to J-65 (*continued*)
floating point operations, J-52 to J-55, **J-54, J-61**
guest OS in, 324
instruction encoding, J-55 to J-56, **J-56, J-57, J-58**
instruction set architecture in, 9–10, B-44 to B-45, **J-42**
integer operations in, J-50 to J-52, **J-52, J-53**
memory addressing in, B-8, C-56
memory protection in, C-49, K-52
operand addressing
 measurements, J-59 to J-62, **J-59 to J-62**
Pacifica revision to, 320, 339
registers and data addressing
 modes, J-47 to J-49, **J-48, J-49, J-50**
RISC instruction sets in, B-3, J-45 to J-65
top ten instructions in, **B-16**
variable instruction encoding in, B-22 to B-23
virtualization and, 320, 321, 339, **340**
- Intel 387, I-33
- Intel ASCI Red Paragon, **E-20, E-44, E-56**
- Intel IA-32 microprocessors, A-46 to A-47, C-49 to C-51, **C-51**
- Intel IA-64
EPIC approach in, 118, G-33, K-24
hardware-based speculation in, 171
historical perspectives on, G-44, K-14 to K-15
ILP performance limitations and, 184
implicit and explicit parallelism in, G-34 to G-37, **G-35, G-36, G-37**
instruction formats in, G-38, **G-39**
page tables in, C-43
register model in, G-33 to G-34
speculation support in, G-38, G-40
- Intel iPSC 860, K-40
- Intel Itanium 1, G-40 to G-41, K-14
- Intel Itanium 2. *See also* Intel IA-64
- functional units and instruction
 issue in, G-41 to G-43, **G-41**
historical perspectives on, G-44
Itanium 1 compared to, G-40 to G-41
peak performance in, **52**
performance measurements of, 179–181, **179, 180, 181, G-43, G-43**
SPEC ratios for, **35, 37**
Sun T1 compared with, 253
- Intel MMX, B-31 to B-32, J-46
- Intel Paragon, K-40
- Intel Pentium
precise exceptions in, A-56
protection in, C-48, C-49 to C-52, **C-51, C-55**
register renaming in, 128
- Intel Pentium 4, 131–138
AMD Opteron compared to, 136–138, **137, 138, 334–335, 334**
clock rate on, 139–141, **139**
hardware prefetching in, 305, **306**
memory hierarchy of, **341**
microarchitecture of, 131–132, **132, 133**
multilevel inclusion in, C-34
performance analysis of, 133–138, **134 to 138, G-43**
prices of, **20**
signal propagation in, 17
SRT division algorithm in, I-56, **I-57**
tournament predictors in, 88
trace caches in, 296
way prediction in, 295
- Intel Pentium 4 Extreme
power efficiency in, 18, 183
SMT performance in, 177, 179–181, **179, 180, 181**
- Intel Pentium 4 Xeon, 215
- Intel Pentium D, 198, 255–257, **255, 256, 257**
- Intel Pentium III, 183
- Intel Pentium M, **20**
- Intel Pentium MMX, **D-11**
- Intel Thunder Tiger4, **E-20, E-44, E-56**
- Intel VT-x, 339–340
- intelligent devices, K-62
- interarrival times, 386
- interconnection networks. *See also* clusters; networks
arbitration in, E-49 to E-50, **E-49, E-56** (*See also* arbitration)
asynchronous transfer mode (*See* asynchronous transfer mode)
buffer organizations, E-58 to E-60
centralized switched, E-30 to E-34, **E-31, E-33, E-48**
characteristics of, **E-20, E-44, E-56**
composing and processing
 messages, E-6 to E-9, **E-7**
compute-optimized, E-88
conceptual illustration of, **E-3**
congestion management, E-64 to E-66
connectivity in, E-62 to E-63
density-optimized vs.
 SPEC-optimized processors, E-85
distributed switched, E-34 to E-39, **E-36, E-37, E-40, E-46**
in distributed-memory multiprocessors, **232**
domains, E-3 to E-5, **E-3, E-4**
Element Interconnect Bus, E-3, E-70, **E-71**
Ethernet (*See* Ethernet)
fault tolerance in, E-66 to E-68, **E-69, E-94**
historical perspectives on, E-97 to E-104
IBM Blue Gene/L eServer, **E-20, E-44, E-56, E-72 to E-74**
InfiniBand, E-4, E-64, E-74 to E-77, **E-75, E-76**
internetworking, E-2, E-80 to E-84, **E-80 to E-84**
I/O subsystem in, E-90 to E-91
latency and effective bandwidth in, E-12 to E-20, **E-13, E-19, E-25 to E-29, E-27, E-28**
memory hierarchy interface
 efficiency, E-87 to E-88
performance of, E-40 to E-44, **E-44, E-52 to E-55, E-54, E-88 to E-92**
protection and user access, E-86 to E-87, **E-87**

- routing in, E-45 to E-48, **E-46**, E-52 to E-55, **E-54**, **E-56**
 - shared-media, E-21 to E-24, **E-22**, E-24 to E-25, E-78
 - smart switches vs. smart interface cards, E-85 to E-86, **E-86**
 - speed of, E-88 to E-89
 - standardization in, E-63 to E-64
 - structure of, E-9 to E-12
 - switch microarchitecture, E-55 to E-58, **E-56**, **E-57**, E-62
 - switched-media, E-21, E-24, E-25
 - switching technique in, E-50 to E-52, **E-56**
 - in symmetric shared-memory multiprocessors, 216–217, **216**
 - topologies in, E-40 to E-44, **E-44**
 - zero-copy protocols, E-8, E-91
 - interference, **D-21**, D-22
 - intermittent faults, 367
 - internal fragmentation, C-46
 - International Mobile Telephony 2000 (IMT-2000), D-25
 - Internet, E-81
 - Internet Archive, 392–397, **394**
 - Internet Protocol, E-81
 - internetworking, E-2, E-80 to E-84, **E-80 to E-84**. *See also* interconnection networks
 - interprocedural analysis, G-10
 - Interrupt Enable (IE) flag, 324
 - interrupts. *See* exceptions
 - invalid exceptions, I-35
 - invalidate protocols. *See* write invalidate protocols
 - inverted page tables, C-43
 - I/O, 371–379. *See also* buses; storage systems
 - asynchronous, 391
 - cache coherence problem, 325–326
 - dependability benchmarks, 377–379, **378**
 - design considerations, 392–393
 - device requests, **A-40**, **A-42**
 - disk accesses at operating systems, 400–401, **401**
 - evaluation of, 394–396
 - historical developments in, K-62 to K-63
 - Internet Archive cluster, 392–397, **394**
 - in multiprogramming workload, 225–227, **227**
 - NetApp FAS6000 filer, 397–398
 - queuing theory, 379–382, **379**, **381**
 - throughput vs. response time, 372–374, **373**, **374**
 - transaction-processing benchmarks, 374–375, **375**
 - virtual caches and, C-37
 - virtual machines and, 320–321, 339
 - write merging and, 301
 - I/O per second (IOPS), 395–396
 - IPC (instructions per clock), 42, 253
 - iPSC 860, K-40
 - Irwin, M. J., I-65
 - ISA. *See* instruction set architecture
 - issue rates, in multiple-issue processors, 182
 - issue slots, MLT and, 174–175, **174**
 - Itanium 1, G-40 to G-41, K-14
 - Itanium 2. *See* Intel Itanium 2
 - iterative arbiter, E-50
 - iterative division, I-27 to I-31, **I-28**
- J**
- Java bytecodes, K-10
 - Java Virtual Machine (JVM), K-10
 - JBOD, 362
 - JIT (just in time) Java compilers, K-10
 - Johnson, R. B., K-59
 - Jouppi, N. P., K-54
 - JTAG networks, H-42, **H-43**
 - jump prediction, 155, 160–162, **160**, **162**
 - jumps, in control flow instructions, B-16 to B-18, **B-17**, B-37 to B-38, **B-38**
 - just in time (JIT) Java compilers, K-10
- K**
- K6, 294
 - Kahan, W., I-1, I-64
 - k-ary n-cubes, E-38
 - Keller, T. W., F-48
 - Kendall Square Research KSR-1, K-41
 - Kennedy, John F., K-1
 - kernel process, 316
 - kernels
 - defined, 29
 - FFT, H-21 to H-29, **H-23 to H-26**, **H-28 to H-32**
 - Livermore FORTRAN, K-6
 - LU, **H-11**, H-21 to H-26, **H-23 to H-26**, **H-28 to H-32**
 - miss rate in multiprogramming example, 227–228, **229**
 - Kilburn, T., C-38, K-52
 - Kroft, D., K-54
 - Kuck, D., F-51, K-23
 - Kung, H. T., I-63
- L**
- L1 cache. *See also* multilevel caches
 - in AMD Opteron, 327, **327**, **328**, 329, **C-55**
 - average memory access time and, 291
 - memory hierarchy and, **292**
 - miss penalties and, 291
 - multilevel inclusion and, 248
 - size of, 294
 - in Sun T1, 251, **251**
 - in virtual memory, C-46 to C-47, **C-47**
 - L2 cache. *See also* multilevel caches
 - in AMD Opteron, 327, **327**, **334**, **C-55**
 - average memory access time and, 291
 - hardware prefetching and, 305
 - memory hierarchy and, **292**
 - miss penalties and, 291
 - multibanked caches, 299, **309**
 - multilevel inclusion and, 248
 - size of, 293
 - speculative execution and, 325
 - in Sun T1, 251–252, **251**
 - in virtual machines, 323, **323**
 - in virtual memory, C-46 to C-47, **C-47**
 - Lam, M. S., 170
 - lanes, F-6, **F-7**, F-29 to F-31, **F-29**, **F-30**
 - LANS (local area networks), E-4, **E-4**, E-77 to E-79, **E-78**, E-99 to E-100. *See also* interconnection networks

- large-scale multiprocessors, K-40 to K-44
 cache coherence implementation, H-34 to H-41
 classification of, H-44 to H-46, **H-45**
 computation-to-communication ratios in, H-10 to H-12, **H-11**
 hierarchical relationships in, H-45, **H-46**
 IBM Blue Gene/L as, H-41 to H-44, **H-43, H-44**
 interprocessor communication in, H-3 to H-6
 limited buffering in, H-38 to H-40
 message-passing vs. shared-memory communication in, H-4 to H-6
 scientific/technical computing in, H-6 to H-12, **H-11**
 synchronization mechanisms in, H-17 to H-21, **H-19, H-21**
 synchronization performance challenges, H-12 to H-16, **H-14, H-15, H-16**
- latency
 defined, 15, 28, C-2
 in distributed-memory multiprocessors, 201
 effective bandwidth and, E-25 to E-29, **E-27, E-28**
 in Element Interconnect Bus, E-72
 in floating-point MIPS pipelining, A-48 to A-49, **A-49, A-50, A-62, A-65**
 hiding, H-4
 improvements in, compared with bandwidth, 15, **16**
 interconnected nodes and, **E-27**
 in interconnection networks, E-12 to E-19, **E-13**, E-25 to E-29, **E-27**
 I/O, 371
 in Itanium 2, **G-41**
 latency overlap, C-20
 main memory, 310
 packet, E-40 to E-41, E-52 to E-53
 in pipelining, 75–79, A-10
 from remote access communication, 203–204
 in shared-memory multiprocessors, H-29
 transport, E-14
 using speculation to hide, 247–248
 in vector processors, F-3, F-4, F-16, F-31 to F-32, **F-31**
- latent errors, 366–367
 law of diminishing returns, Amdahl's Law and, 40
 LCD (liquid crystal displays), D-19
 learning curve, costs and, 19
 least common ancestor, E-48
 least-recently used (LRU) blocks, C-9, **C-10, C-14, C-43**
- Leighton, F. T., I-65
 limit fields, C-50
 line locking, D-4
 linear speedups, 259–260, **260**
 link injection bandwidth, E-17
 link pipelining, E-16, E-92
 link reception bandwidth, E-17
 link registers, 240, J-32 to J-33
 link-level flow control, E-58, E-62, E-65, E-72, E-74
 Linpack benchmark, F-8 to F-9, F-37 to F-38
- Linux
 dependability benchmarks, 377–379, **378**
 Xen VMM compared with, 322–324, **322, 323**
- liquid crystal displays (LCD), D-19
 LISP, J-30
 literal addressing mode, **B-9, B-10 to B-13, B-11, B-12, B-13**
- Little Endian, B-7, B-34, J-49
 Little's Law, 380–381, 385
 livelock, E-45
 liveness, 74
 Livermore FORTRAN Kernels, K-6
- load and store instructions
 in instruction set architecture classification, B-3, **B-4**
 in MIPS architecture, B-36, **B-36**
 in RISC instruction set, A-4
- load buffers, 94–95, **94, 97, 101, 102–103**
 load delays, 2-cycle, A-59, **A-59**
 load interlocks, A-33 to A-35, **A-34, A-63, A-65**
 load vector count and update (VLVCU), F-18
 load-linked (load-locked) instruction, 239–240
 loads, advanced, G-40
 loads, applied, E-53
 load-store architecture, B-3 to B-6, **B-4, B-6**
 load-store ISAs, 9, F-6, **F-7, F-13 to F-14**
 local address space, C-50
 local area networks (LANs), E-4, **E-4, E-77 to E-79, E-78, E-99 to E-100. See also** interconnection networks
 local miss rate, C-30 to C-33, **C-32**
 local nodes, 232, **233**
 local optimizations, B-26, **B-28**
 local scheduling, 116
 locks
 queuing, H-18 to H-20
 spin lock with exponential back-off, H-17 to H-18, **H-17**
- lockup-free (non-blocking) caches, 296–298, **297, 309, K-54**
- logical units, 390–391
 logical volumes, 390–391
 lognormal distribution, 36–37
 loop interchange, 302–303
 loop unrolling
 dependence analysis and, G-8 to G-9
 eliminating dependent computations in, G-11
 global code scheduling and, 116, G-15 to G-23, **G-16, G-20, G-22**
 hardware-controlled prefetching and, 306
 pipeline scheduling and, 75–80, **75**
 recurrences and, G-11 to G-12
 software pipelining as symbolic, G-12 to G-15, **G-13, G-15**
- loop-carried dependences, G-3 to G-5
 loop-level parallelism
 eliminating dependent computations, G-10 to G-12
 finding dependences, G-6 to G-10

