

Contents

1	Low Power Design: An Introduction	1
1.1	The Emergence of Power as an Important Design Metric	1
1.2	Power Efficiency vs. Energy Efficiency	3
1.3	Power-Performance Tradeoff	6
1.4	Power Density	7
1.5	Power and Energy Benchmarks	7
1.6	Power Optimizations at the System Level	7
1.7	Organization of this Book	9
	References	10
2	Basic Low Power Digital Design	11
2.1	CMOS Transistor Power Consumption	11
2.1.1	Switching Power	13
2.1.2	Short Circuit Power	14
2.1.3	Leakage Power	15
2.2	Trends in Power Consumption	18
2.3	Techniques for Reducing Dynamic Power	18
2.3.1	Gate Sizing	20
2.3.2	Control Synthesis	23
2.3.3	Clock Gating	25
2.3.4	Voltage and Frequency Scaling	28
2.4	Techniques for Reducing Short Circuit Power	30
2.5	Techniques for Reducing Leakage Power	31
2.5.1	Multiple Supply Voltage	31
2.5.2	Multiple Threshold Voltage	33
2.5.3	Adaptive Body Biasing	34
2.5.4	Transistor Stacking	34
2.5.5	Power Gating	36
2.6	Summary	37
	References	38

3	Power-efficient Processor Architecture	41
3.1	Introduction	41
3.1.1	Power Budget: A Major Design Constraint	45
3.1.2	Processor Datapath Architecture	48
3.1.3	Power Dissipation	52
3.2	Front-end: Fetch and Decode Logic	54
3.2.1	Fetch Gating	54
3.2.2	Auxiliary Decode Buffer	58
3.3	Issue Queue / Dispatch Buffer	60
3.3.1	Dynamic Adaptation of Issue Queue Size	63
3.3.2	Zero Byte Encoding	64
3.3.3	Banking and Bit-line Segmentation	65
3.3.4	Fast Comparators	66
3.4	Register File	66
3.4.1	Port Reduction and Banking	67
3.4.2	Clustered Organization	70
3.4.3	Hierarchical Organization	70
3.5	Execution Units	72
3.5.1	Clock Gating	73
3.5.2	Operand Isolation/Selective Evaluation	73
3.5.3	Power Gating and Multi-threshold Logic	74
3.6	Reorder Buffer	77
3.6.1	Port Reduction	79
3.6.2	Distributed ROB	80
3.6.3	Dynamic ROB Sizing	81
3.6.4	Zero Bytes and Power Efficient Comparators	81
3.7	Branch Prediction Unit	81
3.7.1	Banking of BHT and BTB	84
3.7.2	Reducing BHT/BTB Lookups	84
3.8	Summary	85
	References	86
4	Power-efficient Memory and Cache	89
4.1	Introduction and Memory Structure	90
4.1.1	Overview	90
4.1.2	Memory Structure	91
4.1.3	Cache Memory	92
4.1.4	Cache Architecture	98
4.1.5	Power Dissipation During Memory Access	100
4.2	Power-efficient Memory Architectures	101
4.2.1	Partitioned Memory and Caches	101
4.2.2	Augmenting with Additional Memories	103
4.2.3	Reducing Tag and Data Array Fetches	105
4.2.4	Reducing Cache Leakage Power	109

4.3	Translation Look-aside Buffer (TLB)	111
4.3.1	TLB Associativity – A Power-performance Trade-off	113
4.3.2	Banking	113
4.3.3	Reducing TLB Lookups	115
4.4	Scratch Pad Memory	116
4.4.1	Data Placement in SPM	117
4.4.2	Dynamic Management of SPM	119
4.4.3	Storing both Instructions and Data in SPM	121
4.5	Memory Banking	121
4.6	Memory Customization	124
4.7	Reducing Address Bus Switching	129
4.7.1	Encoding	129
4.7.2	Data Layout	131
4.8	DRAM Power Optimization	134
4.9	Summary	135
	References	136
5	Power Aware Operating Systems, Compilers, and Application Software	139
5.1	Operating System Optimizations	140
5.1.1	Advanced Configuration and Power Interface (ACPI)	144
5.1.2	Dynamic Voltage and Frequency Scaling	148
5.1.3	I/O Device Power Management	160
5.2	Compiler Optimizations	161
5.2.1	Loop Transformations	162
5.2.2	Instruction Encoding	162
5.2.3	Instruction Scheduling	164
5.2.4	Dual Instruction Set Architectures	164
5.2.5	Instruction Set Extension	168
5.2.6	Power Gating	171
5.2.7	Dynamic Translation and Recompilation	172
5.2.8	Compiler Optimizations Targeting Disks	173
5.3	Application Software	174
5.3.1	Application-aided Power Management	174
5.3.2	DVFS Under Application Control	175
5.3.3	Output Quality Trade-offs	177
5.4	Summary	178
	References	178
6	Power Issues in Servers and Data Centers	183
6.1	Power Efficiency Challenges	183
6.1.1	Nameplate Power Overestimates Actual Power	184
6.1.2	Installed vs. Utilized Capacity	185
6.1.3	Load Variation	185
6.2	Where does the Power go?	186
6.3	Server Power Modeling and Measurement	188

6.4	Server Power Management	190
6.4.1	Frequency Scaling	190
6.4.2	Processor and Memory Packing	193
6.4.3	Power Shifting	196
6.5	Cluster and Data Center Power Management	197
6.5.1	Power Capping/Thresholding	198
6.5.2	Voltage and Frequency Scaling	201
6.6	Summary	204
	References	204
7	Low Power Graphics Processors	207
7.1	Introduction to Graphics Processing	208
7.1.1	Graphics Pipeline	208
7.1.2	Graphics Processor Architecture	216
7.1.3	Power Dissipation in a Graphics Processor	221
7.2	Programmable Units	222
7.2.1	Clock Gating	223
7.2.2	Predictive Shutdown	223
7.2.3	Code Transformation	224
7.3	Texture Unit	228
7.3.1	Custom Memory Architecture – Texture Filter Memory	229
7.3.2	Texture Compression	232
7.3.3	Clock Gating	234
7.4	Raster Operations	235
7.4.1	Depth Buffer Compression	235
7.4.2	Color Buffer Compression	237
7.5	System Level Power Management	238
7.5.1	Power Modes	238
7.5.2	Dynamic Voltage and Frequency Scaling	238
7.5.3	Multiple Power Domains	245
7.6	Summary	245
	References	246
	Index	249