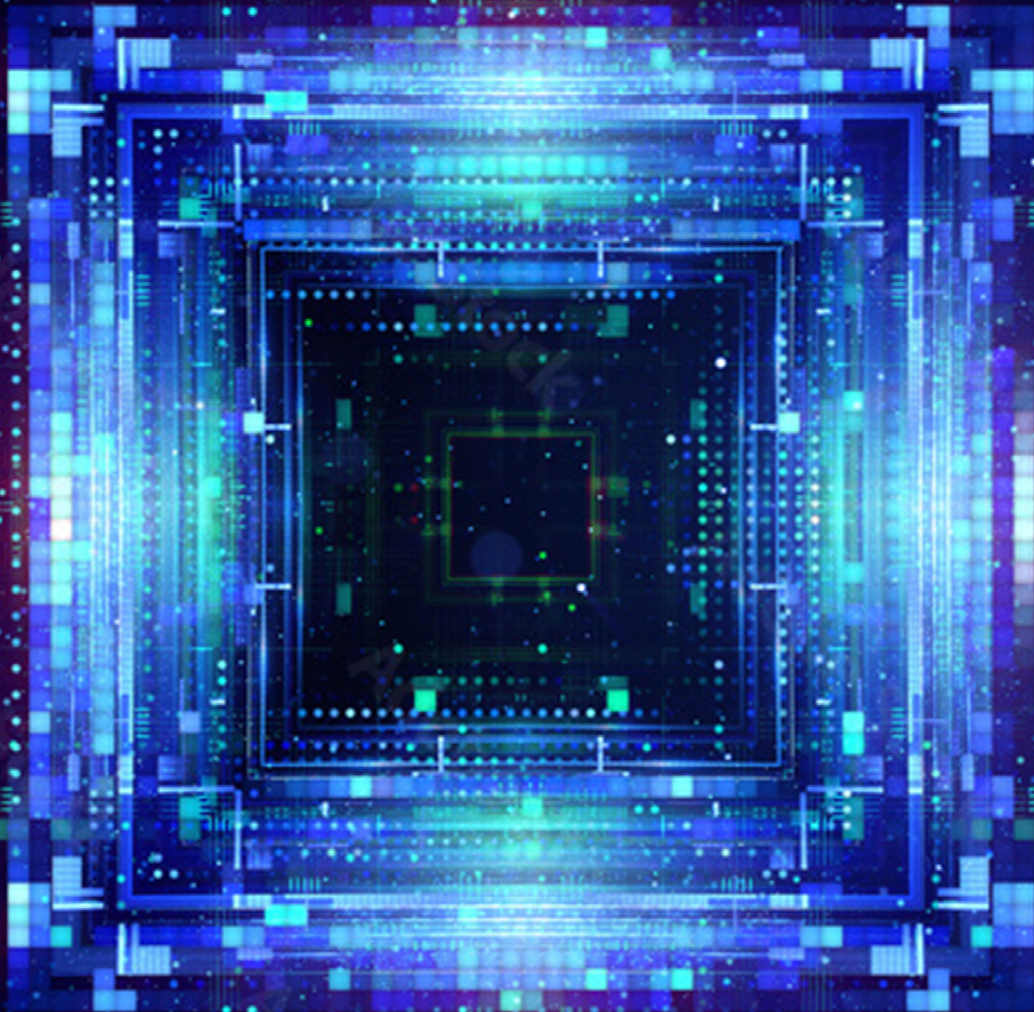


GPU Developments 2022





Copyright Jon Peddie Research 2023. All rights reserved.

Reproduction in whole or in part is prohibited without written permission from Jon Peddie Research.

This report is the property of Jon Peddie Research (JPR) and made available to a restricted number of clients only upon these terms and conditions.

Agreement not to copy or disclose. This report and all future reports or other materials provided by JPR pursuant to this subscription (collectively, “Reports”) are protected by (i) federal copyright, pursuant to the Copyright Act of 1976; and (ii) the nondisclosure provisions set forth immediately following.

License, exclusive use, and agreement not to disclose. Reports are the trade secret property exclusively of JPR and are made available to a restricted number of clients, for their exclusive use and only upon the following terms and conditions. JPR grants site-wide license to read and utilize the information in the Reports, exclusively to the initial subscriber to the Reports, its subsidiaries, divisions, and employees (collectively, “Subscriber”). The Reports shall, at all times, be treated by Subscriber as proprietary and confidential documents, for internal use only. Subscriber agrees that it will not reproduce for or share any of the material in the Reports (“Material”) with any entity or individual other than Subscriber (“Shared Third Party”) (collectively, “Share” or “Sharing”), without the advance written permission of JPR.

Subscriber shall be liable for any breach of this agreement and shall be subject to the cancellation of its subscription to Reports. Without limiting this liability, Subscriber shall be liable for any damages suffered by JPR as a result of any Sharing of any Material, without advance written permission of JPR.