- greatest common divisor test, G-7
 - increasing ILP through, 67–68
 - interprocedural analysis, G-10
 - points-to analysis, G-9 to G-10
 - recurrences, G-5 to G-12
 - type of dependences in, G-2 to G-6
 - loops. *See also* loop unrolling
 - barrier synchronization in, H-14
 - chaining, F-35, **F-35**
 - conditionals in, F-25 to F-26
 - dependence distance in, G-6
 - execution time of vector loop, F-35
 - lossless networks, E-11, E-59, E-65, E-68
 - lossy networks, E-11, E-65
 - LRU (least-recently used) blocks, C-9, **C-10**, C-14, C-43
 - LU kernels
 - characteristics of, H-8, **H-11**
 - on distributed-memory multiprocessors, **H-28 to H-32**
 - on symmetric shared-memory multiprocessors, H-21 to H-26, **H-23 to H-26**
- M**
- M32R
 - addressing modes in, J-5 to J-6, **J-6**
 - architecture overview, **J-4**
 - common extensions in, J-19 to J-24, **J-23, J-24**
 - instructions unique to, J-39 to J-40
 - MIPS core subset in, J-6 to J-16, **J-8, J-9, J-14 to J-17**
 - multiply-accumulate in, J-19, **J-20**
 - MAC (multiply-accumulate), D-5, D-8, J-18, **J-20**
 - machine language programmers, J-84
 - machine memory, defined, 320
 - Macintosh, memory addressing in, K-53
 - magnetic cores, K-4
 - magnetic disks. *See* disk storage; RAID
 - magnetic storage, history of, K-59 to K-61. *See also* storage systems
 - main memory. DRAM chips in, 310
 - Mark-I, K-3
 - MasPar, K-35
 - massively parallel processors (MPP), **H-45**
 - matrix operations, H-7, I-32 to I-33
 - Mauchly, John, K-2 to K-3, K-5
 - maximum transfer unit, E-7, E-76
 - maximum vector length (MVL), F-17, **F-17**
 - Mayer, Milton, E-1
 - McLuhan, Marshall, E-1
 - Mead, C., I-63
 - mean time between failures (MTBF), 26
 - mean time to failure (MTTF), 26–27, 51, 362, 396–397
 - mean time to repair (MTTR), 26–27, 362, 364–365
 - media
 - extensions, D-10 to D-11, **D-11**
 - physical network, E-9
 - shared, E-21, E-23, E-24, E-78
 - switched, E-21, E-24, E-25
 - memory. *See also* caches; virtual memory
 - in embedded computers, 7–8, D-2 to D-3
 - in interconnection networks, 216–217, **216**
 - in vector processors, F-14 to F-16, **F-15**, F-22 to F-23
 - virtual (*See* virtual memory)
 - memory access pipeline cycle, **A-44**, **A-51**, A-52
 - memory access/branch completion cycle
 - in RISC implementation, A-6
 - in shared-memory multiprocessors, H-29 to H-30, **H-32**
 - in unpipelined MIPS implementation, A-28, **A-29**
 - memory accesses per instruction, C-4 to C-6
 - memory addressing, B-7 to B-13
 - addressing modes, B-9 to B-11, **B-9, B-11, B-12**, J-47
 - immediate values, **B-13**
 - in instruction set architectures, 9–10
 - interpreting addresses, B-7 to B-8, **B-8**
 - memory alias analysis, 155, 164–165, **164**
 - memory banks, in vector processors, F-14 to F-16, **F-15**
 - memory consistency models, 243–248
 - coherence and, 207
 - compiler optimization and, 246–247
 - development of, K-44
 - relaxed consistency, 245–246
 - sequential consistency, 243–244, 247
 - synchronized programs and, 244–245
 - using speculation to hide latency in, 247–248
 - memory hierarchy, 287–342, C-1 to C-58. *See also* cache optimizations; virtual memory
 - AMD Opteron data cache example, C-12 to C-14, **C-13, C-15**
 - average memory access time, 290
 - block addressing, 299, **299**, C-8 to C-9, **C-8**
 - block placement in caches, C-7 to C-8, **C-7**
 - block replacement with cache misses, C-9, **C-10**, C-14
 - cache optimization summary, 309, **309**
 - cache organization overview, 288–293, **292**
 - cache performance review, C-3 to C-6, C-15 to C-21, **C-21**
 - cache size and hit time, 293–295, **294, 309**
 - compiler optimizations and, 302–305, **304, 309**
 - compiler-controlled prefetching, 305–309, **309**
 - critical word first and early restart, 299–300, **309**
 - hardware prefetching and, 305, **306, 309**

- memory hierarchy (*continued*)
 - historical perspectives on, K-52 to K-54
 - in IBM Power 5, **341**
 - in Intel Pentium 4, **341**
 - main memory, 310
 - merging write buffers and, 300–301, **301, 309**
 - of microprocessors compared, **341**
 - multibanked caches, 298–299, **299, 309**
 - multilevel inclusion in, C-34
 - nonblocking caches, 296–298, **297, 309**
 - operating system impact on, C-56, **C-57**
 - pipelined cache access and, 296, **309**
 - projections of processor performance, 288, **289**
 - sizes and access times of levels in, **C-3**
 - speculative execution and, 325
 - in Sun Niagara, **341**
 - thrashing, C-25
 - trace caches and, 296, **309**
 - typical multilevel, **288**
 - in virtual memory, C-40, **C-41, C-42 to C-44, C-43, C-46, C-47**
 - way prediction and, 295, **309**
 - writes in caches, C-9 to C-12, **C-10**
- memory indirect addressing mode, **B-9, B-11**
- memory mapping, C-40. *See also* address translations
- memory pipelines, on VMIPS, F-38 to F-40
- memory ports, pipeline stalls and, A-13, **A-14, A-15**
- memory protection, 315–324, C-47 to C-55
 - in 64-bit Opteron, C-53 to C-55, **C-54**
 - architecture provisions for, 316
 - development of, K-52
 - instruction set architecture and, 324–325
 - safe calls from user to OS gates, C-52
 - by virtual machines, 317–324, **322, 323**
 - by virtual memory, 315–317, C-39, C-47 to C-52, **C-51**
- memory reference speculation, G-32, G-40
- memory stall cycles, C-4 to C-6, C-20 to C-21, **C-21**
- memory technology, 310–315. *See also* storage systems
 - DRAM, 311–315, **311, 314**
 - flash memory, 359–360
 - SRAM, 311, F-16
- memory-constrained scaling, H-33 to H-34
- memoryless processes, 384, 386
- memory-memory architecture, B-3, **B-5, B-6**
- memory-memory vector processors, F-4, F-44, F-48
- mesh networks, E-36, **E-40, E-46; E-46**
- MESI protocol, **213**
- Message Passing Interface (MPI), H-5
- message-passing communication, H-4 to H-6
- message-passing multiprocessors, 202
- messages, E-6 to E-8, **E-7, E-76, E-77**
- MFLOPS ratings, F-34 to F-35, F-37, **K-7**
- microinstruction execution, pipelining, A-46 to A-47
- microprocessors
 - costs of, 20, **20**
 - memory hierarchy in, **341**
 - performance milestones in, 15, **16**
 - transistor performance improvements in, 17
- migration of shared data, 207
- MIIPS MDMX, J-16 to J-19, **J-18**
- MIMD. *See* multiple instruction streams, multiple data streams
- minicomputers, beginnings of, 3, 4
- minimal paths, E-45, E-67
- MIPS (Microprocessor without Interlocked Pipeline Stages), J-82, **J-82, K-12**
- MIPS (million instructions per second), 169, A-4 to A-5, **B-19, K-6 to K-7**
- MIPS 1, **J-1, J-6 to J-16, J-7, J-9 to J-13, J-17**
- MIPS 16
 - addressing modes in, J-5 to J-6, **J-6**
 - architecture overview, **J-4**
 - common extensions in, J-19 to J-24, **J-23, J-24**
 - features added to, **J-44**
 - instructions unique to, J-40 to J-42
 - MIPS core subset in, J-6 to J-16, **J-8, J-9, J-14 to J-17**
 - multiply-accumulate in, J-19, **J-20**
 - reduced code size in, B-23
- MIPS 32, **J-80**
- MIPS 64
 - addressing modes in, J-5 to J-6, **J-5**
 - common MIPS extensions in, J-19 to J-24, **J-21 to J-23**
 - instruction set architecture, 10, **11, 12, A-4, B-33, B-34, B-40**
 - unique instructions in, J-24 to J-27, **J-26**
- MIPS architecture, B-32 to B-39
 - addressing modes for, B-34 to B-35
 - ALU instructions, A-4
 - common extensions in, J-19 to J-24, **J-21 to J-24**
 - control flow instructions, B-37 to B-38, **B-38**
 - data types for, B-34
 - in embedded multiprocessors, D-14 to D-15, D-17
 - floating-point operations in, B-38 to B-39, **B-40**
 - instruction format, B-35, **B-35**
 - instruction set usage, 9–10, B-39, **B-41, B-42**
 - operations supported by, B-35 to B-37, **B-36, B-37**
 - processor structure with scoreboard, A-68 to A-69, **A-68**

- recommendations for, B-33
- registers for, B-34
- Tomasulo's approach in, **94**
- unpipelined implementation of, A-26 to A-30, **A-29**
- vector, F-4 to F-6, **F-5, F-7, F-8**
- MIPS M2000, J-81, **J-82**, K-13 to K-14, **K-14**
- MIPS pipeline
 - basic, A-30 to A-33, **A-31, A-32**
 - branch hazards, A-21
 - branches in, A-35 to A-37, **A-38, A-39**
 - control implementation in, A-33 to A-35, **A-34**
 - exceptions in, A-38 to A-41, **A-40, A-42, A-43 to A-45, A-44**
 - floating-point in, A-47 to A-56, **A-48 to A-51, A-57, A-58, A-60 to A-62, A-61 to A-63**
 - ILP limitations in, 167–169
 - instruction set complications, A-45 to A-47
 - loop unrolling in, 76–79
 - MIPS R4000 example, A-56 to A-65, **A-58 to A-65**
 - out-of-order executions in, A-66 to A-67
 - scoreboarding technique in, A-66 to A-75, **A-68, A-71 to A-75**
 - stopping and restarting execution in, A-41 to A-43
- MIPS R1000, 247
- MIPS R2000/3000, A-56
- MIPS R3000, I-12
- MIPS R3010 chip, I-58 to I-60, **I-58, I-59**
- MIPS R4000 pipeline, A-56 to A-65
 - development of, K-19
 - eight-stage structure of, A-56 to A-58, **A-58, A-59, A-60 to A-62, A-61**
 - floating-point pipeline, A-60 to A-62, **A-61, A-62, A-63**
 - forwarding and branch delays in, A-59 to A-60, **A-59, A-60**
 - performance of, A-63 to A-65, **A-64, A-65**
- MIPS R8000, A-43
- MIPS R10000, A-55
- MIPS R12000, 128
- mirroring, 362, **363**, K-61 to K-62
- misalignment, address, B-7 to B-8, **B-8**
- MISD (multiple instruction streams, single data stream), 197
- misprediction rate
 - in Alpha 21264, 89, 140
 - from branch-prediction buffers, 82, **83, 84**
 - for correlating predictors, 86, **87, 88**
 - in Pentium 4, 133–134, **134**
 - in static branch prediction, 81, **81**
 - value prediction and, 130
- miss latency, C-20
- miss penalties
 - block size and, C-26
 - compiler-controlled prefetching, 305–309, **309**
 - CPU time and, C-18
 - critical word first and early restart, 299–300, **309**
 - equation for, 168
 - hardware prefetching and, 305, **306, 309**
 - memory stall cycles and, C-4 to C-6
 - multilevel caches and, 291, C-15 to C-16, C-29 to C-34, **C-32**
 - nonblocking caches and, 296–298, **297, 309**
 - in out-of-order processors, C-19 to C-21, **C-21**
 - read misses and, 291, C-34 to C-35
 - in virtual memory, C-40, C-42
- miss rates
 - associativity and, 291, C-28 to C-29, **C-29**
 - average memory access time and, C-15 to C-16
 - block size and, 291, C-25 to C-28, **C-26, C-27**
 - cache size and, 291, C-28
 - compiler optimizations and, 302–305, **304, 309**
 - compiler-controlled prefetching, 305–309, **309**
 - CPU time and, C-18
 - defined, C-4
 - hardware prefetching and, 305, **306, 309**
 - in instruction vs. data caches, C-14, **C-15**
 - local vs. global, C-30 to C-31
 - main categories of, 290
 - measurement of, C-4 to C-5
 - in multilevel caches, C-30 to C-33, **C-32**
 - process-identifier tags and, C-36, **C-37**
- misses per instruction, 290, C-5 to C-6, C-30 to C-31. *See also* miss rates
- Mitchell, David, K-37
- Mitsubishi M32R. *See* M32R
- mixed caches, C-14
- M/M/I queues, 386
- M/M/m multiple-server model, 388–389, **388**
- MMX, B-31 to B-32, J-46
- Modula-3, **I-12**
- module availability, 26
- module reliability, 26, 49
- modulo scheduling, K-23
- MOESI protocol, **213**
- Moore's Law, 312
- Mosaic, E-98
- Motorola 680x0, **A-40, J-42, K-53**
- Motorola 68882, I-33
- MPI (Message Passing Interface), H-5
- MPP (massively parallel processors), **H-45**
- MSP (Multi-Streaming Processors), F-40 to F-41, **F-41, F-43**
- MTBF (mean time between failures), 26
- MTTF (mean time to failure), 26–27, 51, 362, 396–397
- MTTR (mean time to repair), 26–27, 362, 364–365
- multibanked caches, 298–299, **299, 309**
- multicasting, E-24
- multicomputers, defined, K-39
- multicore processors
 - Element Interconnect Bus, E-70 to E-72, **E-71**
 - MINS compared with, E-92
 - origin of name, 198
 - performance on SPEC benchmarks, 255–257, **255, 256, 257**

