



Jon Peddie

The History of the GPU - New Developments

 Springer

The History of the GPU - New Developments

Jon Peddie

The History of the GPU - New Developments

 Springer

Jon Peddie
Jon Peddie Research
Tiburon, CA, USA

ISBN 978-3-031-14046-4 ISBN 978-3-031-14047-1 (eBook)
<https://doi.org/10.1007/978-3-031-14047-1>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Real-time 3D graphics and consumer gaming markets have been responsible for driving tremendous innovations to feed the insatiable appetite of high-resolution, photo-realistic gaming technologies. Capturing the interest of computer scientists and creative hardware developers around the world, the development of the GPU has led to advancements in the computational capabilities and memory systems to feed them. Advanced algorithms and APIs to manage large, complex data systems—along with the move to general-purpose programming models with exploitation for general-purpose computing, high-performance computing, cryptocurrency, and artificial intelligence—have further propelled the GPU into an unprecedented pace of development.

In the early 1990s when I first became involved with the commercialization of 3D graphics technology, Jon Peddie was already a well-known graphics market analyst. I joined a team of very seasoned hardware and software developers at GE Visual Systems in Daytona Beach, where large-scale military and NASA training systems were developed. We created some of the first consumer commercial uses of the technology with Sega Models 1 and 2 hardware, initially sporting 180 k polygons per second and 1.2 M pixels per second with a resolution of 496×384 in the arcade gaming space. After acquisition by Martin Marietta and then Lockheed Martin, real 3D was formed where I was part of a small team that developed Intel 740 3D architecture that started Intel's 3D rendering roadmaps. Jon has shared some unique perspective on I740 development and Intel's entry into 3D graphics that a quick search will reveal. His second book of this series will cover industry trends and struggles during this period. I joined ATI Technologies in 1999, later acquired by Advanced Micro Devices, Inc. (AMD) where I have had the pleasure of advancing the Radeon product line, console gaming systems, and our latest RDNA/CDNA products that power some of the most exciting developments of the century. Over the years, I have regularly read JPR research report by Jon to understand his broad prospective of relevant emerging trends in our industry. I have had the pleasure of meeting with Jon on several occasions at product introduction events and industry conferences to chat about trends, motivations, technical detail, and the successes in real-time graphics.

In Jon's third book of a three-book series on the History of the GPU, he shares an interesting and knowledgeable history of the chaotic and competitive time that forged today's leaders in the 3D gaming hardware industry. Jon draws on the breath of his relationships formed over the years and his knowledge to break these contributions into six eras of GPU development. In each chapter, Jon not only covers innovations and key products, but also shares his perspective on company strategy, key leaders, and visionary architects during each era of development. I hope that you will thoroughly enjoy this series and the final book while learning about the tremendous growth of technology and the hard work, risk, and determination of those who have contributed to today's GPU success.

Michael Mantor
AMD Chief GPU Architect
and Corporate Fellow

Preface

This is the third book in the three-book series on the History of the GPU.

The first book covered the history of computer graphics controllers and processors from the 1970s leading up to the introduction of the fully integrated GPU first appearing in game consoles in 1996, and then the PC in 1999. The second book in the series covers the developments that led up to the integrated GPU, from the early 1990s to the late 1990s.

The GPU has been employed in many systems (platforms) and evolved since 1996.

This final book in the series covers the second to sixth eras of the development of GPU on the PC platform, and other platforms. Other platforms include workstations, game machines, and others, such as various vehicles—GPUs are used everywhere in almost everything.

Each chapter is designed to be read independently, hence there may be some redundancy. Hopefully, each one tells an interesting story.

In general, a company is discussed and introduced on the year of its formation. However, a company may be discussed in multiple time periods in multiple chapters depending on how significant their developments were and what impact they had on the industry.

History of the GPU		
Steps to Invention Book 1	Eras and Environment Book 2	New Developments Book 3
1. Preface	1. Preface	1. Preface
2. History of the GPU	2. Race to build the first GPU	2. Second Era of GPUs (2001-2006)
3. 1980-1990 Graphics Controllers on Other Platforms	3. GPU Functions	3. Third to Fifth Era of GPUs
4. 1980-1989 Graphics Controllers on PCs	4. Major Era of GPUs	4. Mobile GPUs
5. 1990-1995 Graphics Controllers on PCs	5. First Era of GPUs	5. Game Console GPUs
6. 1990-1999 Graphics Controllers on Other Platforms	6. GPU Environment-Hardware	6. Compute GPUs
7. 1996-1999 Graphics Controller on PCs	7. Application Program Interface (API)	7. Open GPUs
8. What is a GPU	8. GPU Environment-Software Extensions	8. Sixth Era of GPUs

The History of the GPU - New Developments

I mark the GPU’s introduction as the first fully integrated single chip with hardware geometry processing capabilities—transform and lighting. Nvidia gets that honor on the PC by introducing their GeForce 256 based on the NV10 chip in October 1999. However, Silicon Graphics Inc. (SGI) introduced an integrated GPU in the Nintendo 64 in 1996, and ArtX developed an integrated GPU for the PC a month after Nvidia. As you will learn, Nvidia did not introduce the concept of a GPU, nor did they

develop the first hardware implementation of transform and lighting. But Nvidia was the first to bring all that together in a mass-produced single chip device.

The evolution of the GPU however did not stop with the inclusion of the transformation and lighting (T&L) engine because the first era of such GPUs had fixed function T&L processors—that was all they could do and when they were not doing that they sat idle using power. The GPU kept evolving and has gone through six eras of evolution ending up today as a universal computing machine capable of almost anything.

The Author

A Lifetime of Chasing Pixels

I have been working in computer graphics since the early 1960s, first as an engineer, then as an entrepreneur (I founded four companies and ran three others), ending up in a failed attempt at retiring in 1982 as an industry consultant and advisor. Over the years, I watched, advised, counseled, and reported on developing companies and their technology. I saw the number of companies designing or building graphics controllers swell from a few to over forty-five. In addition, there have been over thirty companies designing or making graphics controllers for mobile devices.

I've written and contributed to several other books on computer graphics (seven under my name and six co-authored). I've lectured at several universities around the world, written uncountable articles, and acquired a few patents, all with a single, passionate thread—computer graphics and the creation of beautiful pictures that tell a story. This book is liberally sprinkled with images—block diagrams of the chips, photos of the chips, the boards they were put on, and the systems they were put in, and pictures of some of the people who invented and created these marvelous devices that impact and enhance our daily lives—many of them I am proud to say are good friends of mine.

