

Table of Contents

1. Introduction to computer architecture	9
1.1 Basic terms	9
1.2 Von Neumann architecture	10
1.3 Computer generations	11
2. Performance evaluation	13
2.1 Instruction mixes	13
2.2 Benchmarks	15
2.2.1 Natural benchmarks	15
2.2.2 Synthetic benchmarks	17
3. Memory systems	20
3.1 Memory hierarchy	20
3.2 Interleaved memory	22
3.3 Virtual memory	23
3.3.1 Paged memory system	24
3.3.2 Segmented memory	27
3.3.3 Memory with paged segments	29
3.3.4 Page fault	30
3.4 Cache	31
3.4.1 Locality of references	31
3.4.2 Primary and secondary cache	32
3.4.3 Performance of the memory system	33
3.4.4 Cache addressing	34
3.4.5 Cache consistency	35
4. Pipelining	37
4.1 Principles	37
4.1.1 Pipeline function	37
4.1.2 Optimization of a pipeline	39
4.1.3 Superpipeline processors	41
4.2 Instruction-level pipeline	41
4.2.1 Instruction prefetch	42
4.2.2 Fetch-decode-execute cycle	43

4.2.4. Interrupt handling	45
4.3 Pipeline hazards	45
4.3.1 Structural hazards	46
4.3.2 Data hazards	47
4.3.3 Control hazards	49
4.4 Branch prediction	50
4.4.1 Static branch prediction	50
4.4.2 Delayed branch	50
4.4.3 Dynamic branch prediction	51
4.5 Pipelined arithmetic operations	53
4.5.1 Fixed-point adder	53
4.5.2 Floating-point adder	54
5. Instruction-level parallelism	56
5.1 Superscalar concepts	56
5.2 Fundamental limitations	57
5.2.1 Structural hazards	58
5.2.2 Data hazards	58
5.2.3 Control hazards	59
5.2.4 Types of parallelism	60
5.3 Instruction issue methods	60
5.3.1 In-order issue and in-order completion	61
5.3.2 In-order issue and out-of-order completion	62
5.3.3 Out-of-order issue and out-of-order completion	63
5.4 Register dataflow	64
5.4.1. Software solution	64
5.4.2. Hardware solution	65
5.4.3 Dispatch stack	66
5.5 Storage location reuse	66
5.5.1 Register renaming	67
5.5.2 Partial renaming	68
5.5.3 Memory issues	69
5.6 Instruction fetching and decoding	69
6. Microprocessor architecture	71

6.1 CISC processors	71
6.2 Development of microprocessor technology	71
6.3 Intel 80486	72
6.3.1 Architecture	72
6.3.2 Segmentation	74
6.3.3 Paging	75
6.3.4 On-chip cache	76
6.3.5 Caching translation lookaside buffer entries	77
6.4 Pentium	78
6.4.1 Architecture	78
6.4.2 Separate code and data caches	79
6.4.3 Branch prediction	80
6.4.4 Enhanced 64-bit data bus	80
6.5 Pentium Pro	81
6.5.1 Improvements over the Pentium architecture	81
6.5.2 Processor architecture	81
6.5.3 Speculative execution	83
7. RISC Processors	84
7.1 Evolution of RISC technology	84
7.2 Basic properties of RISC architecture	85
7.3 Example RISC architectures	86
7.4 POWER architecture	88
7.4.1 Main features of POWER architecture	88
7.4.2 PowerPC	90
7.5 UltraSPARC-I	92
7.5.1 Pipeline organization	93
7.5.2 Execution units	94
7.5.3 Memory management unit	95
7.5.4 Cache architecture	96
7.5.5 Graphics unit	96
7.5.6 Operating system support	97
7.6 R10000	97
7.6.1 Execution Units	97
7.6.2 Cache Architecture and TLB	98

7.6.3 Instruction Flow	99
7.6.4 Register Renaming	100
7.6.5 Multi-processor support	101
7.7 PA-8000	101
7.7.1 Processor architecture	102
7.7.2 Off chip primary cache	103
7.7.3 Branch prediction	104
7.7.4 Out-of-order execution	105
7.7.5 Two memory accesses per cycle	105
7.7.6 Multimedia acceleration extensions	106
8 Parallel systems	108
8.1 Basic concepts of parallelism	108
8.1.1 Levels of parallelism	109
8.1.2 Process granularity	110
8.1.3 Extent of parallelism	112
8.2 Flynn's taxonomy	112
8.2.1 SISD systems	113
8.2.2 SIMD systems	113
8.2.3 MISD systems	113
8.2.4 MIMD systems	114
8.2.5 Extension to basic taxonomy	114
8.3 General classification of parallel systems	115
8.4 Amdahl's law	116
8.5 Application areas of parallel systems	118
8.6 Parallel programming	119
8.6.1 Manual parallelization	120
8.6.2. Automatic parallelization	122
8.7 Multithreading	123
9. Interconnection networks	124
9.1 Network function and taxonomy	124
9.2 Some properties of graphs	126
9.3 Static networks	128
9.3.1 Hypercube	129
9.3.2 Lattices	130

9.3.3 Fat tree	131
9.4 Dynamic networks	132
9.4.1 Crossbar switch	132
9.5 Multistage switching networks	134
9.5.1 Building blocks	134
9.5.2 Blocking networks.....	134
9.5.3 Rearrangeable networks	136
9.5.4 Nonblocking networks	136
9.6 Buses	137
9.7 Scalable coherence interface	138
10 SISD parallel systems and vector processors	139
10.1 Very-long instruction word processors	139
10.2 Redundant systems	141
10.2.1 Duplex system	142
10.2.2 TMR system	142
10.2.3 Bi-duplex system	143
10.3 Vector processors	144
10.3.1 Principles of vector processing	144
10.3.2 Vector coprocessors	147
10.3.3 Vector supercomputers	149
11. MIMD systems	152
11.1 Memory access models	152
11.2 Shared memory MIMD systems	153
11.3 Distributed memory MIMD systems	156
11.4 Example architectures of shared memory systems	159
11.4.1 Switch connected systems	159
11.4.2 Bus connected systems	159
11.4.3 Switched memory connected systems	162
11.4.4 NUMA systems	163
11.4.5 COMA systems	165
11.5 Example architectures of distributed memory systems	166
11.5.1 Hypercube connected systems	166
11.5.2 Systems with 2-D interconnections	168
11.5.3 Switch connected systems	172

12. SIMD systems	176
12.1 Principles of SIMD machine operation	176
12.2 SIMD systems with local memory	179
12.2.1 Illiac IV	179
12.2.2 Connection machines	180
12.2.3 MasPar MP-1	182
12.3 SIMD Systems with shared memory	183
12.3.1 BSP system	183
12.3.2 Prime memory system	186
12.4 Associative processors	187
12.4.1 Associative memory organization	187
12.4.2 Associative search algorithms	189
12.4.3 STARAN computer	190
13 Non von Neumann systems	193
13.1 Data flow systems	193
13.1.1 Data flow versus control flow	193
13.1.2 Dedicated DF computers	195
13.1.3 Static DF computers	195
13.1.4 Dynamic DF computers	198
13.1.5 Perspectives of DF computing	199
13.2 Systolic arrays	199
13.2.1 Linear systolic arrays	200
13.2.2 Systolic array for matrix multiplication	203
References	205