Important Note re Sharing of Material. If Subscriber wishes to Share any Material with any Shared Third Party, it shall contact JPR and request permission to do so. JPR shall forthwith prepare an appropriate Sharing agreement, which shall set forth the name and contact information of the Shared Third Party, the terms and conditions of such Sharing, the compensation to JPR for such Sharing, and the agreement by Subscriber and the Shared Third Party that the Shared Third Party shall not, under any circumstances, Share any of the Material with any other entity or person.

Disclaimers. JPR Reports contain “reviews” of various products and information. The content of these Reports represents the interpretation and analysis of statistics and information that are either generally available to the public or released by entities or individuals deemed responsible. The information in the Reports is believed to be reliable but is not guaranteed as to its accuracy or completeness. The Reports do not endorse or attempt to sell any products and should not be interpreted differently. JPR assumes no responsibility for the correct or incorrect usage of any trademarks or service marks.

Jon Peddie Research
4 Saint Gabrielle Court
Tiburon, CA 94920-1619
+1 (415) 435-9368

Table of Contents

INTRODUCTION	15
Where do GPUs come from?	18
Where do GPUs go?.....	19
Integrated vs. discrete performance	21
THE FIRST QUARTER OF 2022.....	25
Q4'21 saw a nominal rise in GPU and PC shipments quarter-to-quarter	25
Quick highlights	26
The kickoff.....	27
AMD.....	28
Intel.....	30
Nvidia.....	33
Qualcomm	38
Ethereum mining GPU days are over—gamers rejoice	42
AMD backfills workstation lineup with the Radeon Pro W6400 and W6300 Pro	43
What do we think?.....	45
Announced at CES, the Nvidia entry-level Ampere RTX 3050 is tested	45
What do we think?.....	49
Nvidia GeForce GT 1010 AIB available	49
What do we think?.....	52
International Solid-State Circuits Conference	52
Intel brings a bonanza to ISSCC 2022	52
Bonanza Mine	55
What do we think?.....	56
Gunnir enters AIB market with Intel GPU	56
What do we think?.....	59
RX 6500 XT: AMD's first entry-level AIB featuring RNDA 2 technology	60
Two sides of the chip business.....	65
Nvidia's Hopper data center GPU.....	67
What do we think?.....	72
Intel announces Xe-based A-series GPU for notebooks.....	73
What do we think?.....	78
THE SECOND QUARTER OF 2022.....	80
Quick highlights	80
GPUs: How did we get here?.....	81
Nvidia's RTX 3090 Ti	92
Once again Nvidia provides us with the most powerful GPU in the solar system	93
What do we think?.....	99
Intel buys Siru	99
So, what's the story with GPUs?	101

Moore Threads announces three new AIBs	108
What do we think?.....	111
AMD introduces trio of new add-in boards for the 6000 series.....	112
What do we think?.....	114
Intel Arcs into notebooks	114
What do we think?.....	115
AIB price inflation on the horizon	115
AMD's FidelityFX Super Resolution 2.0 is a game changer.....	117
What do we think?.....	120
Nvidia leaves a 'paper' trail	120
Nvidia virtually at Computex.....	124
What do we think?.....	131
Arm says no to inferencing	131
What do we think?.....	134
THE THIRD QUARTER OF 2022.....	135
Q2'22 saw a significant decline in shipments qtr-qtr the worst may not be over.....	135
Quick highlights	136
The 8K debate: Is 8K gaming worth the cost?.....	136
The case for 8K	137
What do we think?.....	139
Is 8K visible and desirable?	140
Ray tracing in Innosilicon GPU	143
What do we think?.....	145
Paper launch? Intel releases photos of Arc Pro graphics boards at Siggraph	145
Intel announces path guiding library and other rendering marvels.....	146
What do we think?.....	150
Nvidia Siggraph keynote.....	151
Intel's GPU lineup (we think).....	159
Birin Technology introduces GPU-compute chip.....	161
What do we think?.....	165
SPEC's latest WS graphics benchmark	165
Licensing.....	167
Intel benchmarks the Arc A750 against an RTX 3060	167
Glenfly Arise GT-10C0 AIB.....	170
Intel's long-anticipated entry into the graphics add-in board market has arrived.....	171
Basemark debuts enhanced mobile device benchmark	173
What do we think?.....	175
The hottest chips conference.....	175
Pat Gelsinger's plan: The future	176
The technology pillars.....	176
Scaling up and out.....	176