I laid out the book in such a way (I hope) that you can open it up to any page and start to get the story. You can read it linearly; if you do, you'll probably find new information and probably more than you ever wanted to know. My email address is in various parts of this book, and I try to answer everyone, hopefully with 48 hours. I'd love to hear comments, your stories, and your suggestions.

The following is an alphabetical list of all the people (at least I hope it's all of them) who helped me with this project. A couple of them have passed away, sorry to say. Hopefully, this book will help keep the memory of them and their contributions alive.

Thanks for reading

Jon Peddie—*Chasing pixels, and finding gems*

Acknowledgments and Contributors

The following people helped me with editing, interviews, data, photos, and most of all encouragement. I literally and figuratively could not have done this without them.

Ashraf Eassa—Nvidia
Andrew Wolfe—S3
Anand Patel—Arm
Atif Zafar—Pixilica
Borger Ljosland—Fal anx
Brian Kelleher—DEC, and finally Nvidia
Bryan Del Rizzo—3dfx & Nvidia
Carrell Killebrew—TI/ATI/AMD
Chris Malachowsky—Nvidia
Curtis Priem—Nvidia
Dado Banatao—S3
Dan Vivoli—Nvidia
Dan Wood—Matrox, Intel
Daniel Taranovsky—ATI
Dave Erskine—ATI & AMD
Dave Orton—SGI, ArtX, ATI & AMD
David Harold—Imagination Technologies
Dave Kasik—Boeing
Emily Drake—Siggraph
Edvaed Sergard—Fal anx
Eric Demers—AMD/Qualcomm
Frank Paniagua—Video Logic
Gary Tarolli—3dfx
Gerry Stanley—Real3D
George Sidiropoulos—Think Silicon
Henry Chow—Yamaha & Giga Pixel
Henry Fuchs—UNC
Henry C. Lin—Nvidia
Henry Quan—ATI
Hossain Yassaie—Imagination Technologies
Iakovos Istamoulis—Think Silicon
Ian Hutchinson—Arm
Jay Eisenlohr—Rendition
Jay Torberg—Microsoft
Jeff Bush—Nyuzi
Jeff Fischer—Weitek & Nvidia
Jem Davis—Arm
Jensen Huang—Nvidia
Jim Pappas—Intel
Joe Curley—Tseng/Intel

Jonah Alben—Nvidia
John Poulton—UNC & Nvidia
Karl Gutttag—TI
Karthikeyan (Karu) Sankaralingam—University of Wisconsin-Madison
Kathleen Maher—JPA & JPR
Ken Potashner—S3 & SonicBlue

Tiburon, USA

Jon Peddie

Contents

1 Introduction	1
1.1 Programmable Vertex and Geometry Shaders (2001–2006)	1
1.1.1 Nvidia NV20—GeForce 3 (February 2001)	2
1.1.2 ATI R200 Radeon 8500 (August 2001)	4
1.1.3 Nvidia’s NV25–28—GeForce 4 Ti (February 2002)	11
1.1.4 ATI’s R300 Radeon 9700 and the VPU (August 2002)	13
1.1.4.1 First PC GPU with Eight Pipes	16
1.1.4.2 Z-Buffer	17
1.1.4.3 Video	18
1.1.4.4 Memory Management	19
1.1.4.5 Multiple Displays	20
1.1.4.6 Along Comes a RenderMonkey	22
1.1.4.7 Summary	22
1.1.5 SiS Xabre—September 2002	23
1.1.5.1 SiS 301B Video Processor	26
1.1.5.2 Summary	26
1.1.6 The PC GPU Landscape in 2003	27
1.1.7 Nvidia NV 30–38 GeForce FX 5 Series (2003–2004) ...	27
1.1.7.1 CineFX	28
1.1.7.2 Nvidia Enters the AIB Market with the GeForceFX (2003)	31
1.1.8 ATI R520 an Advanced GPU (October 2005)	31
1.1.8.1 Avivo Video Engine	44
1.1.8.2 Summary	45
1.1.8.3 Nvidia’s NV40 GPU (2005–2008)	45
1.2 Conclusion	49
References	49

- 2 The Third- to Fifth-Era GPUs** 51
 - 2.1 The Third Era of GPUs—DirectX 10 (2006–2009) 51
 - 2.1.1 Nvidia G80 First Unified Shader GPU (2006) 53
 - 2.1.2 Nvidia GT200 Moving to Compute (2008) 54
 - 2.1.2.1 Summary 55
 - 2.1.3 Intel Larrabee to Phi (2006–2009) 57
 - 2.1.4 Intel’s GM45 iGPU Chipset (2007–2008) 62
 - 2.1.5 Intel’s Westmere (2010) Its First iGPU 62
 - 2.2 The Fourth Era of GPUs. October 2009 65
 - 2.2.1 The End of the ATI Brand (2010) 66
 - 2.2.2 AMD’s Turks GPU (2011) 66
 - 2.2.2.1 Summary 68
 - 2.2.3 Nvidia’s Fermi (2010) 69
 - 2.2.3.1 Summary 70
 - 2.2.4 AMD Fusion GPU with CPU (January 2011) 70
 - 2.2.4.1 Summary 76
 - 2.2.5 Nvidia Kepler (May 2013) 76
 - 2.2.6 Intel’s iGPUs (2012–2021), the Lead Up to dGPU 79
 - 2.2.7 Nvidia Maxwell (2014) 82
 - 2.3 The Fifth Era of GPUs (July 2015) 83
 - 2.3.1 AMD’s CGN RX380 (June 2016) 83
 - 2.3.2 Intel’s Kaby Lake G (August 2016) 85
 - 2.3.3 Nvidia 86
 - 2.3.4 AMD’s Navi RDNA Architecture (July 2019) 87
 - 2.3.4.1 Radeon RX 5700 XT AIB (July 2019) 88
 - 2.3.4.2 Summary 91
 - 2.3.4.3 RX 5500 Series (2019) 91
 - 2.3.5 Summary 92
 - 2.3.6 Intel’s Whisky Lake 620 GT2 iGPU (2018) 92
 - 2.3.7 Intel’s Gen 11 iGPU (March 2019) 93
 - 2.3.7.1 Intel’s GPU’s Geometry Engine 94
 - 2.3.7.2 Intel Updates Its Ring Topology 95
 - 2.3.7.3 Coarse Pixel Shading 96
 - 2.3.7.4 Position Only Shading Tile-Based Rendering (POSH) 96
 - 2.3.8 Summary 98
 - 2.4 Conclusion 98
 - References 98
- 3 Mobile GPUs** 101
 - 3.1 Organization 103
 - 3.2 Mobiles: The First Decade (2000–2010) 103
 - 3.3 Imagination Technologies First GPU IP (2000) 104
 - 3.3.1 Summary 107
 - 3.4 Arm’s Path to GPUs (2001) 108