- Multiflow processor, K-22 to K-23
- multigrid methods, H-9 to H-10
- multilevel caches, 291, **C-21**, C-29 to C-34, **C-32**, **C-39**. *See also* caches; L1 cache; L2 cache
- multilevel exclusion, C-34
- multilevel inclusion, 248–249, C-34, K-54
- multilevel page tables, C-53, **C-54**
- multimedia support, J-16 to J-19, **J-18**, J-46
- multipath fading, **D-21**
- multiple instruction streams, multiple data streams (MIMD)
 - advantages of, 198
 - Amdahl's Law and, 258–259
 - centralized shared-memory architectures, 199–200, **200**
 - clusters in, 198
 - distributed-memory architectures, 200–201, **201**
 - historical perspectives on, K-36
 - multicore, 198, 199
- multiple instruction streams, single data stream (MISD), 197
- multiple-issue processors,
 - development of, K-20 to K-23. *See also* superscalar processors; VLIW processors
- multiple-precision addition, I-13
- multiplexers
 - in floating-point pipelining, A-54
 - in MIPS pipelining, **A-31**, A-33, A-35, **A-37**
 - in set-associative caches, C-18
- multiplication
 - faster multiplication with many adders, I-50 to I-54, **I-50** to **I-54**
 - faster multiplication with single adders, I-47 to I-50, **I-48**, **I-49**
 - floating-point, I-17 to I-21, **I-18**, **I-19**, **I-20**
 - higher-radix, I-48 to I-49, **I-49**
 - operands of zero, I-21
 - precision of, I-21
 - radix-2 integer, I-4 to I-7, **I-4**, **I-6**
 - shifting over zeros technique, I-45
 - with single adders, I-47 to I-50, **I-48**
 - speeding up, I-47 to I-50, **I-48**, **I-49**
 - system issues in, I-11
 - of two's complement numbers, I-8
- multiply trees, I-52 to I-53, **I-53**
- multiply-accumulate (MAC), D-5, D-8, J-18, **J-20**
- multiply-step instruction, I-11 to I-12
- multiprocessing. *See also* distributed shared-memory multiprocessors; large-scale multiprocessors; symmetric shared-memory multiprocessors
 - advantages of, 196
 - bus-based coherent, K-38 to K-40
 - cache coherence protocols, 205–208, **206**, 211–215, **213**, **214**
 - centralized shared-memory architectures, 199–200, **200**
 - challenges of, 202–204
 - classes of, K-43
 - data stream numbers in, K-38
 - defined, K-39
 - directory-based coherence in, 234–237, **235**, **236**
 - distributed shared memory in, 230–234, **232**, **233**
 - in embedded systems. D-3, D-14 to D-15
 - historical perspectives on, K-34 to K-45
 - invalidate protocol implementation, 209–211
 - large-scale, K-40 to K-44 (*See also* large-scale multiprocessors)
 - limitations in, 216–217
 - memory consistency models in, 243–248
 - message-passing, 202
 - models for communication and memory architecture, 201–202
 - multilevel inclusion, 248–249
 - multiprogramming and OS workload performance, 227–230, **228**, **229**
 - nonuniform memory access in, 202
 - optimizing software for, 261–262
 - reasons for rise of, 262–264
 - references on, K-44 to K-45
 - SMP performance, 218–227, **222** to **226**
 - snooping protocols, 208–209, **209**, 216–218
 - SPEC benchmark performance, 255–257, **255**, **256**, **257**
 - synchronization in, 237–242, **242**
 - T1 processor performance, 249–254, **250** to **254**
 - taxonomy of, 197–201, **201**
 - in vector processors, F-43
- multiprogrammed workloads, 225–230, **227**, **228**, **229**
- multiprogramming. C-47 to C-48
- multistage interconnection networks, E-30, E-92
- Multi-Streaming Processors (MSP), F-40 to F-41, **F-41**, F-43
- multithreading, 172–179. *See also* thread-level parallelism
 - coarse-grained, 173–174, **174**, K-26
 - development of, K-26 to K-27
 - in the directory controller, H-40
 - fine-grained, 173–175, **174**
 - overview, 172–173, 199
 - parallel processing and, 253, **254**
 - processor comparison for, 179–181, **179**, **180**, **181**
 - processor limitations in, 181–183
 - simultaneous, 173–179, **174**, **178**
 - in Sun T1 processor, 250–252, **251**, **252**
- MVL (maximum vector length), F-17, **F-17**
- MXP processor, D-14 to D-15
- Myrinet switches, K-42
- Myrinet-2000, **E-76**
- ## N
- NAK (negative acknowledgment), H-37, H-39 to H-41
- name, defined, 70
- name dependences, 70–71. *See also* antidependences; output dependences
- NaN (Not a Number), I-14, I-16
- NAS parallel benchmarks, F-51

- NaT (Not a Thing), G-38, G-40
 natural parallelism, 172, D-15
n-body algorithms, H-8 to H-9
n-cube, E-36
 NEC SX/2, F-7, **F-34**, F-49
 NEC SX/6, F-7, F-51
 NEC SX/5, F-7, F-50
 NEC SX/7, 338, **339**
 NEC SX/8, F-7, F-51
 NEC VR 4122, D-13, **D-13**
 NEC VR 5432, D-13, **D-13**
 negative acknowledgment (NAK),
 H-37, H-39 to H-41
 negative numbers, I-12, **I-12**, I-14
 nest page tables, 340
 NetApp FAS6000 filer, 397–398
 Netburst design, 131, 137
 Network Appliance, 365, 391,
 397–398
 network attached storage (NAS)
 devices, 391
 network bandwidth
 congestion management and,
 E-65
 performance and, E-18, E-26 to
 E-27, E-52 to E-55, E-89,
 E-90
 switching and, E-50 to E-52
 topologies and, E-40 to E-41
 network file service (NFS), 376, **376**
 Network File System (NFS), 391
 network interface, E-6, E-62, E-67,
 E-76, E-90
 network interface cards (NIC), 322,
 322, E-87, **E-88**
 Network of Workstations, K-42
 network reconfiguration, E-66
 network-on-chip, E-3
 networks. *See also* interconnection
 networks
 centralized switched, E-30 to
 E-34, **E-31**, **E-33**, E-48
 dedicated link, E-5, E-6, **E-6**
 direct, E-34, **E-37**, E-48, E-67,
 E-92
 distributed switched, E-34 to
 E-39, **E-36**, **E-37**, **E-40**, E-46
 dynamic reconfiguration, E-67
 indirect, E-31, E-48, E-67
 local area, E-4, **E-4**, E-77 to E-79,
 E-78, E-99 to E-100
 lossless, E-11, F-59, E-65, E-68
 lossy, E-11, E-65
 mesh, E-36, **E-40**, E-46, **E-46**
 multistage interconnection, E-30,
 E-92
 on-chip, E-3, **E-4**, E-70, **E-73**,
 E-103 to E-104
 performance and cost of, **E-40**
 shared link, E-5
 shared-media, E-21 to E-24, **E-22**,
 E-78
 storage areas, E-3, E-102 to E-103
 switched point-to-point, E-5
 system areas, E-3, E-72 to E-77,
 E-75 to E-77, E-100 to E-102
 wide area, E-4, **E-4**, **E-75**, E-79,
 E-97
 wireless, D-21 to D-22, **D-21**
 NEWS communication, E-41 to E-42
 Newton's iteration, I-27 to I-29, **I-28**
 NFS (Network File System), 32
 NFS (network file service), 376, **376**
 Ngai, T.-F., I-65
 Niagara, K-26
 NIC (network interface cards), 322,
 322, E-87, **E-88**
 Nicely, Thomas, I-64
 Nintendo-64, F-47
 nodes
 home, 232, **233**
 in IBM Blue Gene/L, H-42 to
 H-44, **H-43**, **H-44**
 interconnected, **E-27**
 local, 232, **233**
 remote, 233, **233**
 X1, F-42, **F-42**
 nonaffine array indexes, G-6
 nonaligned data transfers, J-24 to J-26,
 J-26
 nonatomic operations, 214
 nonbinding prefetches, 306
 nonblocking caches, 296–298, **297**,
 309, K-54
 non-blocking networks, E-32, E-35,
 E-41, E-56
 nonfaulting prefetches, 306
 non-minimal paths, E-45
 nonrestoring division algorithm, I-5 to
 I-7, **I-6**, I-45 to I-47, **I-46**
 nonuniform memory access (NUMA)
 multiprocessors, 202. *See also* distributed
 shared-memory (DSM)
 multiprocessors
 nonunit strides, F-21 to F-22, F-46,
 F-48
 normal distribution, 36
 no-write allocate, C-11 to C-12
 nullification, A-24 to A-25, J-33 to
 J-34
n-way set associative cache placement,
 289, **C-7**, C-8
- O**
 occupancy, message, H-3 to H-4
 Ocean application
 characteristics of, H-9 to H-12,
 H-11
 on distributed-memory
 multiprocessors, H-28, **H-28**
 to **H-32**, H-30
 on symmetric shared-memory
 multiprocessors, H-21 to
 H-26, **H-23 to H-26**
 OCN (on-chip networks), E-3, **E-4**,
 E-70, **E-73**, E-103 to E-104
 octrees, H-9
 off-load engines, E-8, E-77, E-92
 offset, in RISC architectures, A-4 to
 A-5
 OLTP. *See* online transaction
 processing
 Omega, E-30, **E-31**
 on-chip multiprocessing, 198, 205. *See also*
 multicore processors
 on-chip networks (OCN), E-3, **E-4**,
 E-70, **E-73**, E-103 to E-104
 one's complement system, I-7
 one-third-distance rule of thumb,
 401–403, **403**
 one-way conflict misses, C-25
 online transaction processing (OLTP)
 benchmarks for, 374–375, **375**
 performance evaluation for, 46,
 47, **48**
 in shared-memory
 multiprocessors, 220–224,
 221, **222**
 OOO (out-of-order) processors, miss
 penalty and, C-19 to C-21,
 C-21
 opcode field
 encoding instruction sets, B-21 to
 B-24, **B-22**