It's all about the foundry.....	178
What do we think?.....	178
GPU compute.....	179
Biren BR100.....	179
The incumbents.....	184
AMD Instinct MI200.....	184
Intel's Ponte Vecchio.....	187
Nvidia's Hopper.....	188
Summary.....	192
MetaX Tech plans to have a gaming GPU for the domestic market by 2025.....	193
Intel's Arctic Sound-M Flex Series AIBs.....	194
EVGA quits Nvidia.....	197
What do we think?.....	201
SPEC updates benchmark for SolidWorks.....	202
Intel's wonderful Flex AIBs.....	203
What do we think?.....	205
Feel the love for Nvidia's latest GPUs: Ada Lovelace.....	206
How close were the RTX 4090 leaks?.....	216
Gaming monitors shipments to see 10% decline YoY to 20.5 million units in 2022.....	218
What do we think?.....	220
THE FOURTH QUARTER OF 2022.....	221
GPUs experience biggest qtr-to-qtr drop in Q3'22 since the 2009 recession.....	221
GPUs are damn tricky.....	223
The RTX 4090, a successful debut of the Ada Lovelace series.....	224
What do we think?.....	235
Intel has released the long-awaited Intel Arc A750 and A770.....	236
SPEC updates its test suite.....	237
What do we think?.....	240
AMD slam-dunks GPU-Compute test.....	240
What do we think?.....	242
Immortalis GPU in new MediaTek 9200?.....	242
What do we think?.....	243
China's Biren to feel the heat of US sanctions.....	243
What do we think?.....	245
Rumor: Apple Mac Pro to get big GPU.....	246
AMD's RDNA 3 chiplet GPU and RX 7900 XTX.....	247
Intel's XeSS is agnostic but is it asymmetric?.....	251
What do we think?.....	254
Basemark introduced platform agnostic ray tracing benchmark.....	254
AIB shortage tuns into a glut.....	262
Everybody is a gamer.....	263

Moore Threads adds S3000 server GPU based on Chunxiao architecture	264
The improved RADV open-source Radeon Vulkan driver.....	267
So what?	268
Nvidia satisfies US export controls with GPU for China	268
Biren thwarted from building mythical chip.....	269
MediaTek Dimensity 9200 will get the rays right	271
What do we think?.....	274
Intel Maxes their data center GPUs	275
What do we think?.....	278
Real-time ray tracing on a smartphone—really?	278
What do we think?.....	281
The improved RADV open-source Radeon Vulkan driver.....	282
So what?	282
Changes in interfaces, changes in attitudes.....	283
USB4	283
DisplayPort 2.1	284
Thunderbolt 5.....	284
AMD ups the ante for gaming displays	284
The Intel Arc A700 arrives and acquits itself handily	286
Nvidia 4000 pro graphics results—a preview	289
MetaX Tech plans to have a gaming GPU for the domestic market by 2025.....	292
Moore Treads AIB available in China	293
What do we think?.....	295
Nvidia RTX4080 Founder’s Edition AIB	295
What do we think?.....	299
The Nvidia 4090 versus everything	300
AMD introduces its flagship add-in board from the new line of RDNA 3-based graphics cards	302
WHAT’S NEW FOR 2023	313
Nvidia will offer RTX4070 Ti	313
Faster monitors.....	315
A shameless plug—buy a book.....	315
Index.....	316

Table of Figures

Figure 1. The first-ever PC AIB report—circa 1987	15
Figure 2. PC suppliers not including intellectual property providers	16
Figure 3. A silicon wafer with dozens of die chips.....	18
Figure 4. GPU designers and developers.....	19
Figure 5. X86-based GPU taxonomy.....	20