3.4.1	Falanx	110
3.4.2	Mali Family (2005)	111
3.4.3	More Cores	113
3.4.4	Balanced, Scalable, and Fragmented	115
3.4.5	More Designs	116
3.5	Fujitsu’s MB86292 GPU (2002–)	117
3.5.1	IMB86R01 Jade	118
3.5.2	Several Name Changes	119
3.6	Nvidia’s Tegra—From PDAs to Autonomous Vehicles and consoles (2003–)	120
3.6.1	Tegra is Born	126
3.6.2	Nvidia Enters the Automotive Market (2009)	128
3.7	Bitboys 3.0 (2002–2011)	130
3.7.1	End Game: Bitboys’ VG (2003)	131
3.8	Qualcomm’s Path to the Snapdragon GPU (2004–)	139
3.8.1	The Adreno GPU (2006)	141
3.9	SECOND DECADE of Mobile GPU Developments (2010 and on)	143
3.10	Siru (2011–2022)	143
3.10.1	Samsung	145
3.11	Texas Instruments OMAP (1999–2012)	146
3.12	Arm’s Midgard (2012)	148
3.12.1	Arm’s Bifrost (2018)	153
3.12.2	Arm’s Valhall (2019)	156
3.12.2.1	AR and VR	157
3.12.3	Valhall Architecture	157
3.12.3.1	ML and Display	160
3.12.3.2	Mali-D77 Display Processor (2019)	161
3.12.4	Arm Epilogue	163
3.12.5	Second Epilogue	164
3.13	Nvidia Leaves Smartphone Market, 2014	165
3.13.1	Xavier Introduced (2016)	166
3.14	Qualcomm Snapdragon 678 (2020)	167
3.15	Qualcomm Snapdragon 888 (2020)	170
3.16	Apple’s M1 GPU and SoC (2020)	172
3.16.1	Apple’s M1 Pro GPU (2021)	174
3.16.2	Apple’s M1 Ultra (2022)	177
3.16.3	Summary	182
3.17	Conclusion	182
	References	183

- 4 Game Console GPUs** 187
 - 4.1 Sony PlayStation 2 (2000) 187
 - 4.2 Microsoft Xbox (2001) 191
 - 4.2.1 Epilogue 194
 - 4.3 Sony PSP (2004) 195
 - 4.4 Xbox 360—Unified Shaders and Integration (November 2005) 197
 - 4.4.1 The Xbox 360 GPU 199
 - 4.4.2 The Many Versions of Xbox 360 200
 - 4.4.3 Updated Xbox 360—Integrated SoC (August 2010) 200
 - 4.5 Nintendo Wii (November 2006) 202
 - 4.6 Sony PlayStation 3 (2006) 203
 - 4.7 Nintendo 3DS (June 2011) 205
 - 4.8 Sony PS Vita (December 2011) 207
 - 4.9 Eighth-Generation Consoles (2012) 209
 - 4.10 Nintendo Wii U (November 2012) 211
 - 4.11 CPUs with GPUs Lead to Powerful Game Consoles (2013) 212
 - 4.12 Nvidia Shield (January 2013–2015) 212
 - 4.12.1 A Grid Peripheral? 213
 - 4.12.2 But Was It Disruptive? 214
 - 4.13 Sony PlayStation 4 (November 2013) 216
 - 4.14 Microsoft Xbox One (November 2013) 217
 - 4.15 Nvidia Shield 2 (March 2015) 219
 - 4.16 Playmaji Polymega (February 2017) 221
 - 4.17 Nintendo Switch (March 2017) 223
 - 4.18 Atari VCS (June 2017) 224
 - 4.19 Zhongshan Subor Z-Plus Almost Console (2018–2020) 226
 - 4.20 Sony PlayStation 5 (November 2020) 227
 - 4.21 Microsoft Xbox Series X and S (November 2020) 230
 - 4.22 Valve Steam Deck Handheld (July 2021) 232
 - 4.23 Qualcomm Handheld Dec (2021) 234
 - 4.24 Conclusion 235
 - References 236
- 5 Compute Accelerators and Other GPUs** 239
 - 5.1 Sun’s XVR-4000 Zulu (2002) the End of an Era 240
 - 5.2 SiliconArts Ray Tracing Chip and Intellectual Property (IP) (2019) 242
 - 5.2.1 RayCore 1000 242
 - 5.2.2 RayCore 2000 244
 - 5.2.3 RayCore Lite 244
 - 5.2.4 Road Map 245
 - 5.2.5 Summary 248
 - 5.3 Intel Xe Architecture-Discrete GPU for High-Performance Computing (HPC) (2019) 248

- 5.4 Compute GPU Zhaoxin (2020) 255
- 5.5 MetaX (2020–) 258
 - 5.5.1 MetaX Epilogue 261
- 5.6 XiangDiXian Computing Technology (2020) 262
- 5.7 Bolt Graphics (2021–) 262
- 5.8 Jingjia Micro Series GPUs (2014) 264
- 5.9 Alphamosaic to Pi via Broadcom (2000–2021) 267
- 5.10 The Other IP Providers 270
 - 5.10.1 AMD 2004 271
 - 5.10.2 Digital Media Professionals Inc. (DMP Inc.) 2002 271
 - 5.10.3 Imagination Technologies 2002 274
 - 5.10.4 Think Silicon (2007) 275
 - 5.10.5 VeriSilicon 281
- 5.11 Nvidia’s Ampere (May 2020) 287
 - 5.11.1 A Supercomputer 290
- 5.12 Imagination Technologie’s Ray Tracing IP (2021) 293
 - 5.12.1 Summary 298
- 5.13 Nvidia’s Mega Data Center GPU Hopper (2022) 298
 - 5.13.1 Summary 302
- 5.14 Conclusion 303
- References 303
- 6 Open GPU Projects (2000–2018) 305**
 - 6.1 Open Graphics Project (2000) 306
 - 6.2 Nyuzi/Nyami (2012) 308
 - 6.3 MIAOW (2015) 309
 - 6.4 GPUOpen (2015) 310
 - 6.5 SCRATCH (2017) 312
 - 6.6 Libre-GPU (2018) 313
 - 6.7 Vortex: RISC-V GPU (2019) 315
 - 6.8 RV64X (2019) 316
 - 6.9 Conclusion 320
 - References 321
- 7 The Sixth Era GPUs: Ray Tracing and Mesh Shaders 323**
 - 7.1 Miners and Taking a Breath 324
 - 7.2 Nvidia’s Turing GPU (September 2018) 326
 - 7.2.1 Ray Tracing 328
 - 7.2.2 Hybrid-Rendering: AI-Enhanced Real-Time Ray Tracing 329
 - 7.2.2.1 Variable Rate Shading 330
 - 7.2.2.2 Nvidia’s New DLSS (March 2020) 331
 - 7.2.2.3 Mesh Shaders 333
 - 7.2.3 Summary 334
 - 7.3 Intel–Xe GPU (2018) 335