- opcode field (*continued*)
 - in MIPS instruction, **B-35**, **B-35**
 - operand type and, B-13
 - Open Systems Interconnect (OSI), E-81, **E-82**
 - OpenMP consortium, H-5
 - operands
 - address specifiers for, B-21
 - decimal, B-14
 - in instruction encoding, **B-22**
 - instruction set architecture
 - classification and, B-3, **B-4**, **B-5**, **B-6**
 - in Intel 80x86, J-59 to J-62, **J-59 to J-62**
 - in MIPS, 10
 - shifting, J-36
 - type and size of, B-13 to B-14, **B-15**
 - in VAX, J-67 to J-68, **J-68**
 - operating systems
 - asynchronous I/O and, 391
 - disk accesses in, 400–401, **401**
 - memory hierarchy performance and, C-56, **C-57**
 - multiprogrammed workload performance, 225–230, **227**, **228**, **229**
 - page size changes and, C-56 to C-57
 - Ultrix, **C-37**
 - user access to, C-52
 - in virtual machines, 318, 319, 320
 - operation faults, 367, **370**
 - operations, in instruction sets, B-14 to B-16, **B-15**, **B-16**
 - operator dependability, 369–371
 - Opteron processor. *See* AMD Opteron processor
 - order of instruction exceptions, A-38 to A-41, **A-40**, **A-42**
 - organization, defined, 12
 - orthogonal architectures, B-30, J-83, K-11
 - OS. *See* operating systems
 - OSI (Open Systems Interconnect), E-81, **E-82**
 - Otellini, Paul, 195
 - out-of-order completion, 90–91, A-54, A-66
 - out-of-order execution, 90–91, A-66 to A-67, A-75 to A-76, C-3. *See also* scoreboarding
 - out-of-order (OOO) processors, miss penalties and, C-19 to C-21, **C-21**
 - output dependences, 71, K-23
 - output-buffered switches, E-57, E-59
 - overflow, I-8, I-10 to I-12, **I-11**, I-20
 - overhead
 - occupancy and, H-4
 - packet switching and, E-51
 - receiving, E-14, E-17, E-63, E-76, E-88, E-92
 - routing algorithms and, E-48
 - sending, E-14, E-16, E-63, E-76, E-92
 - overlapping triplets, I-49, **I-49**
 - overlays, C-39
 - owner of a cache block, 211, 231, **235**
- P**
- Pacifica, 320, 339
 - packets
 - in asynchronous transfer mode, E-79
 - discarding, E-65
 - in Element Interconnect Bus, E-71 to E-72
 - headers, E-6, E-48, E-52, E-57 to E-58, E-60, E-72
 - in IBM Blue Gene/L 3D Torus, E-72
 - in InfiniBand, E-76
 - latency, E-12, **E-13**
 - size of, E-18, **E-19**
 - switching, E-50, E-77
 - trailers, E-6, **E-7**, E-61
 - transport, E-8 to E-9, E-94
 - packing operations, B-14
 - Padua, D., F-51
 - page allocation, 262
 - page coloring, C-37
 - page faults, C-3, C-40
 - page offsets, 291, C-38, C-42
 - page remapping, in virtual machines, 324
 - page sizes, C-45 to C-46, C-56 to C-57
 - page tables
 - inverted, C-43
 - memory protection and, 316–317
 - multilevel, C-53 to C-54, **C-54**
 - nested, 340
 - page sizes and, C-45, C-53
 - paging of, C-44
 - process protection and, C-48
 - shadow, 320
 - in virtual memory mapping, **C-43**
 - page table entries (PTEs)
 - in AMD Opteron, 326–327, C-54
 - in Intel Pentium, C-50
 - in virtual memory, C-43, **C-43**
 - paged segments, C-41, **C-42**
 - page-level protection, C-36
 - pages
 - in 64-bit Opteron memory management, C-53 to C-55, **C-54**
 - in virtual memory, C-30, C-41 to C-42, **C-41**, **C-42**
 - paired single operations, B-39, D-10 to D-11
 - PAL code, J-28
 - Panda, D. K., **E-77**
 - Paragon, K-40
 - parallel, defined, 68
 - parallel processing
 - historical perspectives on, K-34 to K-36
 - in large-scale multiprocessors, H-2
 - performance with scientific applications, H-33 to H-34
 - parallelism. *See also* instruction-level parallelism; thread-level parallelism
 - Amdahl's Law and, 258–259
 - challenges of, 202–204
 - data dependences and, 68–70
 - data-level, 68, 197, 199
 - at the detailed digital design level, 38
 - explicit, G-34 to G-37, **G-35**, **G-36**, **G-37**
 - hardware vs. software approach to, 66
 - historical perspectives on, K-24 to K-25, K-34 to K-36
 - implicit, G-34
 - at the individual processor level, 37

- multithreading and, 253, **254**
- natural, 172, D-15
- in scoreboarding, A-74
- at the system level, 37
- taxonomy of, 197–201, **200, 201**
- in vector processing, F-29 to F-31, **F-29, F-30**
- paravirtualization, 321–324
- PA-RISC
 - common MIPS extensions in, J-19 to J-24, **J-21 to J-23**
 - conditional branch options in, **B-19**
 - extended precision in, I-33
 - features added to, **J-44**
 - instructions unique to, J-33 to J-36, **J-34**
 - MIPS core subset in, J-6 to J-16, **J-7, J-9 to J-13, J-17**
- PA-RISC 1.1, **J-4**
- PA-RISC 2.0, J-5 to J-6, **J-5**
- PA-RISC MAX2, J-16 to J-19, **J-18**
- partial store order, 246
- partitioned add operations, D-10
- Pascal, integer division and remainder in, **I-12**
- passes, optimizing, B-25, **B-25**
- path loss, **D-21**
- Patterson, D. A., K-12 to K-13
- payload, E-6, E-61
- PCI-Express (PCIe), E-29, E-63
- PC-relative addressing, B-10, B-18
- PC-relative control flow instructions, B-17
- PDP-11
 - address size in, C-56
 - memory caches in, K-53
 - memory hierarchy in, K-52
 - Unibus, K-63
- peak performance, 51, **52**
- peer-to-peer architectures, D-22
- peer-to-peer communication, E-81 to E-82
- Pegasus, K-9
- Pentium. *See* Intel Pentium; Intel Pentium 4; Intel Pentium 4 Extreme
- Pentium chip, division bug in, I-2, I-64 to I-65
- Pentium D, 198
- Pentium III, 183
- Pentium M, **20**
- Pentium MMX, **D-11**
- Perfect Club benchmarks, F-51
- perfect-shuffle permutation, E-30
- performance. *See also* benchmarks:
 - cache performance; processor performance
 - Amdahl's Law and, 184
 - average memory access time and, C-17 to C-19
 - bandwidth and, E-16 to E-19, **E-19, E-25 to E-29, E-28, E-89, E-90**
 - of branch schemes, A-25 to A-26, **A-26**
 - cache misses and, C-17 to C-19
 - cache size and, H-22, H-24, **H-24, H-27, H-28**
 - of commercial workload, 220–230, **221 to 229**
 - of compilers, B-27, **B-29**
 - contention and, E-25, E-53
 - of desktop computers, 44–46, **45, 46**
 - development of measures of, K-6 to K-7
 - of DRAM, 312–315, **313, 314**
 - effective bandwidth and, E-16 to E-19, **E-19, E-25 to E-29, E-28, E-89, E-90**
 - Ethernet, E-89, **E-90**
 - of floating-point operations, 3
 - flow control and, E-17
 - I/O, 371–379, **372 to 376, 378**
 - of multicore processors, 255–257, **255, 256, 257**
 - of multiprocessors, 218–230, 249–257
 - of online transaction processing, 46, **47, 48**
 - peak, 51, **52**
 - pipeline stalls and, A-11 to A-13
 - real-time, 7, D-3
 - of scientific applications, H-21 to H-26, **H-23 to H-26**
 - of servers, 46–48, **47, 48**
 - simultaneous multithreading and, 177–179, **178**
 - of superscalar processors, **16, 179–181, 179, 180, 181**
 - topology and, E-40 to E-44, **E-44, E-52**
 - transistors and, 17–19
 - of vector processors, F-34 to F-38, **F-35, F-40, F-44 to F-45, F-45**
 - virtual channels and, E-93
 - in VMIPS, F-36 to F-38
- periodic functions, I-32
- permanent failures, E-66
- permanent faults, 367
- PetaBox GB2000, 393, **394**
- phase-ordering problem, B-26
- phases (passes), optimizing, B-25, **B-25**
- phits, E-60, E-62, E-71
- physical caches, defined, C-36
- physical channels, E-47
- physical memory, in virtual machines, 320
- physical volumes, 390–391
- pi (p) computation, I-32
- PID (process-identifier tags), C-36, **C-37**
- piggyback acknowledgment field, E-84
- Pinkston, T. M., E-104
- pin-out constraint, E-39, E-71, E-89
- pipe stages, A-3, A-7
- pipeline bubbles, A-13, A-20. *See also* pipeline stalls
- pipeline depths, F-12 to F-13
- pipeline hazards, A-11 to A-26. *See also* dependences
 - control hazards, A-11, A-21 to A-26, **A-21 to A-26**
 - data hazards, A-11, A-15 to A-21, **A-16 to A-21**
 - detection of, A-33 to A-35, **A-34**
 - in floating-point pipelining, A-49 to A-54, **A-51, A-57, A-58, A-61 to A-65, A-61 to A-63**
 - load interlocks, A-33 to A-35, **A-34**
 - in longer latency pipelines, A-49 to A-54, **A-50, A-51**
 - multicycle operations and, A-46 to A-47
 - performance of pipelines with stalls, A-11 to A-13
 - structural hazards, A-11, A-13 to A-15, A-64, **A-65**
- pipeline latches, A-30, **A-36**
- pipeline registers, A-8 to A-10, **A-9, A-30, A-35**

- pipeline reservation tables, K-19
- pipeline scheduling, loop unrolling
 - and, 75–80, **75**, 117–118
- pipeline stalls
 - bubbles, A-13, A-20
 - data hazards requiring stalls, A-19 to A-20, **A-20**, **A-21**
 - diagrams of, A-13, **A-15**
 - in floating-point pipelines, A-51, **A-51**
 - minimizing by forwarding, A-17 to A-18, **A-18**, A-35, **A-36**, **A-37**
 - in MIPS pipelines, A-33 to A-35, **A-34**
 - in MIPS R4000 pipeline, A-63 to A-65, **A-63**, **A-64**
 - performance and, A-11 to A-13
 - in SMPs, **222**
 - in vector processors, F-9 to F-10
- pipelined cache access, 296, **309**
- pipelined circuit switching, E-50, E-71
- pipelines, self-draining, K-21
- pipelining, A-2 to A-77
 - in addition, I-25
 - basic MIPS, A-30 to A-33, **A-31**, **A-32**
 - condition codes in, A-5, A-46
 - data dependences and, 69–70
 - depth of, A-12
 - dynamic scheduling in, A-66 to A-75, **A-68**, **A-71**, **A-73** to **A-75**
 - in embedded systems, D-7 to D-10, **D-7**
 - encoding instruction sets and, B-21 to B-22
 - exceptions in, A-38 to A-41, **A-40**, **A-42**
 - five-stage pipeline for RISC processors, A-6 to A-10, **A-7**, **A-8**, **A-9**, **A-21**
 - floating-point, A-47 to A-56, **A-48** to **A-51**, **A-57**, **A-58**, A-60 to A-62, **A-61** to **A-63**
 - freezing/flushing, A-22
 - historical perspectives on, K-10, K-18 to K-27
 - increasing instruction fetch bandwidth in, 121–127, **122**, **124**, **126**
 - in interconnection networks, E-12, E-25, E-51 to E-52, E-60, E-65, E-70
 - interlocks, A-20, A-33 to A-35, **A-34**, A-52, F-9 to F-10
 - in Itanium 2 processor, G-42
 - link, E-16, E-92
 - microinstruction execution, A-46 to A-47
 - MIPS branches in, A-35 to A-37, **A-38**, **A-39**
 - multicycle operations and, A-46 to A-47
 - in multiplication, I-51
 - overview of, 37, A-2 to A-3
 - in Pentium 4, 131–132, **132**, **133**
 - performing issues in, A-10 to A-11
 - SMP stalls in, **222**
 - software, D-10, G-12 to G-15, **G-13**, **G-15**
 - stopping and restarting execution, A-41 to A-43
 - superpipelining, A-57
 - in switch microarchitecture, E-60 to E-61, **E-60**
 - in vector processors, F-31 to F-32, **F-31**
- Pleszkun, A. R., A-55, K-22
- pointers
 - current frame, G-33 to G-34
 - dependences and, G-9
 - function, B-18
 - urgent pointer field, E-84
 - in VAX, J-71
- points-to analysis, G-9 to G-10
- point-to-point links, 390, E-24, E-29, E-79
- poison bits, G-28, G-30 to G-32
- Poisson distribution, 384–390, **388**
- Poisson processes, 384
- polycyclic scheduling, K-23
- Popok, Gerald, 315
- POPF, 338
- position independence, B-17
- postbytes, J-57, **J-57**
- POWER, **J-44**
- power
 - in cell phones, D-22, D-24
 - dynamic, 18–19
 - EEMBC benchmarks for consumption of, D-13, **D-13**
 - in multiple-issue processors, 182
 - redundancy of supplies, 27–28
 - reliability of, 49
 - static, 19
 - transistor and wire scaling and, 17–19
- Power processors, 128
- Power2 processor, 130, A-43
- Power4 processor, **52**
- Power5 processor. *See* IBM Power5 processor
- PowerEdge 1600SC, **323**
- PowerEdge 2800, **47**, **48**, **49**
- PowerEdge 2850, **47**, **48**, **49**
- PowerPC
 - addressing modes in, J-5 to J-6, **J-5**
 - AltiVec in, F-47
 - common extensions in, J-19 to J-24, **J-21** to **J-23**
 - conditional branch options in, **B-19**
 - features added to, **J-44**
 - instructions unique to, J-32 to J-33
 - MIPS core subset in, J-6 to J-16, **J-7**, **J-9** to **J-13**, **J-17**
 - multimedia support in, J-16 to J-19, **J-18**
 - performance per watt in, D-13, **D-13**
 - reduced code size in, B-23
- PowerPC 620, A-55
- PowerPC AltiVec, B-31, **D-11**
- precise exceptions, A-43, A-54 to A-56
- Precision Workstation 380, **45**
- predicated instructions, G-23 to G-27
 - annulling instructions, G-26
 - in ARM, J-36
 - concept behind, G-23
 - conditional moves in, G-23 to G-24
 - exceptions in, G-25 to G-26
 - in Intel IA-64, G-38, **G-39**
 - limitations of, G-26 to G-27
 - moving time-critical, G-24 to G-25
- predication, D-10
- predicted-not-taken scheme, A-22, **A-22**, **A-25**, A-26, **A-26**
- predicted-taken scheme, A-23, **A-25**, A-26, **A-26**