Figure 6. Improvement in GFLOPS of GPUs over time	22
Figure 7. Quarterly shipments and market share percentages and year-to-year results.....	26
Figure 8. Laura Smith introduced the Ryzen 7 5800X (Courtesy AMD).....	29
Figure 9. The company is introducing eight new 12th gen CPUs (Courtesy Intel).....	31
Figure 10. Bryant said 70% of users access an app across mobile devices, and 90% of those people use multiple screens to accomplish a single task	31
Figure 11. Lisa Pierce, Intel’s VP of graphics software and engineer, said Intel will be over 50 new notebook and desktop PCs	32
Figure 12. Market figures as seen by Nvidia (Courtesy Nvidia)	33
Figure 13. Jeff Fisher demonstrates the benefits of going to the gym as he holds up the 25-pound 40 TFLOPS RTX 3090 Ti with just one hand	35
Figure 14. Ali Kani, Nvidia’s GM & VP of automotive, said autonomous vehicles are the most intense challenge to AI.....	36
Figure 15. Going for a little ride with Nvidia (Courtesy Nvidia).....	36
Figure 16. Advancement in autonomous vehicles increases GPU and infrastructure requirements (Courtesy Nvidia).....	38
Figure 17. One technology roadmap (Courtesy Qualcomm).....	39
Figure 18. Snapdragon Ride platform vision system in vehicles in 2024 4-nm vision (Courtesy Qualcomm)	40
Figure 19. Snapdragon for everything (Courtesy Qualcomm)	41
Figure 20. Cristiano Armon’s grand plan for Qualcomm’s takeover of the world (Courtesy Qualcomm)	41
Figure 21. AMD’s benchmarks show the RDNA 2 over CGN and RDNA (Courtesy AMD).....	43
Figure 22. The Radeon Pro W6000 boards (Courtesy AMD)	44
Figure 23. Frame rate performance Evil Genius 2.....	47
Figure 24. Frame rate performance Chernobyl Lite	48
Figure 25. The RTX 3050 performed well in the Pmark due to its frame rate performance, low price point and low wattage.	48
Figure 26. In raw TFLOPS the Nvidia RTX 3050 did very well	49
Figure 27. Nvidia’s GP108-200-A1 GPU circa 2021 (Courtesy Nvidia).....	50
Figure 28. Nvidia’s reference design version of the GT 1010 AIB (Courtesy Nvidia).....	51
Figure 29. Lenovo’s GT1010 AIB (Courtesy https://market.m.taobao.com/).....	51
Figure 30. Intel’s Ponte Vecchio AIB is named after a medieval stone bridge that connects the Piazza della Signoria on one side of the Arno River in Florence, Italy, with the Palazzo Pitti on the other side (Photo by Matt Hardy from Pexels)	53
Figure 31. Intel’s Ponte Vecchio GPU with 3D and 2D system partitioning with Foveros and embedded interconnect bridges (EMIB) (Courtesy ISSCC).....	54
Figure 32. Intel’s tiny low-power bitcoin ASIC (Courtesy Bonanza).....	55
Figure 33. Intel’s Bonanza Mine system is a self-contained unit comprising four hash boards, an Intel FPGA-based control unit, programmable power supply, and four fans, delivering hash rate of 40THash/s under a thermal design power envelope of 3600W (Courtesy ISSCC).....	56
Figure 34. Gunnir’s Lanji DG1-based AIB (Courtesy Gunnir).....	57
Figure 35. Gunnir AIBs on Taobao’s web page (Courtesy Taobao)	58
Figure 36. Comparison of Intel GPUs (Courtesy VideoCardz.com).....	59
Figure 37. Gunnir’s Iris Xe Max Index V2 (Courtesy Gunnir)	60
Figure 38. AMD RX6600 XT (Courtesy Gigabyte).....	61

Figure 39. Frames per second running Evil Genius 2.....	63
Figure 40. Frame per second running Godfall.....	63
Figure 41. Pmark FPS scores.....	64
Figure 42. Pmark FPS scores with FidelityFX enabled.....	64
Figure 43. The RX 6500 XT (bottom) compared the Gigabyte RX 6600 XT Eagle (Courtesy AMD).....	65
Figure 44. Imagination Technologies' business is taking off, resulting in a hiring boom (Courtesy Imagination Technologies).....	66
Figure 45. H100 compute improvement summary (Courtesy Nvidia).....	67
Figure 46. Nvidia's Hopper subsystem board (Courtesy Nvidia).....	68
Figure 47. Time to train the Mixture of Experts transformer network for H100 versus A100 (Courtesy Nvidia).....	69
Figure 48. Nvidia's H100 Hopper AIB with NVLink (upper left) supports a unified cluster of eight GPUs.(Courtesy Nvidia).....	70
Figure 49. Nvidia's DGX H100 supercomputer.(Courtesy Nvidia).....	71
Figure 50. Nvidia's Earth 2 super supercomputer (Courtesy Nvidia).....	71
Figure 51. Intel's Xe road map—the names, however, are subject to change (Courtesy Intel)....	74
Figure 52. There are 16 256-bit vector engines in each XVE and 16 1,014-bit matrix engines in each XMX (the little purple box next to the XVE (Courtesy Intel).....	75
Figure 53. Xe Arc matrix engine operations (Courtesy Intel).....	76
Figure 54. Intel's new Dynamic Power Share technique (Courtesy Intel).....	77
Figure 55. The whole package is summed up this illustration. (Courtesy Intel).....	78
Figure 56. Intel's future AIB? (Courtesy Intel).....	79
Figure 57: Quarterly shipments, market share percentages, and year-to-year results.....	80
Figure 58. Ikonas graphics system.....	83
Figure 59. Ikonas matrix multiplier.....	84
Figure 60. Ikonas RDS-3000.....	85
Figure 61. Pixel-Planes 4 (circa 1986).....	87
Figure 62. AT&T Pixel Machine.....	89
Figure 63. IBM's Professional Graphics Adapter, also known as the Professional Graphics Controller. (Courtesy Wikipedia).....	91
Figure 64. EVGA's RTX 3090 Ti FTW3 Ultra Gaming AIB in its natural environment.....	93
Figure 65. The ASUS MARS II released in 2011 was truly a behemoth.....	94
Figure 66. The RTX 3090 Ti and the RTX 3080 Ti—the size difference is obvious.....	95
Figure 67. Pmark comparison.....	97
Figure 68. The RTX 3090 Ti produces over 50 FPS in 4K gaming with ray tracing enabled.....	97
Figure 69. The RTX 3090 Ti performs very well in ray tracing-based tests.....	98
Figure 70. Metro Exodus Enhanced.....	98
Figure 71. GPU taxonomy.....	103
Figure 72. What a retail store might have looked like before the scalpers and miners.....	105
Figure 73. A GPU has many different names depending on its usage.....	107
Figure 74. Moore Threads launched its first GPU, Sudi GPU, based on its unified system architecture, MUSA (Source: Moore Threads).....	109
Figure 75. AMD's AIB line up.....	113
Figure 76. High-end AIB retail prices over time.....	116
Figure 77. Pricing is outstripping volume.....	117