- 7.3.1 Intel’s X^e Max (2020) 336
- 7.3.2 Intel’s dGPU Family (2021) 340
- 7.3.3 DG1 341
 - 7.3.3.1 Hello Arc, Goodbye DG 342
 - 7.3.3.2 Intel’s Supersampling (XeSS) 344
- 7.3.4 Summary 345
- 7.4 AMD Navi 21 RDNA 2 (October 2020) 346
 - 7.4.1 AMD Ray Tracing (October 2020) 349
 - 7.4.2 FidelityFX Super Resolution (March 2021) 351
 - 7.4.3 Summary 354
- 7.5 Innosilicon (2021) 354
 - 7.5.1 The GPU Population Continued to Expand in 2021 354
 - 7.5.2 Summary 358
- 7.6 Conclusion 359
- References 359
- 8 Concluding Remarks 361**
 - References 364
- Appendix A: Acronyms 365**
- Appendix B: Definitions 369**
- Index 405**

List of Figures

Fig. 1.1 Nvidia NV20-based GeForce 3 AIB with AGP4x buss. Courtesy tech Power Up 4

Fig. 1.2 ATI R200-based Radeon 8500 AIB. Courtesy tech Power Up 5

Fig. 1.3 ATI R200 block diagram. The chip had 60 million transistors, four-pixel shaders, two vertex shaders, two texture-mapping units, and four ROP engines 5

Fig. 1.4 Tessellation can reduce or expand the number of triangles (polygons) in a 3D model to improve realism or increase performance 6

Fig. 1.5 Normal(s) generation within a TruForm N-patch. Courtesy of ATI 7

Fig. 1.6 Generation of control points with N-patches. Courtesy of ATI 9

Fig. 1.7 Subdivision and tessellation add realism. Courtesy of ATI 10

Fig. 1.8 ATI's TruForm was a preprocessor in an expanding chain of graphics functions 10

Fig. 1.9 VisionTek Nvidia NV25-based GeForce Ti 4200 AIB. Courtesy of Hyins for Wikipedia 11

Fig. 1.10 Nvidia GeForce 4 pipeline 12

Fig. 1.11 ATI R300 Radeon 9700 AIB. Notice heatsinks on the memory and similar layout to Nvidia NV25-based GeForce Ti 4200 AIB, in Fig. 1.9. Courtesy of Wikimedia 14

Fig. 1.12 ATI R300 block diagram. The display interface included a multi-input LUTDAC 14

Fig. 1.13 ATI's R300 vertex setup engine (one of four) 15

Fig. 1.14 ATI's R300 pixel shader engine the chip had eight of these "pipes" 17

Fig. 1.15 ATI R300 video processing engine block diagram 19

Fig. 1.16 ATI R300 video processing engine showing all the outputs 20

Fig. 1.17 Xabre 600 AIB with similar layout to ATI and Nvidia. Courtesy of Zoltek 24

Fig. 1.18 SiS’s Xabre vertex shader data flow between CPU and GPU ... 25

Fig. 1.19 SiS’s competitive market position 25

Fig. 1.20 Nvidia’s NV30-based GeForce Fx 5900 with heat sink and fan removed. Courtesy of iXBT 28

Fig. 1.21 Nvidia NV30 block diagram 29

Fig. 1.22 *Final Fantasy* used subdivision rendering for skin tone. Courtesy of Nvidia [14] 30

Fig. 1.23 ATI R520 ring bus memory controller. The GDDR is connected at the four ring stops. (Source ATI) 34

Fig. 1.24 ATI R520 block diagram 35

Fig. 1.25 ATI R520 thread size and dynamic branching efficiency was improved with ultra-threading. Courtesy of ATI 37

Fig. 1.26 ATI R520 vertex shader engine 38

Fig. 1.27 Making things look brighter than they are. Courtesy of ATI 39

Fig. 1.28 Inside the abandoned church with HDR on. Courtesy of Valve 40

Fig. 1.29 Inside the abandoned church with HDR off. Courtesy of Valve 40

Fig. 1.30 Different modes of anti-aliasing. Courtesy of Valve 41

Fig. 1.31 ATI’s special class of bump mapping 42

Fig. 1.32 ATI’s Ruby red CrossFire—limited production. Courtesy of ATI 43

Fig. 1.33 Nvidia NV40 Curie vertex and fragment processor block diagram 45

Fig. 1.34 Nvidia’s NV40 curie-based GeForce 6800 Xt AIB. Courtesy tech Power Up 47

Fig. 1.35 Nvidia curie block diagram 48

Fig. 2.1 Tony Tamasi. Courtesy of Nvidia 52

Fig. 2.2 GPU architecture progression, first and second era. Courtesy of Tony Tamasi 52

Fig. 2.3 Evolution from first-era to third-era GPU design 53

Fig. 2.4 Nvidia’s G80 unified shader GPU—a sea of processors 54

Fig. 2.5 Nvidia GeForce 8800 Ultra with the heatsink removed showing the 12 memory chips surrounding the GPU. Courtesy of Hyins—Public Domain, Wikimedia 55

Fig. 2.6 Nvidia’s GT200 streaming multiprocessor 56

Fig. 2.7 Evolution of Nvidia’s logo, 1993 to 2006 (left) and 2006 on (right). Courtesy of Nvidia 57