- prefetching
 - in AMD Opteron, 330
 - compiler-controlled, 305–309, **309**
 - development of, K-54
 - hardware, 305, **306, 309**
 - instruction set architecture and, B-46
 - integrated instruction fetch units and, 126
 - in RISC desktop architectures, **J-21**
 - prefixes, in instructions, J-51, J-55
 - present bits, C-50
 - price-performance, in desktop computers, 5, **5**
 - prices vs. costs, 25–28
 - primitives, 239–240, H-18 to H-21, **H-21**
 - principle of locality, 38, 288, C-2
 - private data, 205
 - probability mass function, 384
 - procedure invocation options, B-19 to B-20
 - process switch, 316, C-48
 - processes
 - defined, 199, 316, C-47 to C-48
 - protection of, C-48 to C-49
 - process-identifier tags (PID), C-36, **C-37**
 - processor consistency, 245
 - processor cycle, A-3
 - processor performance, 28–44. *See also* benchmarks
 - Amdahl's Law and, 39–42, 184
 - average memory access time and, C-17 to C-19
 - benchmarks in, 29–33, **31, 35**
 - of desktop systems, 45–46, **45, 46**
 - equation for, 41–44
 - execution time in, 28–29
 - focusing on the common case, 38
 - parallelism and, 37–38
 - peak performance, 51, **52**
 - price and, 45–46, **45, 46**
 - principle of locality in, 38
 - real-time, D-3 to D-5
 - summarizing benchmark results, 33–37, **35**
 - using benchmarks to measure, 29–33, **31**
 - using parallelism to improve, 37–38
 - processor-dependent optimizations, B-26, **B-28**
 - processors. *See also* digital signal processors; multiprocessing; superscalar processors; vector processors; VLIW processors; *names of specific processors*
 - array, K-36
 - directory-based multiprocessors, H-29, **H-31**
 - in embedded computers, 7–8
 - importance of cost of, 49–50
 - massively parallel, **H-45**
 - microprocessors, 5, 15, **16, 17, 20, 20, 341**
 - multicore, 255–257, **255, 256, 257**, E-70 to E-72, E-92
 - out-of-order, C-19 to C-21, **C-21**
 - performance growth since mid-1980s, 2–4, **3**
 - Single-Streaming, F-40 to F-41, **F-41, F-43**
 - VPU, D-17 to D-18
 - producer-server model, 371, **372**
 - profile-based predictors, 161, **162**
 - program counter (PC), B-17. *See also* PC-relative addressing
 - program order. *See also* out-of-order completion; out-of-order execution
 - control dependences and, 72–74
 - data hazards and, 71
 - memory consistency and, 243–246
 - in shared-memory multiprocessors, 206
 - propagation delay, E-10, E-13, E-14, E-25, E-40
 - protection. *See also* memory protection
 - in 64-bit Opteron memory management, C-53 to C-55, **C-54, C-55**
 - call gates, C-52
 - capabilities in, C-48 to C-49
 - in Intel Pentium, C-48 to C-52, **C-51**
 - rings in, C-48
 - protocol families, E-81
 - protocol fields, E-84
 - protocol stacks, E-83, **E-83**
 - protocols, E-8, E-62, E-77, E-91, E-93. *See also* names of specific protocols
 - PS2. *See* Sony Playstation 2
 - PTE. *See* page table entries
- ## Q
- QR factorization method, H-8
 - QsNet, **E-76**
 - quad precision, I-33
 - queue, 380
 - queue depth, 360, **360**
 - queue discipline, 382
 - queuing locks, H-18 to H-20
 - queuing theory, 379–390
 - examples of, 387
 - overview, 379–382, **379, 381**
 - Poisson distribution of random variables in, 382–390, **388**
- ## R
- racers, 218, 245
 - radio waves, D-21 to D-22, **D-23**
 - RAID (redundant arrays of inexpensive disks). *See also* disk arrays
 - availability benchmark, 377, **378**
 - development of, K-61 to K-62
 - levels of, 362–366, **363, 365**
 - logical units in, 391
 - reliability, 400
 - RAID-DP (row-diagonal parity), 365–366, **365**
 - RAMAC-350, K-59 to K-60, K-62 to K-63
 - RAMBUS, 336
 - random block replacement, C-9, **C-10**
 - random variables, distributions of, 382–390, **388**
 - RAS (row access strobe), 311–312, **313**
 - Rau, B. R., K-23
 - RAW (read after write) hazards
 - in floating-point MIPS pipelines, **A-50, A-51 to A-53, A-51**
 - hardware-based speculation and, 112, 113
 - as ILP limitations, 71

- RAW (read after write) hazards
(continued)
 load interlocks and, A-33
 in scoreboarding, A-69 to A-70, A-72
 Tomasulo's approach and, 92
- read miss
 directory protocols and, 231, **233**, 234–237, **236**
 miss penalty reduction and, 291, C-34 to C-35, **C-39**
 in Opteron data cache, C-13 to C-14
 in snooping protocols, 212, **213**, **214**
- real addressing mode, J-45, **J-50**
- real memory, in virtual machines, 320
- real-time constraints, D-2
- real-time performance, 7, D-3
- rearrangeably non-blocking networks, E-32
- receiving overhead, E-14, E-17, E-63, E-76, E-92
- reception bandwidth, E-18, E-26, E-41, E-55, E-63, E-89
- RECN (regional explicit congestion notification), E-66
- reconfiguration, E-45
- recovery time, F-31, **F-31**
- recurrences, G-5, G-11 to G-12
- red-black Gauss-Seidel multigrid technique, H-9 to H-10
- Reduced Instruction Set Computer architectures. *See* RISC (Reduced Instruction Set Computer) architectures
- redundant arrays of inexpensive disks. *See* RAID
- redundant quotient representation, I-47, I-54 to I-55, **I-55**
- regional explicit congestion notification (RECN), E-66
- register addressing mode, **B-9**
- register fetch cycle, A-5 to A-6, A-26 to A-27, **A-29**
- register indirect addressing mode jumps, B-17 to B-18, **B-18**
 in MIPS data transfers, B-34
 overview of, **B-9**, B-11, **B-11**
- register prefetch, 306
- register pressure, 80
- register renaming
 finite registers and, 162–164, **163**
 in ideal processor, 155, 157
 name dependences and, 71
 reorder buffers vs., 127–128
 in Tomasulo's approach, 92–93, 96–97
- register rotation, G-34
- register stack engine, G-34
- register windows, J-29 to J-30
- register-memory ISAs, 9, B-3, **B-4**, **B-5**, **B-6**
- register-register architecture, B-3 to B-6, **B-4**, **B-6**
- registers
 base, A-4
 branch, J-32 to J-33
 count, J-32 to J-33
 current frame pointer, G-33 to G-34
 finite, effect on ILP, 162–164, **163**
 floating-point, A-53, B-34, B-36
 general-purpose, B-34
 history and future files, A-55
 in IBM Power5, 162
 instruction encoding and, B-21
 in instruction set architectures, 9, **9**
 integer, B-34
 in Intel 80x86, J-47 to J-49, **J-48**, **J-49**
 in Intel IA-64, G-33 to G-34
 link, 240, J-32 to J-33
 loop unrolling and, 80
 in MIPS architecture, B-34
 in MIPS pipeline, A-30 to A-31, **A-31**
 number required, B-5
 pipeline, A-8 to A-10, **A-9**
 predicate, G-38, **G-39**
 in RISC architectures, A-4, A-6, A-7 to A-8, **A-8**
 in scoreboarding, A-71, A-72
 in software pipelining, G-14
 in Tomasulo's approach, 93, **99**
 in VAX procedure, J-72 to J-76, **J-75**, J-79
 vector-length, F-16 to F-18
 vector-mask, F-26
 in VMIPS, F-6, **F-7**
 VS, F-6
- regularity, E-33, E-38
- relative speedup, 258
- relaxed consistency models, 245–246
- release consistency, 246, K-44
- reliability
 Amdahl's Law and, 49
 benchmarks of, 377–379, **378**
 defined, 366–367
 "five nines" claims of availability, 399–400, **400**
 implementation location and, 400
 in interconnection networks, E-66
 module, 26
 operator, 369–371
- relocation, C-39
- remote memory access time, H-29
- remote nodes, 233, **233**
- renaming. *See* register renaming
- renaming maps, 127–128
- reorder buffers (ROB)
 development of, K-22
 in hardware-based speculation, 106–114, **107**, **110**, **111**, **113**, G-31 to G-32
 renaming vs., 127–128
 in simultaneous multithreading, 175
- repeat (initiation) intervals, A-48 to A-49, **A-49**, **A-62**
- repeaters, E-13
- replication of shared data, 207–208
- requested protection level, C-52
- request-reply, E-45
- reservation stations, 93, **94**, 95–97, **99**, **101**, 104
- resource sparing, E-66, E-72
- response time. *See also* execution time; latency
 defined, 15, 28, 372
 throughput vs., 372–374, **373**, **374**
- restarting execution, A-41 to A-43
- restoring division algorithm, I-5 to I-7, **I-6**
- restricted alignment, B-7 to B-8, **B-8**
- resuming events, A-41, A-42
- return address predictors, 125, **126**, K-20
- returns, procedure, B-17 to B-19, **B-17**
- reverse path, in cell phone base stations, D-24
- rings, C-48, E-35 to E-36, **E-36**, **E-40**, E-70

- ripple-carry addition, I-2 to I-3, **I-3**, **I-42**, **I-44**
 - RISC (Reduced Instruction Set Computer) architectures, J-1 to J-90
 - ALU instructions in, A-4
 - classes of instructions in, A-4 to A-5
 - digital signal processors in embedded, J-19
 - five-stage pipeline for, A-6 to A-10, **A-7**, **A-8**, **A-9**, **A-21**
 - historical perspectives on, 2, K-12 to K-15, **K-14**
 - lineage of, **J-43**
 - MIPS core extensions in, J-19 to J-24, **J-21** to **J-24**
 - MIPS core subsets in, J-6 to J-16, **J-7** to **J-16**
 - multimedia extensions in, J-16 to J-19, **J-18**
 - overview of, A-4 to A-5, **J-42**
 - pipelining efficiency in, A-65 to A-66
 - reduced code size in, B-23 to B-24
 - simple implementation without pipelining, A-5 to A-6
 - unique instructions in, J-24 to J-42, **J-26**, **J-31**, **J-34**
 - virtualization of, 320
 - RISC-I/RISC-II, K-12 to K-13
 - ROB. *See* reorder buffers
 - rotate with mask instructions, J-33
 - rounding
 - double, I-34, I-37
 - in floating-point addition, I-22
 - in floating-point division, I-27, I-30
 - in floating-point multiplication, I-17 to I-18, **I-18**, **I-19**, **I-20**
 - floating-point remainders, I-31
 - fused multiply-add, I-32 to I-33
 - in IEEE floating-point standard, I-13 to I-14, **I-20**
 - precision and, I-34
 - underflow and, I-36
 - round-off errors, D-6, **D-6**
 - round-robin, E-49, E-50, E-71, E-74
 - routing
 - adaptive, E-47, E-53 to E-54, **E-54**, E-73, E-93 to E-94
 - algorithm for, E-45, E-52, E-57, E-67
 - deterministic, E-46, E-53 to E-54, **E-54**, E-93
 - packet header information, **E-7**, E-21
 - in shared-media networks, E-22 to E-24, **E-22**
 - switch microarchitecture and, **E-57**, E-60 to E-61, **E-61**
 - in switched-media networks, E-24
 - routing algorithm, E-45, E-52, E-57, E-67
 - row access strobes (RAS), 311–312, **313**
 - row major order, 303
 - row-diagonal parity (RAID-DP), 365–366, **365**
 - Rowen, C., I-58
 - RP3, K-40
 - RS 6000, K-13
- S**
- SAGE, K-63
 - Saltzer, J. H., E-94
 - SAN (system area networks), E-3, E-72 to E-77, **E-75** to **E-77**, E-100 to E-102. *See also* interconnection networks
 - Santayana, George, K-1
 - Santoro, M. R., I-26
 - Sanyo VPC-SX500 digital camera, D-19, **D-20**
 - SAS (Serial Attach SCSI), 361, **361**
 - SATA disks, 361, **361**
 - saturating arithmetic, D-11, J-18 to J-19
 - scalability, 37, 260–261, K-40 to K-41
 - scaled addressing mode, **B-9**, **B-11**, J-67
 - scaled speedup, 258–259, H-33 to H-34
 - scaling, 17–19, 259, H-33 to H-34
 - Scarott, G., K-53
 - scatter-gather operations, F-27 to F-28, F-48
 - scheduling, historical perspectives on, K-23 to K-24
 - Schneck, P. B., F-48
 - scientific/technical computing, H-6 to H-12
 - Barnes application, H-8 to H-9, **H-11**
 - computation-to-communication ratio in, H-10 to H-12, **H-11**
 - on distributed-memory multiprocessors, H-26 to H-32, **H-28** to **H-32**
 - FFT kernels, H-7, **H-11**, H-21 to H-29, **H-23** to **H-26**, **H-28** to **H-32**
 - LU kernels, H-8, **H-11**, H-21 to H-26, **H-23** to **H-26**, **H-28** to **H-32**
 - need for more computation in, 262
 - Ocean application, H-9 to H-12, **H-11**
 - parallel processor performance in, H-33 to H-34
 - on symmetric shared-memory multiprocessors, H-21 to H-26, **H-23** to **H-26**
 - scoreboarding, A-66 to A-75
 - basic steps in, A-69 to A-70, A-72, **A-73**, **A-74**
 - costs and benefits of, A-72 to A-75, **A-75**
 - data structure in, A-70 to A-72, **A-71**
 - development of, 91, K-19
 - goal of, A-67
 - in Intel Itanium 2, G-42, G-43
 - structure of, A-67 to A-68, **A-68**, **A-71**
 - scratch pad memory (SPRAM), D-17, **D-18**
 - SCSI (small computer systems interface), 360–361, **361**, K-62 to K-63
 - SDRAM (synchronous DRAM), 313–314, 338, **338**
 - SDRWAVE, I-62
 - sector-track cylinder model, 360–361
 - security. *See* memory protection
 - seek distances and times, 401–403, **402**, **403**
 - segments
 - segment descriptors, C-50 to C-51, **C-51**
 - in virtual memory, C-40 to C-42, **C-41**, **C-42**, C-49
 - self-draining pipelines, K-21