Figure 78. AMD’s FSR (2.0) is in the output stage as a post processor (Courtesy AMD)	118
Figure 79. Compared to FSR 1.0, FSR 2.0 is earlier in the frame pipeline (Courtesy AMD).....	118
Figure 80. Comparison of native to scaled at 4K—if it doesn’t look different, that’s the whole idea (Courtesy AMD and Arkane Studios).....	120
Figure 81. A character can learn more than one action at a time using AI reinforcement learning (Courtesy Nvidia).....	121
Figure 82. New research into VR glasses (Courtesy Nvidia).....	122
Figure 83. Indirectly lit teapots made of metal, ceramic, and glass (Courtesy Nvidia).....	123
Figure 84. A text-driven method allows shifting a generative model to new domains, without collecting a single image (Courtesy Nvidia).....	124
Figure 85. Ian Buck.....	125
Figure 86. The Hopper GPU has a new Transformer Engine with FP8 Tensor Core	125
Figure 87. Brian Kelleher	126
Figure 88. Nvidia’s road map (Courtesy Nvidia)	126
Figure 89. Grace CPU (left) and GPU processor (Courtesy Nvidia).....	127
Figure 90. Ying Yin Shih.....	127
Figure 91. An Nvidia refence design server chassis (Courtesy Nvidia).....	128
Figure 92. Deepu Talla	128
Figure 93. Jeff Fisher	129
Figure 94. Improvement in aiming due to latency reduction (Courtesy Nvidia and Kovaak).....	130
Figure 95. Tion Thomas.....	130
Figure 96. The company also claims their new GPUs will deliver 2× architectural ML improvements (Courtesy Arm)	133
Figure 97. Arm’s road map (Courtesy Arm)	133
Figure 98. Quarterly shipments and market share percentages, as well as year-to-year results..	135
Figure 99. 8K resolutions are much bigger than HD and perfect for VR (Courtesy PCMag).....	137
Figure 100. Nvidia showcasing its DLSS technology by comparing the different resolutions (Courtesy Nvidia).....	138
Figure 101. 8K requires big monitor sizes. More pixels, more screen (Courtesy Sony).....	140
Figure 102. Seeing more with 8K.....	141
Figure 103. Innosilicon’s 2021 Fenghua-based Fantasy 1 AIB (Courtesy Innosilicon).....	143
Figure 104. Innosilicon’s Fenghua 2 dGPU (Courtesy Innosilicon)	144
Figure 105. There will be three versions of the Pro AIBs, according to the web page, 3.5 TFLOPS (SP) for A40, 4.8 TFLOPS (SP) for the A50, and 3.5 TFLOPS (SP) for the A30 (Courtesy Intel)	146
Figure 106. An interior scene with complex, diffuse, multi-bounce global illumination (Courtesy Intel).....	148
Figure 107. A flooded version of Intel’s famous Sponza rendered with Blender’s Cycles (Courtesy Intel).....	149
Figure 108. Underwater rendered with Blender’s Cycles (Courtesy Intel)	150
Figure 109. Sanja Fidler, Nvidia’s VP of AI Research (Courtesy Nvidia).....	152
Figure 110. Rev Lebedev, VP Omniverse & Simulation technology (Courtesy Nvidia).....	153
Figure 111. Steven Parker, Vice President, Professional Graphics (Courtesy Nvidia)	154
Figure 112. Nvidia has introduced the MDL distiller (Courtesy Nvidia).....	154
Figure 113. OpenVDB hierarchical tree-structure (Courtesy Nvidia).....	155