Fig. 2.8 Daniel Pohl demonstrating Quake running ray-traced in real time 58

Fig. 2.9 Intel Larrabee AIB. Courtesy of the VGA Museum 58

Fig. 2.10 General organization of the Larrabee many-core architecture 59

Fig. 2.11 Larrabee’s simplified DirectX 10 pipeline. The gray components were programmable by the user, and blue were fixed. Omitted from the diagram are memory access, stream output, and texture-filtering stages 59

Fig. 2.12 Larrabee CPU core and associated system blocks. The CPU was a Pentium processor in-order design, plus 64-bit instructions, multi-threading, and a wide vector processor unit (VPU) 59

Fig. 2.13 Intel’s G45 chipset 63

Fig. 2.14 Block diagram of an iGPU within a CPU 64

Fig. 2.15 Intel’s Westmere dual-chip package. Courtesy of Intel 64

Fig. 2.16 Intel’s Ironlake-integrated HD GPU 65

Fig. 2.17 AMD graphics logos, circa 1985, 2006, 2010. Courtesy of AMD 66

Fig. 2.18 AMD’s Turks entry-level GPU (2011) 69

Fig. 2.19 Portion of the Llano chip. Courtesy of AMD 73

Fig. 2.20 Comparison of GPU balance philosophy of semiconductor suppliers 73

Fig. 2.21 AMD’s APU road map. Courtesy of AMD 74

Fig. 2.22 AMD’s integrated Llano CPU–GPU 75

Fig. 2.23 Nvidia’s GeForce GTX 780. Courtesy of Wikipedia GBPublic_PR 77

Fig. 2.24 Nvidia demo of a crumbling building. Courtesy of Nvidia 77

Fig. 2.25 Intel’s Gen 11 Tiger Lake CPU with iGPU 80

Fig. 2.26 Intel’s SuperFin transistor. Courtesy of Intel 80

Fig. 2.27 Die shot of Intel’s 11th Gen Core processor showing the amount of die used by the GPU. Courtesy of Intel 81

Fig. 2.28 Raja Koduri, Intel’s Chief Architect and Senior Vice President. Courtesy of Intel 82

Fig. 2.29 Nvidia Maxwell GPU running voxel illumination. Courtesy of Nvidia 83

Fig. 2.30 AMD’s CGN CU block diagram 85

Fig. 2.31 AMD revealed their GPU roadmap. Courtesy of AMD 85

Fig. 2.32 Intel multi-chip Kaby Lake G. The chip on the left is the 4 GB HMB2, the middle chip is the Radeon RX Vega, and the chip on the right is the eighth-gen core. Courtesy of Intel 86

Fig. 2.33 Nvidia’s GPU roadmap. Courtesy of Nvidia 87

Fig. 2.34 Block diagram of the AMD Navi 10, one of the first GPUs powered by the RDNA architecture 89

Fig. 2.35 AMD’s RDNA command processor and scan converter 90

Fig. 2.36 AMD’s RDNA compute unit front-end and SIMD 90

Fig. 2.37 Intel GT2 iGPU block diagram 93

Fig. 2.38 Intel Gen 11 iGPU block diagram 95

Fig. 2.39 CPS added two more steps in the GPU’s pipeline 96

Fig. 2.40 Geometry with red boxes is sufficiently far from the camera, and therefore, it is of minor importance to the overall image. Thus, the color shading frequency could be reduced (using CPS with no noticeable effect on the visual quality or the frame rate). Courtesy of Intel 97

Fig. 2.41 Position only tile-based rendering (PTBR) block diagram 97

Fig. 3.1 The rise and fall of mobile graphics chip and intellectual property (IP) suppliers versus market growth 102

Fig. 3.2 Mobile devices 102

Fig. 3.3 Sources of mobile GPUs 103

Fig. 3.4 Big jump in GPU power efficiency. Courtesy of Imagination Technologies 104

Fig. 3.5 Tile region protection isolates critical functions from each other. Courtesy of Imagination Technologies 105

Fig. 3.6 Imagination’s BXT MC4 block diagram. Courtesy of Imagination Technologies 106

Fig. 3.7 The B boxes of imagination. Courtesy of Imagination Technologies 106

Fig. 3.8 In 2020 imagination had a broadest range of IP GPU designs available. Courtesy of Imagination Technologies 107

Fig. 3.9 Mali in Arm, circa 2006 109

Fig. 3.10 Falanx Arm Mali block diagram 112

Fig. 3.11 Arm Mali’s graphics stack with MIDlets 114

Fig. 3.12 The Mali-400 could share the load on fragments 116

Fig. 3.13 Fujitsu MB86292 GPU 117

Fig. 3.14 Fujitsu’s MB86R01 SoC Jade 119

Fig. 3.15 MediaQ MQ-200 block diagram 121

Fig. 3.16 MediaQ MQ-200 drawing engine 122

Fig. 3.17 Symbolic block diagram of the Nvidia TEGRA 6x0 (2007) 126

Fig. 3.18 Nvidia’s Tegra road map (2011) 127

Fig. 3.19 Nvidia offered its X-Jet software development toolkit (SDK) software stack for automotive development on the Jetson platform 129

Fig. 3.20 Mercedes concept car of the future. Courtesy of Nvidia 130

Fig. 3.21 In the back row from left: Petri Norlaund, Kaj Tuomi, and Mika Tuomi from Bitboys. In the front row, Falanx, from left: unknown (guy in blue jeans), Mario Blazevic, Jørn Nystad, Edvard Sjørgård, and Borgar Ljosland. Courtesy of Borgar Ljosland 132

Fig. 3.22 Bitboys’ Acceleon handheld prototype and the art it is rendering. Courtesy of Petri Nordlund 133

Fig. 3.23 Bitboys’ G40 mobile GPU organization 134

Fig. 3.24 Mikko Sarri 2009. Courtesy Mikko Sarri 138

Fig. 3.25 Ideal’s Joe Palooka punching bag. *Source* thepeoplehistory.com 139

Fig. 3.26 Qualcomm’s SMS6550 SoC 140

Fig. 3.27 Qualcomm’s Snapdragon SoC with Adreno GPU 142

Fig. 3.28 Mikko Alho. Courtesy of Siru 143

Fig. 3.29 The many lives of the Bitboys 144

Fig. 3.30 Texas Instruments early OMAP SoCs 147

Fig. 3.31 Arm Mali-T658 organization 149

Fig. 3.32 Mali-T658 program management 151

Fig. 3.33 Pipelines in Arm’s Mali architecture 151

Fig. 3.34 Arm Midgard block diagram 152

Fig. 3.35 Jem Davis VP Arm. Courtesy of Arm 2020 152

Fig. 3.36 The 12-year history of Arm Mali architectures over time 154

Fig. 3.37 Arm’s Mali-G76’s core design block diagram 154

Fig. 3.38 Improved gaming performance with Mali-G76. Courtesy
of Arm 155

Fig. 3.39 Arm’s comparison of performance between the Mali-G52
and the new Mali-G57. Courtesy of Arm 156