- self-routing property, E-48
- semantic clash, B-41
- semantic gap, B-39, B-41, K-11
- sending overhead, E-14, E-16, E-63, E-76, E-92
- sense-reversing barriers, H-14 to H-15, **H-15**, H-21, **H-21**
- sentinels, G-31, K-23
- sequential consistency, 243–244, K-44
- sequential interleaving, 299, **299**
- serial advanced technology attachment (SATA), 361, **361**, E-103
- Serial Attach SCSI (SAS), 361, **361**
- serialization, 206–207, H-16, H-37
- serpentine recording, K-59
- serve-longest-queue, E-49
- server benchmarks, 32–33
- servers
 - characteristics of, **D-4**
 - defined, 380, **381**
 - downtime costs, **6**
 - instruction set principles in, B-2
 - memory hierarchy in, **341**
 - operand type and size in, B-13 to B-14
 - performance and
 - price-performance of, 46–48, **47, 48**
 - price range of, 5, 6–7
 - requirements of, 6–7
 - transaction-processing, 46–48, **47, 48**
 - utilization, 381, 384–385, 387
- Service Level Agreements (SLA), 25–26
- Service Level Objectives (SLO), 25–26
- service specification, 366
- set-associative caches
 - defined, 289
 - miss rate and, C-28 to C-29, **C-29**
 - n-way cache placement, 289, **C-7, C-8**
 - parallelism in, 38
 - structure of, C-7 to C-8, **C-7, C-8**
- sets, defined, C-7
- settle time, 402
- SFS benchmark, 376
- SGI Altix 3000, **339**
- SGI Challenge, K-39
- SGI Origin, H-28 to H-29, **H-31**, K-41
- shadow fading, **D-21**
- shadow page tables, 320
- shadowing (mirroring), 362, **363**, K-61 to K-62
- shared data, 205
- shared link networks, E-5
- shared memory. *See also* distributed shared-memory multiprocessors; symmetric shared-memory multiprocessors
 - communication, H-4 to H-6
 - defined, 202
 - multiprocessor development, K-40
 - synchronization, J-21
- shared-media networks, E-21 to E-25, **E-22**, E-78
- shared-memory communication, H-4 to H-6
- shifting over zeros technique, I-45 to I-47, **I-46**
- shortest path, E-45, E-53
- sign magnitude system, I-7
- signal processing, digital, D-5 to D-7, **D-6**
- signals, in embedded systems, D-2
- signal-to-noise ratio (SNR), D-21, **D-21**
- signed numbers, I-7 to I-10, I-23, **I-24**, I-26
- signed-digit trees, I-53 to I-54
- sign-extended offsets, A-4 to A-5
- significands, I-15
- Silicon Graphics MIPS 1. *See* MIPS 1
- Silicon Graphics MIPS 16. *See* MIPS 16
- SIMD (single instruction stream, multiple data streams)
 - compiler support for, B-31 to B-32
 - defined, 197
 - in desktop processors, D-11, **D-11**
 - in embedded systems, D-10, **D-16**
 - historical perspectives on, K-34 to K-36
 - Streaming SIMD Extension, B-31
- simultaneous multithreading (SMT), 173–179
 - approaches to superscalar issue slots, 174–175, **174**
 - design challenges in, 175–177
 - development of, K-26 to K-27
 - potential performance advantages
 - from, 177–179, **178**
 - preferred-thread approach, 175–176
- single extended precision, **I-16**, I-33
- single instruction stream, multiple data streams (SIMD). *See* SIMD
- single instruction stream, single data streams (SISD), 197, K-35
- single-chip multiprocessing, 198. *See also* multicore processors
- single-precision numbers
 - IEEE standard on, **I-16**, I-33
 - multiplication of, I-17
 - representation of, I-15 to I-16
 - rounding of, I-34
- Single-Streaming Processors (SSP), F-40 to F-41, **F-41**, F-43
- SISD (single instruction stream, single data streams) 197, K-35
- Sketchpad, K-26
- SLA (Service Level Agreements), 25–26
- sliding window protocol, E-84
- SLO (Service Level Objectives), 25–26
- Slotnick, D. L., K-35
- small computer systems interface (SCSI), 360–361, **361**, K-62 to K-63
- Smalltalk, J-30
- smart switches, E-85 to E-86, **E-86**
- Smith, A. J., K-53 to K-54
- Smith, Burton, K-26
- Smith, J. E., A-55, K-22
- SMP. *See* symmetric shared-memory multiprocessors
- SMT. *See* simultaneous multithreading
- snooping protocols, 208–218
 - cache coherence implementation
 - and, H-34
 - development of, K-39 to K-40, **K-39**
 - examples of, 211–215, **213, 214, 215, K-39**
 - implementation of, 209–211, 217–218
 - limitations of, 216–217, **216**
 - overview, 208–209, **209**
- SNR (signal-to-noise ratio), D-21, **D-21**

- SoC (system-on-chip). D-3, D-19,
D-20, E-23, E-64
- soft real-time systems. 7, D-3
- software
- optimization, 261–262, 302–305,
304, **309**
 - pipelining, D-10, G-12 to G-15,
G-13, **G-15**
 - speculation, control dependences
and, 74
- Solaris, 377–379, **378**
- Sony Playstation 2 (PS2)
- block diagram of, **D-16**
 - embedded microprocessors in,
D-14
 - Emotion Engine in, D-15 to D-18,
D-16, **D-18**
 - Graphics Synthesizer, **D-16**, D-17
to D-18
 - vector instructions in, F-47
- source routing, E-48
- SPARC
- addressing modes in, J-5 to J-6,
J-5
 - architecture overview, **J-4**
 - common extensions in, J-19 to
J-24, **J-21** to **J-23**
 - conditional branch options in,
B-19
 - exceptions and, A-56
 - extended precision in, I-33
 - features added to, **J-44**
 - instructions unique to, J-29 to
J-32, **J-31**
 - MIPS core subset in, J-6 to J-16,
J-7, **J-9** to **J-13**, **J-17**
 - multiply-step instruction in, I-12
 - register windows in, J-29 to J-30
- SPARC VIS, **D-11**, J-16 to J-19, **J-18**
- SPARCLE processor, K-26
- sparse array accesses, G-6
- sparse matrices, in vector mode, F-26
to F-29
- spatial locality, 38, 288, C-2, C-25
- SPEC (Standard Performance
Evaluation Corporation)
- evolution of, 29–32, **31**, K-7
 - Perfect Club and, F-51
 - reproducibility of, 33
 - SPEBWeb, 32, 249
 - SPEC CPU2000, 31, **35**
 - SPEC CPU2006, 30, **31**
 - SPEC89, 30
 - SPEC92, **157**
 - SPEC2000, 331–335, **332**, **333**,
334
 - SPECfp, E-87
 - SPEChpc96, F-51
 - SPECint, E-87
 - SPECMail, 376
 - SPEC-optimized processors, E-85
 - SPECrate, 32
 - SPECRatio, 34–37, **35**
 - SPECSFS, 32, 376
 - Web site for, 30
- special-purpose register computers.
B-3
- spectral methods, computing for, H-7
- speculation. *See also* hardware-based
speculation
- compiler, G-28 to G-32
 - development of, K-22
 - dynamic scheduling in, 104
 - memory latency hiding by,
247–248
 - misspeculation rates in the
Pentium 4, 134–136, **135**
 - multiple instructions and,
118–121, **120**, **121**
 - optimizing amount of, 129
 - register renaming vs. reorder
buffers, 127–128
 - software, 74
 - through multiple branches, 129
 - value prediction and, 130
- speculative code scheduling, K-23
- speculative execution, 325
- speed of light, E-11
- speedups
- Amdahl's law and, 39–41,
202–203
 - in buffer organizations, E-58 to
E-60
 - cost-effectiveness and, 259, **260**
 - execution time and, 257–258
 - linear, 259–260, **260**
 - as performance measure in
parallel processors, H-33 to
H-34
 - from pipelining, A-3, A-10 to
A-13
 - relative vs. true, 258
 - scaled, 258–259, H-33 to H-34
 - from SMT, 177–178, **178**
- superlinear, 258
 - switch microarchitecture and,
E-62
- spin locks
- coherence in implementation of,
240–242, **242**
 - with exponential back-off, H-17
to H-18, **H-17**
- spin waiting, 241, **242**
- SPRAM (scratch pad memory), D-17,
D-18
- spread spectrum, D-25
- square root computations, I-14, I-31,
I-64, J-27
- squared coefficient of variance, 383
- SRAM (static RAM), 311, F-16
- SRC-6 system, F-50
- SRT division, I-45 to I-47, **I-46**, I-55
to I-58, **I-57**
- SSE (Streaming SIMD Extension),
B-31
- SSE/SSE2, J-46
- stack architecture
- extended, J-45
 - high-level languages and, B-45
 - historical perspectives on, B-45,
K-9 to K-10
 - in Intel 80x86, J-52
 - operands in, B-3 to B-5, **B-4**
- stalls. *See also* dependences; pipeline
stalls
- bubbles, A-13, A-20, E-47, E-53
 - control, 74
 - data hazard, A-19 to A-20, **A-20**,
A-21, A-59, **A-59**
 - forwarding and, A-17 to A-18,
A-18
 - reservation stations and, 93, **94**,
95–97, **99**, **101**, 104
 - write, C-11
- standard deviation, 36
- Stanford DASH multiprocessor, K-41
- Stanford MIPS computer, K-12 to
K-13, K-21
- start-up time, in vector processors,
F-11 to F-12, **F-13**, F-14,
F-20, F-36
- starvation, E-49
- state transition diagrams, 234–236,
235, **236**
- static branch prediction, 80–81, **81**,
D-4