Figure 114. Example of compression and memory reduction using NeuralVDB (Courtesy Nvidia)	156
Figure 115. Nvidia also introduced a suite of tools for Omniverse (Courtesy Nvidia)	156
Figure 116. Nvidia's Omniverse roadmap (Courtesy Nvidia).....	157
Figure 117. Nvidia's partner list for Omniverse (Courtesy Nvidia).....	157
Figure 118. Simon Yuen, director of avatar technology at Nvidia, thinks avatars will be everywhere (Courtesy Nvidia).....	158
Figure 119. Nvidia's Ace AI avatar suite for recommenders (Courtesy Nvidia)	159
Figure 120. Zhang Wen, founder, chairman and CEO of Birin Technology.....	161
Figure 121. The Biren BR100 specifications (Courtesy Biren).....	162
Figure 122. Biren's GPU-compute modules (Courtesy Biren).....	162
Figure 123. The Biren is constructed from two chiplet devices (Courtesy Biren)	163
Figure 124. Hong Zhou, co-founder and CTO showed the BR104 GPU (Courtesy Biling).....	164
Figure 125. Biren's BR104-based IB (Courtesy Biren).....	165
Figure 126. SPEC's latest graphics benchmark viewset, created from API traces of the latest version of 3ds Max (Courtesy SPEC).....	166
Figure 127. Tom Petersen describes a thin API (Courtesy Intel)	168
Figure 128. Intel's Arc A750 performance in games (Source Intel)	169
Figure 129. Intel's Arc A750 at various resolutions (Source Intel).....	169
Figure 130. Intel's Arc A750 AIB (Courtesy Intel).....	170
Figure 131. Glenfly Arise GT-10C0 AIB (Courtesy Glenfly).....	171
Figure 132. Intel's A380 AIB (Courtesy Intel).....	173
Figure 133. Screenshot from Basemark's VRS benchmark for mobile (Courtesy Basemark) ...	174
Figure 134. At Hot Chips, Gelsinger presented the chip as a harbinger of the future (Courtesy Intel presentation)	177
Figure 135. Biren BR100 block diagram (Courtesy Biren).....	180
Figure 136. Biren's BR100 SPC architecture block diagram (Courtesy Biren).....	181
Figure 137. Biren's SPC GEMM block diagram (Courtesy Biren).....	182
Figure 138. The BR100 employs a NUMA and a UMA memory (Courtesy Biren).....	183
Figure 139. With a flexible memory structure, Biren thinks they can approach near-memory computing (Courtesy Biren)	183
Figure 140. AMD's multi-chip fabric construct (Courtesy AMD).....	184
Figure 141. Block diagram of AMD's MI250X MCM (Courtesy AMD).....	185
Figure 142. AMD's MI250X unified computing node (Courtesy AMD)	186
Figure 143. Intel's Ponte Vecchio 4X subsystem with X ^e Links (Courtesy Intel)	187
Figure 144. Nvidia's Hopper board (Courtesy Nvidia)	188
Figure 145. Nvidia's H100 Hopper tensor core cGPU (Courtesy Nvidia).....	189
Figure 146. Nvidia's Hopper SM architecture (Courtesy Nvidia).....	190
Figure 147. Nvidia's thread block cluster example (Courtesy Nvidia)	192
Figure 148. The Flex GPU is described by Intel as a midpoint between desktop GPUs and the high-end GPUs like the also delayed Ponte Vecchio (Courtesy Intel)	196
Figure 149. EVGA's first Nvidia cooler (JPR).....	198
Figure 150. EVGA RTX 3090 (Courtesy EVGA).....	200
Figure 151. Nvidia cuts price on RTX 3090 Ti on BestBuy	201
Figure 152. The benchmark measures application performance for workstations running SolidWorks 2022 (Courtesy SPEC).....	202