Fig. 3.40 Arm’s Valhall shader architecture block diagram 158

Fig. 3.41 Arm Valhall microarchitecture 158

Fig. 3.42 Arm’s Mali Valhall architecture. Courtesy of Arm 159

Fig. 3.43 Arm has the whole suite of engines for 5G AI, ML,
and VR. Courtesy of Arm 160

Fig. 3.44 Arm’s D77 display processor block diagram 161

Fig. 3.45 Arm said an SoC that could drive the level of performance
for wearable VR HMDs did not exist (in 2019). That
presented a significant challenge to SoC vendors who need
to achieve the above requirements. Courtesy of Arm 162

Fig. 3.46 Nvidia’s route to and from the mobile market 165

Fig. 3.47 Nvidia’s high-level (circa 2017) Xavier block
diagram—DLA is the deep learning accelerator 167

Fig. 3.48 Nvidia’s Xavier-based Pegasus board (circa 2018) offered
320 TOPS and the ability to run deep neural networks
at the same time. Courtesy of Nvidia 168

Fig. 3.49 Nvidia’s Tegra SoC roadmap 2022. Courtesy of Nvidia 168

Fig. 3.50 Qualcomm Snapdragon 6xx 170

Fig. 3.51 Block diagram of Apple’s M1 PC SoC 172

Fig. 3.52 Floor plan of Apple’s M1 substrate with chip and memory.
Courtesy of Apple 173

Fig. 3.53 Apple’s M-Series SoCs. Courtesy of Apple 174

Fig. 3.54 The CPUs of the M1 Pro. Courtesy of Apple 175

Fig. 3.55 The M1 Max offers 4x faster GPU performance than M1.
Courtesy of Apple 176

Fig. 3.56 The M1 Mx with its unified embedded memory. Courtesy
of Apple 176

Fig. 3.57 Performance comparison. Courtesy of Apple 177

Fig. 3.58	Apple's UltraFusion packaging architecture connects two M1 Max die to create the M1 Ultra. Courtesy of Apple	178
Fig. 3.59	Apple said the 20-core CPU of the M1 Ultra could deliver 90% higher multi-threaded performance than the fastest 2022 16-core PC desktop chip in the same power envelope. Courtesy of Apple	178
Fig. 3.60	Apple claimed its M1 Ultra 64-core GPU produced faster performance than the highest-end PC GPU available while using 200 fewer watts of power. Courtesy of Apple	179
Fig. 3.61	Apple M1 compared to AMD Ryzen chip size. Courtesy of Max Tech/YouTube [49]	181
Fig. 4.1	Rise and fall of console supplier versus market growth	188
Fig. 4.2	Number of consoles offered per year over time	188
Fig. 4.3	Sony PlayStation 2 block diagram	189
Fig. 4.4	Original Xbox team Ted Hase, Nat Brown, Otto Berkes, Kevin Bachus, and Seamus Blackley. Courtesy of Microsoft	191
Fig. 4.5	Xbox block diagram with Nvidia IGP	193
Fig. 4.6	<i>Halo</i> , developed by Bungie, was an exclusive Xbox title and credited with the machine's success. Courtesy of Microsoft	193
Fig. 4.7	Sony PSP block diagram	195
Fig. 4.8	Ken Kutaragi at E3 2003 telling the audience about the PSP	196
Fig. 4.9	Microsoft Xbox 360 block diagram	197
Fig. 4.10	Microsoft Xbox 360 GPU block diagram	198
Fig. 4.11	Microsoft's Xbox 360 Vejle SoC block diagram	201
Fig. 4.12	Microsoft Xbox 360 SoC chip floor plan. Courtesy of Microsoft	202
Fig. 4.13	Nintendo Wii Hollywood chip	203
Fig. 4.14	IBM technologist Dr. Lisa Su holds the new Cell microprocessor. The processor was jointly developed by IBM, Sony, and Toshiba. IBM claimed the Cell provided vastly improved graphics and visualization capabilities, in many cases 10 times the performance of PC processors. Courtesy of Business Wire	205
Fig. 4.15	Nintendo 3DS handheld game machine. Courtesy of Nintendo	206
Fig. 4.16	DMP PICO GPU	207
Fig. 4.17	Sony PlayStation Vita. Courtesy of Sony	208
Fig. 4.18	Imagination technologies' SGX543 IP GPU	209
Fig. 4.19	CPUs plus caches take up approximately 15% of the chip area. The GPUs (center) take up about 33% of the 348 mm ² die area; the rest of the chip area was the memory. Courtesy of Wikipedia	210
Fig. 4.20	Block diagram of AMD Liverpool (PS4) and Durango (Xbox One) APU	210