- static scheduling, A-66
 - steady state, **379**, **380**
 - sticky bits, I-18, **I-19**
 - Stop & Go flow control, E-10
 - storage area networks, E-3, E-102 to E-103
 - storage systems, 357–404. *See also*
 - disk storage; I/O
 - asynchronous I/O, 391
 - block servers vs. filers, 390–391
 - dependability benchmarks, 377–379, **378**
 - disk arrays, 362–366, **363**, **365**
 - disk storage improvements, 358–361, **359**, **360**, **361**
 - faults and failures in, 366–371, **369**, **370**
 - filers, 391, 397–398
 - flash memory, 359–360
 - Internet Archive, 392–397, **394**
 - I/O performance, 371–379, **372** to **376**, **378**
 - point-to-point links and switches in, 390, **390**
 - queuing theory, 379–382, **379**, **381**
 - sector-track cylinder model, 360–361
 - Tandem disks, 368–369, **370**
 - Tertiary Disk project, 368, **369**, 399, **399**
 - throughput vs. response time, 372–374, **373**, **374**
 - transaction-processing benchmarks, 374–375, **375**
 - StorageTek 9840, K-59
 - store buffers, 94–95, **94**, 97, **101**, 102–104
 - store conditional instruction, 239–240
 - store-and-forward switching, E-50, E-79
 - streaming buffers, K-54
 - Streaming SIMD Extension (SSE), B-31
 - Strecker, W. D., C-56, J-65, J-81, K-11, K-12, K-14, K-52
 - Stretch (IBM 7030), K-18
 - stride, F-21 to F-23
 - strided addressing, B-31
 - strip mining, F-17 to F-18, **F-17**, F-39
 - striping, 362–364, **363**
 - strong typing, G-10
 - structural hazards, A-11, A-13 to A-15, A-70. *See also* pipeline hazards
 - subset property, 248
 - subtraction, I-22 to I-23, I-45
 - subword parallelism, J-17
 - Sun Java Workstation W1100z, 46–47, **46**
 - Sun Microsystems, fault detection in, 51–52
 - Sun Microsystems SPARC. *See* SPARC
 - Sun Microsystems UNIX, C-36 to C-37
 - Sun Niagara processor, 300, **341**
 - Sun T1
 - directories in, 208, 231
 - multicore, 198, 205
 - multithreading in, 250–252, **251**
 - organization of, 249, **250**, **251**
 - overall performance of, 253–257, **253** to **257**
 - Sun Ultra 5, **35**
 - Sun UltraSPARC, **E-73**
 - Super Advanced IC, **D-20**
 - superblocks, G-21 to G-23, **G-22**
 - supercomputers, 7
 - SuperH
 - addressing modes in, J-5 to J-6, **J-6**
 - architecture overview, **J-4**
 - common extensions in, J-19 to J-24, **J-23**, **J-24**
 - conditional branch options in, **B-19**
 - instructions unique to, J-38 to J-39
 - MIPS core subset in, J-6 to J-16, **J-8**, **J-9**, **J-14** to **J-17**
 - multiply-accumulate in, J-19, **J-20**
 - reduced code size in, B-23 to B-24
 - superlinear speedup, 258
 - "superpages," C-57
 - superpipelining, A-57
 - superscalar (multiple-issue) processors
 - characteristics of, **115**
 - development of, **16**, K-21 to K-22, K-25 to K-26
 - in embedded systems, D-8
 - goals of, 114
 - ideal, 155–156, **157**
 - increasing instruction fetch
 - bandwidth in, 121–127, **122**, **124**, **126**
 - issue slots in, 174–175, **174**
 - limitations of, 181–183
 - SMT performance comparison on, 179–181, **179**, **180**, **181**
 - speculation in, 118–121, **120**, **121**
 - types of, 114
 - vectorization of, F-46 to F-47
- supervisor process, 316
 - Sur, S., **E-77**
 - Sutherland, Ivan, K-26
 - SVlex, **F-7**
 - Swartzlander, E., I-63
 - switch degree, E-38
 - switch microarchitecture, E-55, E-60
 - switch statements, register indirect jumps for, B-18
 - switched point-to-point networks, E-5
 - switched-media networks, E-21, E-24, E-25
 - switches
 - context, 316
 - input-buffered, E-57, E-59, E-62, E-73
 - input-output-buffered, E-57, **E-57**, E-60, **E-61**, E-62
 - microarchitecture, E-55 to E-58, **E-56**, **E-57**, E-62
 - output-buffered, E-57, E-59
 - pipelining, E-60 to E-61, **E-61**
 - point-to-point, 390, **390**
 - process, 316, C-48
 - smart, E-85 to E-86, **E-86**
 - switching
 - buffered wormhole, E-51
 - circuit, E-50, E-64
 - cut-through, E-50, E-60, E-74
 - defined, E-22
 - network performance and, E-52
 - packet, E-50, E-77
 - pipelined circuit, E-50, E-71
 - in shared-media networks, E-23
 - store-and-forward, E-50, E-79
 - in switched-media networks, E-24
 - technique of, E-50, E-52

- virtual cut-through, E-51, E-73, E-92
 - wormhole, E-51, E-58, E-92
 - syllables, G-35
 - symmetric shared-memory
 - multiprocessors (SMPs), 205–218
 - architecture of, 200, **200**
 - cache coherence protocols in, 205–208, **206**
 - coherence in, 205–208
 - commercial workload
 - performance in, 220–224, **221 to 226**
 - in large-scale multiprocessors, **H-45**
 - limitations of, 216–217, **216**
 - scientific application performance on, H-21 to H-26, **H-23 to H-26**
 - shared vs. private data in, 205
 - snooping protocol example, 211–215, **213, 214, 215**
 - snooping protocol implementation in, 208–211, 217–218
 - symmetry, E-33, E-38
 - Synapse N + 1, K-39, **K-39**
 - synchronization, 237–242
 - barrier, H-13 to H-16, **H-14, H-15, H-16**
 - development of, K-40, K-44
 - hardware primitives, 238–240, H-18 to H-21, **H-21**
 - implementing locks using coherence, 240–242, **242**
 - memory consistency and, 244–245
 - performance challenges in
 - large-scale multiprocessors, H-12 to H-16, **H-14, H-15, H-16**
 - sense-reversing barriers, H-14 to H-15, **H-15**
 - serialization in, H-16
 - software implementations, H-17 to H-18, **H-17**
 - synchronous DRAM (SDRAM), 313–314, 338, **338**
 - synchronous events, A-40, A-41, **A-42**
 - synchronous I/O, 391
 - synonyms, 329, C-36
 - synthetic benchmarks, 29
 - system area networks (SAN), E-3, **E-4**, E-72 to E-77, **E-75 to E-77**, E-100 to E-102. *See also* interconnection networks
 - system calls, 316
 - system-on-chip (SoC), D-3, D-19, **D-20**, E-23, E-64
- T**
- tag field, C-8, **C-8**
 - tags
 - function of, 289, C-8
 - in Opteron data cache, C-12 to C-13
 - process-identifier, C-36, **C-37**
 - in snooping protocols, 210–211
 - in SPARC architecture, J-30, **J-31**
 - tail duplication, G-21
 - tailgating, F-39 to F-40
 - Takagi, N., I-65
 - Tandem disks, 368–369, **370**
 - Tanenbaum, A. S., K-11
 - TB-80 cluster, **394**, 396–397
 - TCP/IP, E-81, E-83, **E-84**, E-95
 - TDMA (time division multiple access), D-25
 - telephone company failures, 371
 - temporal locality, 38, 288, C-2
 - Tera processor, K-26
 - terminating events, A-41, **A-42**
 - Tertiary Disk project, 368, **369**, 399, **399**
 - test-and-set synchronization primitive, 239
 - TFLOPS multiprocessor, K-37 to K-38
 - Thinking Machines, K-35, K-61
 - Thinking Multiprocessor CM-5, K-40
 - Thornton, J. E., K-10
 - thread-level parallelism (TLP), 172–179. *See also* multiprocessing; multithreading
 - defined, 172
 - instruction-level parallelism vs., 172
 - in MIMD computers, 197
 - processor comparison for, 179–181, **179, 180, 181**
 - processor limitations in, 181–183
 - reasons for rise of, 262–264
 - simultaneous multithreading in, 173–179, **174, 178**
 - threads, 172, 199
 - three-hop miss, **H-31**
 - three-phased arbitration, E-49
 - throttling, E-10, E-53
 - throughput. *See also* bandwidth; effective bandwidth
 - congestion management and, E-54, E-65
 - defined, 15, 28, 372, E-13
 - deterministic vs. adaptive routing and, E-53 to E-54, **E-54**
 - I/O, 371
 - Thumb. *See* ARM Thumb
 - Thunder Tiger4, **E-20, E-44, E-56**
 - TI 320C6x, D-8 to D-10, **D-9, D-10**
 - TI 8847 chip, I-58, **I-58, I-59**, I-61
 - TI ASC, F-44, F-47
 - TI TMS320C55, D-6 to D-8, **D-6, D-7**
 - time division multiple access (TDMA), D-25
 - time of flight, E-13, E-25
 - time per instruction, in pipelining, A-3
 - time-constrained scaling, 259, H-33 to H-34
 - time-domain filtering, D-5
 - time-sharing, C-48
 - time-to-live field, E-84
 - TLB. *See* translation lookaside buffers
 - TLP. *See* thread-level parallelism
 - Tomasulo's approach to dynamic scheduling, 92–104
 - advantages of, 98, 104
 - algorithm details, 100–101, **101**
 - basic processor structure in, 94, **94**
 - dynamic scheduling using, 92–97
 - hardware-based speculation in, 105–114
 - instruction steps in, 95
 - loop-based example, 102–104
 - multiple issue and speculation example, 118–121, **120, 121**
 - register renaming in, 127–128
 - reorder buffer in, 106–114, **107, 110, 111, 113**
 - reservation stations in, 93, **94**, 95–97, **99, 101**, 104
 - software pipelining compared to, G-12

I-36 Index

- topology, E-29 to E-44
 - in centralized switched networks, E-30 to E-34, **E-31**, **E-33**
 - defined, E-21
 - in distributed switched networks, E-34 to E-39, **E-36**, **E-37**, **E-40**
 - network performance and, E-40 to E-44, **E-44**, E-52
 - torus
 - in IBM Blue Gene/L, E-53 to E-55, **E-54**, E-63, E-72 to E-74
 - overview of, E-36 to E-38
 - performance and cost of, **E-40**
 - total ordering in, E-47
 - total store ordering, 245
 - tournament predictors, 86–89, **160**, 161, **162**, K-20
 - toy programs, 29
 - TP (transaction-processing)
 - benchmarks, 32–33, 374–375, **375**
 - TPC (Transaction Processing Council), 32, 374–375, **375**
 - TPC-A, 32
 - TPC-App, 32
 - TPC-C, 32, 46–47
 - TPC-H, 32
 - TPC-W, 32
 - trace caches, 131, **132**, **133**, 296, **309**
 - trace compaction, G-19
 - trace scheduling, G-19 to G-21, **G-20**
 - trace selection, G-19
 - traffic intensity, 381
 - Transaction Processing Council (TPC), 32, 374–375, **375**
 - transaction time, 372
 - transaction-processing benchmarks, 32–33, 374–375, **375**
 - transaction-processing servers, 46–48, **47**, **48**
 - transactions, steps in, 372, **373**
 - transcendental functions, I-34, **J-54**
 - transfers, instructions as, B-16
 - transient failures, E-66
 - transient faults, 367, 378–379
 - transistors, performance scaling in, 17–19
 - translation buffers (TB). *See* translation lookaside buffers
 - translation lookaside buffers (TLB)
 - in AMD Opteron, 326–327, **327**, **328**, C-55, C-55
 - cache hierarchy and, 291, **292**
 - development of, K-52
 - in MIPS 64, K-52
 - misses and, C-45
 - speculation and, 129
 - virtual memory and, 317, 320, **323**, C-36, C-43 to C-45, C-45
 - Transmission Control Protocol, E-81
 - transmission speed, E-13
 - transmission time, E-13 to E-14
 - transport latency, E-14
 - trap handlers, I-34 to I-35, I-36, J-30
 - trap instructions, A-42
 - tree height reduction, G-11
 - tree-based barriers, H-18, **H-19**
 - trees
 - binary tree multipliers, I-53 to I-54
 - combining, H-18
 - fat, **E-33**, E-34, E-36, E-38, **E-40**, E-48
 - multiply, I-52 to I-53, **I-53**
 - octrees, H-9
 - signed-digit, I-53 to I-54
 - tree height reduction, G-11
 - tree-based barriers, H-18, **H-19**
 - Wallace, I-53 to I-54, **I-53**, I-63
 - Trellis codes, D-7
 - trigonometric functions, I-31 to I-32
 - TRIPS Edge processor, E-63
 - Trojan horses, C-49, C-52
 - true sharing misses, 218–219, 222, **224**, **225**
 - true speedup, 258
 - tunnel diode memory, K-53
 - Turing, Alan, K-4
 - Turn Model, E-47
 - two-level predictors, 85
 - two-phased arbitration, E-49
 - two's complement system, I-7 to I-10
 - two-way conflict misses, C-25
 - two-way set associative blocks, C-7, C-8
 - TX-2, K-26
 - type fields, E-84
- ## U
- Ultracomputer, K-40
 - Ultra-SPARC desktop computers, K-42
 - Ultrix operating system, **C-37**, **E-69**
 - UMA (uniform memory access), 200, **200**, **216**, 217
 - unbiased exponents, I-15
 - uncertainty, code, D-4
 - underflow, I-15, I-36 to I-37, I-62
 - unicasting, E-24
 - Unicode, B-14, B-34
 - unified caches, C-14, **C-15**
 - uniform memory access (UMA), 200, **200**, **216**, 217. *See also* symmetric shared-memory multiprocessors
 - unit stride addressing, B-31
 - UNIVAC I, K-5
 - unpacked numbers, I-16
 - unpacking operations, B-14
 - up*/down* routing, E-48, E-67
 - upgrade misses, H-35
 - upgrade requests, 219
 - urgent pointer fields, E-84
 - use bits, C-43 to C-44
 - user maskable events, A-41, **A-42**
 - user miss rates, 228, **228**, **229**
 - user nonmaskable events, A-41, **A-42**
 - user productivity, transaction time and, 372–374, **373**, **374**
 - user-level communication, E-8
 - user-requested events, A-40 to A-41, **A-42**
- ## V
- valid bits, C-8
 - value prediction, 130, 154–155, 170, K-25
 - variable-length encoding, 10, B-22 to B-23, **B-22**
 - variables, register types and, B-5
 - variance, 383
 - VAX, J-65 to J-83
 - addressing modes in, J-67, J-70 to J-71