Figure 153. Intel's new campaign (Courtesy Intel)	204
Figure 154. Intel's media stack (Courtesy Intel)	205
Figure 155. Intel's Flex AIBs (Courtesy Intel).....	206
Figure 156. Photo of Ada Lovelace GPU die (Courtesy Nvidia).....	208
Figure 157. The AD100 has 269 times more transistors than the A100 (Courtesy Nvidia).....	209
Figure 158. Nvidia's Ada Lovelace power-performance efficiency compared to previous generations (Courtesy Nvidia)	210
Figure 159. Nvidia's new 3 rd -gen cores (Courtesy Nvidia).....	212
Figure 160. Nvidia's RTX 4090 AIB (Courtesy Nvidia)	213
Figure 161. Lady Lovelace (Image courtesy Oxford Today)	215
Figure 162. Dell's 34-inch curved gaming monitor (Courtesy Dell).....	219
Figure 163. Shipments of gaming monitors over time and forecasted (Courtesy TrendForce) ..	219
Figure 164. Quarterly shipments and market share percentages, and year-to-year results.....	221
Figure 165. CPU shipments platform share and units	222
Figure 166. The Ada Lovelace GPU die (Courtesy Nvidia).....	224
Figure 167. The RTX 4090 Founder's Edition at home in its natural environment.....	227
Figure 168. Average frame rates in 4K with Vulkan API	229
Figure 169. Percent change for the RTX 4090 over the RX 6900 XT and RTX 3090.....	229
Figure 170. Average frame rates in 4K with DX12 API	230
Figure 171. Percent change for the RTX 4090 over the RX 6900 XT and RTX 3090.....	231
Figure 172. Percent change for the RTX 4090 over the RX 6900 XT and RTX 3090.....	232
Figure 173. Scores for the RTX 4090, RTX 3090, and RX 6900 XT in 3DMark.....	232
Figure 174. Percent change for the RTX 4090 over the RX 6900 XT and RTX 3090 in 3DMark	233
Figure 175. Pmark equation.....	233
Figure 176. Pmark comparison.....	234
Figure 177. Nvidia's Ada Lovelace power-performance efficiency compared to previous generations (Courtesy Nvidia)	235
Figure 178. Comparison of Intel A770 and A750 AIBs (Courtesy Intel)	236
Figure 179. Intel's Arc A770 AIB (Courtesy Intel).....	237
Figure 180. SPEC's Sandra 2020 GPU-COMPUTE test results	241
Figure 181. MediaTek's new SoC (Courtesy MediaTek).....	242
Figure 182. Biren BR100 specifications (Courtesy Biren).....	244
Figure 183. Biren claims a faster GPU-compute chip than Nvidia (Courtesy Biren)	245
Figure 184. AMD RDNA3 block diagram (Courtesy AMD).....	247
Figure 185. AMD's RDNA AI accelerators within the compute engine (Courtesy AMD)	248
Figure 186. AMD's RDNA 3 ray-tracing unit within the compute unit (Courtesy AMD)	249
Figure 187. AMD RX 7900 XRX (Courtesy AMD)	250
Figure 188. XeSS uses recent frame data to estimate current frames (Source Intel).....	252
Figure 189. Intel's XeSS quality vs performance results (Source Intel)	253
Figure 190. Intel's XeSS control panel (Source Intel).....	254
Figure 191. Imagination was a pioneer in real-time hardware accelerated ray tracing (Source Imagination Technologies)	255
Figure 192. Who needs wants ray tracing? (Source Imagination Technologies)	256
Figure 193. The levels of ray tracing (Source Imagination Technologies)	256
Figure 194. Imagination Technologies ray tracing demos (Source Imagination Technologies).....	257

Figure 195. Summary of Imagination Technologies point of view (Source Imagination Technologies).....	257
Figure 196. Real-time ray tracing is now available on PCs and mobile devices (Source Basemark)	258
Figure 197. There aren't too many games yet on mobiles that employ ray tracing (Source Basemark)	258
Figure 198. The GPU Score Basemark reports includes advanced functions like global illumination (Source Basemark)	259
Figure 199. The ray tracing benchmark will run across all popular APIs and operating systems (Source Basemark).....	259
Figure 200. Basemark's three ray-tracing tests (Source Basemark).....	260
Figure 201. Two of Basemark's three ray tracing tests (Source Basemark).....	260
Figure 202. Basemark's third ray tracing test (Source Basemark).....	261
Figure 203. Vietnam's Lê Thành, the self-proclaimed 'King of VGA, made a comedy video about selling AIBs by the pound.....	262
Figure 204. Henry Cavill lifting a gaming PC off a table (Provided by Digital Trends)	264
Figure 205. Moore Threads' S80 gaming AIB (Source: Moore Threads).....	266
Figure 206. Chip and specs (Source: Moore Threads)	267
Figure 207. Nvidia's A800 (Courtesy Nvidia)	269
Figure 208. Alleged benchmark results (Source: Biren)	270
Figure 209. MediaTek's Dimensity 9220 specifications (Source: MediaTek).....	272
Figure 210. MediaTek Dimensity 9200 block diagram (Courtesy MediaTek).....	273
Figure 211. Ray tracing on a handheld device.....	274
Figure 212. Intel's Max GPU overview (Courtesy Intel)	275
Figure 213. Intel's Max 1350 and 1550 OAM (Courtesy Intel).....	276
Figure 214. Intel's inter-module communications system Xe Link (Courtesy Intel).....	277
Figure 215. Intel's next generation data center GPU (Courtesy Intel)	277
Figure 216. Intel Tech At SC22 Overview (Courtesy Intel).....	278
Figure 217. Real-time ray tracing on a mobile device with hardware acceleration demonstrated to the public for the first time (Source: Qualcomm).....	279
Figure 218. Qualcomm Snapdragon 8 sketch block diagram (Source: Qualcomm)	280
Figure 219. Readers may remember that USB4 was developed from Thunderbolt technology that Intel made available	284
Figure 220. AMD said the 4K performance will be fast enough for professional esports gaming, allowing the use of better displays	285
Figure 221. Intel Arc 770 and 750 AIBs.....	286
Figure 222. Intel's Arc A750 is a solid, economical performer	288
Figure 223. Resizable BAR on and off.....	289
Figure 224. Hexagon built their product MSC Apex Generative Design from the ground up using GPUs. Not only is the finished product faster, but it also combines the functions of design, meshing, and analysis in one product. (Source: Hexagon).....	290
Figure 225. Abaqus has several features that benefit from GPU acceleration. For instance, Abaqus gets a performance boost when the AMS Eigensolver comes into play for full vehicle models with a large number of nodes. (Source: Dassault Systèmes)	291
Figure 226. An analysis of stress points in a wheel design	292
Figure 227. Moore Thread's S80 circuit board (Source Expeview).....	294