Fig. 4.21	Tiny but mighty, AMD's Jaguar-based APU powered the most popular eighth-generation game consoles. Courtesy of AMD	211
Fig. 4.22	Nvidia's Shield game controller/player. Courtesy of Nvidia	213
Fig. 4.23	Nvidia's grid. Courtesy of Nvidia	214
Fig. 4.24	An Nvidia Shield look-alike, the MOGA Pro controller with smartphone holder. Courtesy of MOGA	215
Fig. 4.25	Sony's eighth-generation PlayStation 4 with controller changed the design rules for consoles. Courtesy of Sony Computer Entertainment	216
Fig. 4.26	Xbox One system architecture	218
Fig. 4.27	Internals of Xbox One's 5+ billion transistor SoC	218
Fig. 4.28	Nvidia's Shield Console in its holder with controller. Courtesy of Nvidia	220
Fig. 4.29	Nvidia Tegra X1 block diagram	221
Fig. 4.30	Artist's rendition of the Polymega system. The final version was a dark, flat gray. Courtesy of Polymega	222
Fig. 4.31	Nintendo's Switch with controls attached. Courtesy of Nintendo	223
Fig. 4.32	Nintendo Switch desk mount. Courtesy of Nintendo	223
Fig. 4.33	Nintendo console introduction timeline	224
Fig. 4.34	Feargal Mac (left) of Atari and former Microsoft games executive Ed Fries. Courtesy of Dean Takahashi	225
Fig. 4.35	Atari 2600 and VCS. Courtesy of Wikipedia	225
Fig. 4.36	Xiaobawang Zhongshan Subor Z-plus console. Courtesy of Xiaobawang	227
Fig. 4.37	Sony PlayStation 5 block diagram	228
Fig. 4.38	PlayStation introduction timeline	228
Fig. 4.39	Microsoft Xbox Series X block diagram	230
Fig. 4.40	Microsoft's Xbox series APU. Courtesy of Microsoft	231
Fig. 4.41	Microsoft Xbox series introduction timeline	231
Fig. 4.42	Valve's Steam Deck game console. Courtesy of Valve	234
Fig. 4.43	Qualcomm handheld, game console, reference design	235
Fig. 5.1	The GPU scales faster than any other processor	240
Fig. 5.2	Sun microsystem's XVR-4000 graphics subsystem. Courtesy of forms.irixnet.org	241
Fig. 5.3	SiliconArt's RayCore 1000 block diagram	243
Fig. 5.4	Autodesk 3DS Max 2019 test with two omni-directional lights and 12,268 triangles. Courtesy of SiliconArts	244
Fig. 5.5	SiliconArt's RayCore 2000 block diagram	245
Fig. 5.6	SiliconArt's RayTree structure	247
Fig. 5.7	SiliconArt's RayTree architecture	247
Fig. 5.8	Features of intel's ponte vecchio GPU. Courtesy of Intel	249
Fig. 5.9	Aurora exploited a lot of intel technology. Courtesy of Intel	249

Fig. 5.10 The X^e architecture is scaled by ganging together tiles of primary GPU cores. Courtesy of Intel 249

Fig. 5.11 Intel’s X^e-HPC 2-stack shows the configurable and scalable aspects of its X^e-core design. Courtesy of Intel 250

Fig. 5.12 The X^e link allowed even more extensive subsystems to be created. Courtesy of Intel 251

Fig. 5.13 Ponte Vecchio with >100 billion transistors. 47 active tiles and five process nodes. Courtesy of Intel 252

Fig. 5.14 Ponte Vecchio chips in a carrier from the fab. Courtesy of Stephen Shankland/CNET 253

Fig. 5.15 Intel’s Ponte Vecchio circuit board revealing the tiles in the package. Courtesy of Intel 253

Fig. 5.16 Intel’s accelerated compute system. Courtesy of Intel 254

Fig. 5.17 Intel’s oneAPI allowed heterogeneous processors to communicate and cooperate 254

Fig. 5.18 Zhaoxin’s road map showed a dGPU. Courtesy of CNTechPost 257

Fig. 5.19 Texas CPU to new GPU—a long, tortuous path 258

Fig. 5.20 GlenFly’s AIB running the Unigine Heaven benchmark. Courtesy of Glenfield GlenFly Technology 259

Fig. 5.21 Muxi’s CEO Chen Weiliang has worked in the GPU field for 20 years. Courtesy of Muxi 260

Fig. 5.22 Darwesh Singh, CEO and founder bolt graphics. Courtesy of Singh 263

Fig. 5.23 Bolt’s HPC GPU. Courtesy of Bolt 263

Fig. 5.24 Bolt targeted industries characterized by exponentially expanding workloads. Courtesy of Bolt Graphics 264

Fig. 5.25 Jing Jiawei, founder of Changsha Jingjia Microelectronics Co., Ltd. Courtesy of Changsha Jingjia 265

Fig. 5.26 Jingjia micro’s JM7200-based PCIe AIB. Courtesy of Changsha Jingjia Microelectronics Co. 266

Fig. 5.27 Alphamosaic’s Dr. Robert Swann shows off the VC02’s development board 268

Fig. 5.28 Raspberry Pi 4 model B development board. Courtesy of Miiicihiaieil Hieinizilieir for Wikimedia Commons 269

Fig. 5.29 Doom III running on a Raspberry Pi 4. Courtesy of Hexus 270

Fig. 5.30 Rendering examples using only OpenVG features. Courtesy of DMP Inc. 273

Fig. 5.31 DP CEO Tatsuo Yamamoto and his dog Momo at E3 274

Fig. 5.32 Think silicon founders George Sidropoulos and Iakovos Stamoulis. Courtesy of Think Silicon 275

Fig. 5.33 Think silicon’s whiteboard from 2015. Courtesy of Think Silicon 276

Fig. 5.34 Think silicon’s Nema pico GPU 277

Fig. 5.35 The think silicon team. Courtesy of Think Silicon 278

Fig. 5.36 Comparison of think silicon’s Nema and Neox GPUs. Courtesy of Think Silicon 279

Fig. 5.37 System diagram using the think silicon IP blocks 279

Fig. 5.38 An example of an SoC with Neox IP cores 280

Fig. 5.39 Think silicon’s application and device range. Courtesy of Think Silicon 282

Fig. 5.40 Vivante’s Vega IP GPU 283

Fig. 5.41 VeriSilicon’s GPU could scale from IoT and wearables to AI training systems. Courtesy of VeriSilicon 286

Fig. 5.42 Nvidia GPU growth in transistors and die size over time 288

Fig. 5.43 The Nvidia GA100 streaming multiprocessor 291

Fig. 5.44 Nvidia’s A100 Ampere chip on a circuit board 292

Fig. 5.45 Nvidia’s DGX A100 supercomputer. Courtesy of Nvidia 293

Fig. 5.46 Conventional ray tracing organization 294

Fig. 5.47 Imagination technologie’s Photon RAC 295

Fig. 5.48 Imagination technologie’s GPU with RAC 296

Fig. 5.49 Rays per second monitor in PVRTune. Courtesy of Imagination Technologies 297

Fig. 5.50 Imagination created the industry’s first real-time ray tracing silicon in 2014. It showed the R6500 test chip code named Plato. Courtesy of Imagination Technologies 297

Fig. 5.51 Nvidia’s hopper subsystem board. Courtesy of Nvidia 299

Fig. 5.52 Time to train the mixture of experts transformer network for H100 versus A100. Courtesy of Nvidia 300

Fig. 5.53 Nvidia’s H100 Hopper AIB with NVLinks (upper left) supports a unified cluster of eight GPU. Courtesy of Nvidia 301