- architecture summary, **J-42**, J-66 to J-68, **J-66**
- CALLS instruction in, B-41 to B-43
- code size overemphasis in, B-45
- condition codes in, J-71
- conditional branch options in, **B-19**
- data types in, **J-66**
- encoding instructions in, J-68 to J-70, **J-69**
- exceptions in, **A-40**, A-45 to A-46
- frequency of instruction distribution, J-82, **J-82**
- goals of, J-65 to J-66
- high-level language architecture in, K-11
- historical floating point formats in, I-63
- memory addressing in, B-8, B-10
- operand specifiers in, J-67 to J-68, **J-68**
- operations, J-70 to J-72, **J-71**, **J-73**
- pipelining microinstruction execution in, A-46
- sort procedure in, J-76 to J-79, **J-76**, **J-80**
- swap procedure in, J-72 to J-76, **J-74**, **J-75**, J-79
- VAX 11/780, 2, 3, K-6 to K-7, K-11
- VAX 8600, A-76
- VAX 8700
 - architecture of, K-13 to K-14, **K-14**
 - MIPS M2000 compared with, J-81, **J-82**
 - pipelining cost-performance in, A-76
- VAX 8800, A-46
- vector architectures
 - advantages of, B-31 to B-32, F-47
 - compiler effectiveness in, F-32 to F-34, **F-33**, **F-34**
 - in Cray X1, F-40 to F-41, **F-41**
 - in embedded systems, D-10
- vector instructions, 68
- vector length
 - average, F-37
 - control, F-16 to F-21, **F-17**, **F-19**, F-35
 - optimization, F-37
- registers, F-16 to F-18
- vector loops. execution time of, F-35. *See also* loop unrolling
- vector processors, F-1 to F-51
 - advantages of, F-2 to F-4
 - basic architecture of, F-4 to F-6, **F-5**, **F-7**, **F-8**
 - chaining in, F-23 to F-25, **F-24**, **F-35**
 - characteristics of various, **F-7**
 - conditionally executed statements in, F-25 to F-26
 - Cray X1, F-40 to F-44, **F-41**, **F-42**
 - Earth Simulator, F-3 to F-4
 - historical perspectives on, F-47 to F-51
 - instructions in, **F-8**
 - load-store units in, F-6, **F-7**, F-13 to F-14
 - memory systems in, F-14 to F-16, **F-15**, F-22 to F-23, F-45
 - multiple lanes in, F-29 to F-31, **F-29**, **F-30**
 - multi-streaming, F-43
 - operation example, F-8 to F-10
 - peak performance in, F-36, F-40
 - performance measures in, F-34 to F-35, **F-35**
 - pipelined instruction start-up in, F-31 to F-32, **F-31**
 - scalar performance and, F-44 to F-45, **F-45**
 - sparse matrices in, F-26 to F-29
 - sustained performance in, F-37 to F-38
 - vector execution time, F-10 to F-13, **F-13**
 - vector stride in, F-21 to F-23
 - vector-length control, F-16 to F-21, **F-17**, **F-19**, F-35
- vector-mask control, F-25 to F-26, F-28
- vector-mask registers, F-26
- vector-register processors
 - characteristics of various, **F-7**
 - components of, F-4 to F-6, **F-5**, **F-7**, **F-8**
 - defined, F-4
 - vector-length control in, F-16 to F-21, **F-17**, **F-19**
- VelociTI 320C6x processors, D-8 to D-10, **D-9**, **D-10**
- versions, E-84
- very long instruction word processors. *See* VLIW processors
- victim blocks, 301, 330
- victim buffers, 301, 330, C-14
- victim caches, 301, K-54
- virtual addresses, C-36, **C-54**
- virtual caches, C-36 to C-38, **C-37**
- virtual channels
 - head-of-line blocking and, E-59, **E-59**
 - in IBM Blue Gene/L ED Torus, E-73
 - in InfiniBand, E-74
 - performance and, E-93
 - routing and, E-47, E-53 to E-55, **E-54**
 - switching and, E-51, E-58, E-61, E-73
- virtual cut-through switching, E-51, E-73, E-92
- virtual functions, register indirect jumps for, B-18
- Virtual Machine Control State (VMCS), 340
- virtual machine monitors (VMMs)
 - instruction set architectures and, 319–320, 338–340, **340**
 - Intel 80x86, 320, 321, 339, **340**
 - overview of, 315, 318
 - page tables in, 320–321
 - requirements of, 318–319
 - Xen VMM example, 321–324, **322**, **323**
- virtual machines (VM), 317–324
 - defined, 317
 - impact on virtual memory and I/O, 320–321
 - instruction set architectures and, 319–320
 - overview of, 317–318
 - Xen VMM example, 321–324, **322**, **323**
- virtual memory, C-38 to C-55
 - in 64-bit Opteron, C-53 to C-55, **C-54**, **C-55**
 - address translations in, C-40, C-44 to C-47, **C-45**, **C-47**
 - block replacement in, C-43 to C-44
 - caches compared with, C-40, **C-41**

- virtual memory (*continued*)
 - defined, C-3
 - development of, K-53
 - function of, C-39
 - in IBM 370, J-84
 - impact of virtual machines on, 320–321
 - in Intel Pentium, C-48, C-49 to C-52, **C-51**
 - mapping to physical memory, C-39, **C-40**
 - in memory hierarchy, C-40, **C-41**, C-42 to C-44, **C-43**
 - miss penalties in, C-40, C-42
 - in Opteron, C-53 to C-55, **C-54**, **C-55**
 - page sizes in, C-45 to C-46
 - paged vs. segmented, C-40 to C-42, **C-41**, **C-42**
 - protection and, 315–317, 324–325, C-39
 - relocation in, C-39, **C-40**
 - size of, C-40
 - translation lookaside buffers and, 317, 320, **323**, C-36, C-43 to C-45, **C-45**
 - virtual output queues (VOQ), E-60, E-66
 - virtually indexed, physically tagged optimization, 291–292, C-38, C-46
 - VLIW Multiflow compiler, **297**
 - VLIW processors, 114–118. *See also* Intel IA-64
 - characteristics of, 114–115, **115**
 - in embedded systems, D-8 to D-10, **D-9**, **D-10**
 - EPIC approach in, G-33
 - historical perspectives on, K-21
 - overview of, 115–118, **117**
 - VLVCU (load vector count and update), F-18
 - VM. *See* virtual machines
 - VME racks, 393, **394**
 - VMIPS
 - architecture of, F-4 to F-6, **F-5**, **F-7**, **F-8**
 - memory pipelines on, F-38 to F-40
 - multiple lanes in, F-29 to F-31, **F-29**, **F-30**
 - operation example, F-8 to F-10
 - peak performance in, F-36
 - processor characteristics in, **F-7**
 - sustained performance in, F-37 to F-38
 - vector length control in, F-19 to F-20
 - vector stride in, F-22
 - VMM. *See* virtual machine monitors
 - voltage, adjustable, 18
 - von Neumann, J., 287, I-62, K-2 to K-3
 - von Neumann computers, K-3
 - VPU processors, D-17 to D-18
 - VS registers, F-6
 - VT-x, 339–340
- W**
- wafer yield, 23–24
 - wafers, costs of, 21–22, **23**
 - waiting line, 380. *See also* queuing theory
 - Wall, D. W., 154, 169–170, K-25
 - Wallace trees, I-53 to I-54. **I-53**, I-63
 - wall-clock time, 28
 - WAN (wide area networks), E-4, **E-4**, **E-75**, E-79, E-97 to E-99. *See also* interconnection networks
 - Wang, W.-H., K-54
 - WAR (write after read) hazards
 - hardware-based speculation and, 112
 - as ILP limitations, 72, 169
 - in pipelines, 90
 - in scoreboarding, A-67, A-69 to A-70, A-72, **A-75**
 - Tomasulo's approach and, 92, 98
 - wavelength division multiplexing (WDM), E-98
 - WAW (write after write) hazards
 - in floating-point pipelines, A-50, A-52 to A-53
 - hardware-based speculation and, 112
 - as ILP limitations, 71, 169
 - in pipelines, 90
 - in scoreboarding, A-67, A-69, A-75 to A-76
 - Tomasulo's approach and, 92, 98–99
 - way prediction, 295, **309**
 - Wayback Machine, 393
 - WB. *See* write-back cycles
 - WCET (worst case execution time), D-4
 - weak ordering, 246, K-44
 - Web server benchmarks, 32–33, 377
 - Web sites
 - availability of, 400
 - on multiple-issue processor development, K-21
 - for OpenMP consortium, H-5
 - for SPEC benchmarks, 30
 - for Transaction Processing Council, 32
 - weighted arithmetic mean time, 383
 - Weitek 3364 chip, I-58, **I-58**, **I-60**, I-61
 - West, N., I-65
 - Whetstone synthetic program, K-6
 - Whirlwind project, K-4
 - wide area networks (WAN), E-4, **E-4**, **E-75**, E-79, E-97 to E-99. *See also* interconnection networks
 - Wilkes, Maurice, 310, B-1, K-3, K-52, K-53
 - Williams, T. E., I-52
 - Wilson, R. P., 170
 - Winchester disk design, K-60
 - window (instructions)
 - effects of limited size of, 158–159, **159**, 166–167, **166**
 - defined, 158
 - limitations on size of, 158
 - in scoreboarding, A-74
 - in TCP, **E-84**
 - windowing, E-65
 - wireless networks, D-21 to D-22, **D-21**
 - within vs. between instructions, A-41, **A-42**
 - Wolfe, M., F-51
 - word count field, **C-51**, C-52
 - word operands, B-13
 - working set effect, H-24
 - workloads, execution time of, 29
 - World Wide Web, 6, E-98
 - wormhole switching, E-51, E-58, E-88, E-92 to E-93
 - worst case execution time (WCET), D-4
 - write allocate, C-11 to C-12
 - write back, in virtual memory, C-44
 - write buffers

- defined, C-11
 - function of, 289, 291
 - merging, 300–301, **301**, **309**
 - read misses and, 291, C-34 to C-35
 - in snooping protocols, 210
 - write invalidate protocols
 - in directory-based cache coherence protocols, **233**, 234
 - example of, 212, **213**, **214**
 - implementation of, 209–211
 - overview, 208–209, **209**
 - write merging, 300–301, **301**, **309**
 - write miss
 - directory protocols and, 231, **233**, 234–237, **235**, **236**
 - in large-scale multiprocessors, H-35, H-39 to H-40
 - sequential consistency and, 244
 - in snooping protocols, 212–214, **213**, **214**
 - in spinning, 241, **242**
 - write allocate vs. no-write allocate, C-11 to C-12
 - write result stage of pipeline, 96, 100–101, **103**, 108, 112
 - write serialization, 206–207
 - write speed, C-9 to C-10
 - write stalls, C-11
 - write update (broadcast) protocol, 209, 217
 - write-back caches
 - advantages and disadvantages of, C-10 to C-12
 - cache coherence and, H-36
 - consistency in, 289
 - defined, C-10
 - directory protocols and, 235, **236**, 237
 - invalidate protocols and, 210, 211–212, **213**, **214**
 - in Opteron microprocessor, C-14
 - reducing cost of writes in, C-35
 - write-back cycles (WB)
 - in floating-point pipelining, **A-51**, A-52
 - in RISC instruction set, A-6
 - in unpipelined MIPS implementation, A-28, **A-29**
 - writes, to disks, 364
 - write-through caches
 - advantages and disadvantages of, C-11 to C-12
 - defined, C-10
 - invalidate protocols and, 210, 211, 212
 - I/O coherency and, 326
 - write buffers and, C-35
 - Wu, Chuan-Lin, E-1
- ## X
- X1 nodes, F-42, **F-42**
 - Xen VMM, 321–324, **322**, **323**
 - Xeon-MP, 198
 - XIE, **F-7**
 - XIMD architecture, K-27
 - Xon/Xoff flow control, E-10
- ## Y
- Yajima, S., I-65
 - Yamamoto, W., K-27
 - Yasuura, H., I-65
 - yields, 19–20, **20**, 22–24
- ## Z
- zero
 - finding zero iteration, I-27 to I-29, **I-28**
 - in floating-point multiplication, I-21
 - shifting over, I-45 to I-47, **I-46**
 - signed, I-62
 - zero-copy protocols, E-8, E-91
 - zero-load, E-14, E-25, E-52, E-53, E-92
 - zSeries, F-49
 - Zuse, Konrad, K-4



About the CD

The CD that accompanies this book includes:

- *Reference appendices.* These appendices—some guest authored by subject experts—cover a range of topics, including specific architectures, embedded systems, and application-specific processors.
- *Historical Perspectives and References.* Appendix K includes several sections exploring the key ideas presented in each of the chapters in this text. References for further reading are also provided.
- *Search engine.* A search engine is included, making it possible to search for content in both the printed text and the CD-based appendices.

Appendices on the CD

- **Appendix D:** Embedded Systems
- **Appendix E:** Interconnection Networks
- **Appendix F:** Vector Processors
- **Appendix G:** Hardware and Software for VLIW and EPIC
- **Appendix H:** Large-Scale Multiprocessors and Scientific Applications
- **Appendix I:** Computer Arithmetic
- **Appendix J:** Survey of Instruction Set Architectures
- **Appendix K:** Historical Perspectives and References

Srinivas Institute of Technology

Acc. No.17049.....

Call No.