Fig. 5.54 Nvidia’s DGX H100 supercomputer. Courtesy of Nvidia 301

Fig. 5.55 Nvidia’s Earth 2 supercomputer. Courtesy of Nvidia 302

Fig. 6.1 Timothy Miller. Courtesy of University Binghamton 307

Fig. 6.2 OGP test board. Courtesy of en. wikipedia 308

Fig. 6.3 Karu Sankaralingam. Courtesy of University of Wisconsin–Madison 309

Fig. 6.4 MIAOW block diagram 310

Fig. 6.5 AMD RDNA generation series block diagram 311

Fig. 6.6 Two different trimmed architectures were generated for two distinct soft kernels. Courtesy of Pedro Duarte and Gabriel Falcao from the Universities of Coimbra and Lisboa 312

Fig. 6.7 During compile-time, the instructions present in kernel A indicated that only scalar and vectorized integer FUs should be instantiated on the reconfigurable fabric. Courtesy of Pedro Duarte and Gabriel Falcao from the Universities of Coimbra and de Lisboa 313

Fig. 6.8 Luke Kenneth Casson Leighton. Courtesy of Leighton 314

Fig. 6.9 The libre-SOC hybrid 3D CPU-VPU-GPU 314

Fig. 6.10 Vortex block diagram 316

Fig. 6.11 Dr. Atif Zafar. Courtesy of Zafar 317

Fig. 6.12 RV64X block diagram 319

Fig. 6.13 RV64X’s scalable design 319

Fig. 7.1 Nvidia’s Turing TU102 GPU die photo and block diagram.
Courtesy of Nvidia 327

Fig. 7.2 Ray tracing features supported in Nvidia’s Turing GPU.
Courtesy of Nvidia 329

Fig. 7.3 Nvidia’s hybrid-rendering technology combining the ray
tracing capabilities of the RT cores and the image denoising ... 330

Fig. 7.4 Data flow of Nvidia’s DLSS 2.0 process. Courtesy of Nvidia ... 332

Fig. 7.5 Nvidia’s DLSS used motion vectors to improve
the supersampling of the enhanced image. Courtesy
of Nvidia 333

Fig. 7.6 An Nvidia demo of ray tracing used in a game. Courtesy
of Nvidia 334

Fig. 7.7 History of Intel graphics devices 338

Fig. 7.8 Intel’s product range for GPUs. Courtesy of Intel 339

Fig. 7.9 EMIB created a high-density connection between the Stratix
10 FPGA and two transceiver dies. Courtesy of Intel 339

Fig. 7.10 Intel plans to span the entire dGPU market. Courtesy of Intel ... 340

Fig. 7.11 Pat Gelsinger, Intel’s CEO. Courtesy of Intel 341

Fig. 7.12 Intel’s DG1 AIB. Courtesy of Intel 342

Fig. 7.13 Intel’s X^e Arc HPG road map circa 2021. Courtesy of Intel 343

Fig. 7.14 Intel’s HPG 344

Fig. 7.15 Classic performance versus quality relationship 346

Fig. 7.16 Intel had a new SDK for its recent supersampling
and scaling algorithm. Courtesy of Intel 346

Fig. 7.17 Example of the quality of Intel’s X^eSS—notice the Caution
sign. Courtesy of Intel 347

Fig. 7.18 AMD RDNA 2 Big Navi GPU 348

Fig. 7.19 AMD RDNA 2 compute unit 348

Fig. 7.20 AMD’s intersection block diagram 351

Fig. 7.21 AMD’s FSR pipeline 352

Fig. 7.22 Innosilicon’s series one family. Courtesy of Innosilicon 356

Fig. 7.23 An HDMI, display port with a VGA connector on the back
of the AIB. Courtesy of Innosilicon 356

Fig. 7.24 Fantasy one type B AIBs. Courtesy of Innosilicon 357

Fig. 7.25 Innosilicon’s innolink IP Chiplet block diagram. Courtesy
of Innosilicon 357

Fig. 7.26 Innosilicon’s roadmap. Courtesy of Innosilicon 358

Fig. 8.1 Slick car from Forza Horizon 5. Courtesy of Xbox Game
Studios 362

Fig. 8.2 Game characters in the 1990s: Doom and Tome raider.
Source Wikipedia 362

Fig. 8.3	Final Fantasy 2001 and Tomb Raider 2013. Courtesy of Wikipedia and Crystal Dynamics	362
Fig. 8.4	Death Standing 2020 and enemies. Courtesy of Sony Interactive Entertainment and Unity	363
Fig. 8.5	Computational fluid dynamics is used to model and test in a computer to find problems and opportunities. Courtesy of Siemens	363

List of Tables

Table 1.1	Second-era GPUs	2
Table 1.2	Comparison of three generations of ATI chips	21
Table 1.3	Comparison of DirectX features	21
Table 2.1	Compute capability of Fermi and Kepler GPUs	78
Table 3.1	Configuration and performance parameters for Mali family of graphics cores	111
Table 3.2	Mali function list	112
Table 3.3	2008 Mali-400 MP GPU process-fill rate	115
Table 3.4	Arm Mali Midgard arithmetic unit per pipeline (per core)	150
Table 3.5	Comparison of Mali-G72 to Mali-G76	155
Table 3.6	Nvidia Tegra SoC product line	169
Table 4.1	Game consoles introduced after the GPU	189
Table 4.2	Nvidia's Shield Console specifications. Courtesy of Nvidia	220
Table 4.3	Atari VCS specifications	226
Table 4.4	Comparison of Sony PlayStation 5 and Microsoft Xbox Series X key specifications	229
Table 4.5	Microsoft's Xbox Series X and S game consoles	232
Table 4.6	Steam Deck's specifications compared to Nintendo's Switch	233
Table 5.1	SiliconArts feature set	243
Table 5.2	RayCore versus CPU ray tracing	244
Table 5.3	SiliconArt's RayCore Lite specifications	246
Table 5.4	SiliconArt's RayCore MC specifications	246
Table 5.5	Bolts specifications	264
Table 5.6	Comparison JM9000-series GPUs to Nvidia GTX1000-series GPU	266
Table 5.7	A comparison of features of Nema and Neox	281
Table 5.8	Nvidia's ampere A100 specifications	289
Table 5.9	Nvidia ampere GPUs	292
Table 5.10	Nvidia's hopper H100 GPU compared to previous GPUs	299

Table 7.1	Intel’s 2020 discrete mobile GPU	337
Table 7.2	Intel’s Arc alchemist mobile dGPU product line	345
Table 7.3	AMD’s Radeon series AIBs	350
Table 7.4	AMD’s quality settings versus performance	353