Figure 228. Comparison of price-performance for Nvidia AIBs.....	297
Figure 229. Pmark for various performance criteria.....	298
Figure 230. Pmark results for FPS, TFLOPS, and scores.....	301
Figure 231. Price–performance of five AIBs.....	301
Figure 232. AMD’s new RX 7900 XTX and RX 7900 XT.....	302
Figure 233. AMD’s Radeon RDNA 3.0 chip	303
Figure 234. The AMD RX 7900 XTX in its natural environment.....	306
Figure 235. Average frame rates in 4K and 1440p with Vulkan API, with ray tracing when available	307
Figure 236. Average frame rates in 4K and 1440p with DirectX 12 API, with ray tracing where available	308
Figure 237. Scores for the 3D Mark test suite	308
Figure 238. The PowerColor Hellhound RX 7900 XTX.....	309
Figure 239. Pmark scores for averages of parameters	310
Figure 240. These synthetic ray-tracing benchmarks were performed using the AMD reference add-in board; all other tests in this article were completed with the PowerColor Hellhound RX 7900 XTX	311
Figure 241. Gigabyte’s ARC AIBS (Source Gigabyte).....	312
Figure 242. PC sales fell 2022	313
Figure 243. The History of the GPU—three volumes	315

Table of Tables

Table 1. GPU suppliers	16
Table 2: PC dGPU shipment market shares.....	26
Table 3. AMD’s comparison of its latest products (Courtesy AMD).....	45
Table 4. AIB Specs	46
Table 5. Pmark Variables.....	47
Table 6. Pmark Comparison	47
Table 7. Benchmark scores for Nvidia GT 1010.(Courtesy Geekbench via @BenchLeaks).....	52
Table 8. AIB specifications.....	61
Table 9. Test suite	62
Table 10. Pmark variables.....	62
Table 11. Pmark scores comparison	62
Table 12. Nvidia’s Hopper H100 GPU compared to previous GPUs (Courtesy Nvidia).....	68
Table 13. Intel Arc GPU specifications (Intel)	76
Table 14. PC dGPU shipment market shares.....	81
Table 15. AIB specifications.....	95
Table 16. Test suite	96
Table 17. Results and Pmark variables.....	96
Table 18. AIB specifications.....	113
Table 19. AMD’s FSR 2.0 Quality mode results. (Courtesy AMD)	119
Table 20. AMD’s FSR 2.0 Performance mode results (Courtesy AMD).....	119
Table 21. Those crazy Ks	142
Table 22. A possible lineup of the coming Intel Arc GPU lineup (Courtesy here and there)	160

Table 23. Specifications of the ASRock Challenger ITX vs. other AIBs in the entry-level segment	172
Table 24. Intel's expected Arc lineup	172
Table 25. Comparison of key specifications of compute GPUs	179
Table 26. Intel's Arctic Sound history	194
Table 27. Intel's Flex data center AIBs	195
Table 28. Intel Flex 170 and 140 specifications (Courtesy Intel).....	204
Table 29. Comparison of Ada Lovelace to last-gen Ampere GPU.....	207
Table 30. GeForce RTX 40 series specifications.....	211
Table 31. Sources for the table are linked in the first row	217
Table 32. Comparison of Ada Lovelace to last-generation Ampere GPU	226
Table 33. AIB comparison	228
Table 34. Vulcan test suite.....	228
Table 35. DirectX 12 test suite	230
Table 36. DirectX 11 test suite	231
Table 37. Synthetic benchmark test suite	232
Table 38. Pmark comparison variables.....	233
Table 39. Snapdragon 8 Gen 2 overall performance	281
Table 40. Tests use in the AIB evaluations.....	287
Table 41. Pmark test results	287
Table 42. Test results	288
Table 43. Comparison of two generations of Nvidia RTX AIBs	296
Table 44. Test results of three Nvidia RTX AIBs	297
Table 45. Power consumption across common use cases (Source Nvidia)	298
Table 46. Power consumption can vary by game and resolution (Source Nvidia).....	299
Table 47. Comparison of the flagship RDNA 3.0 AIB to the previous-generation RDNA 2.0; the numbers for the RX 7900 XTX are reflecting the reference add-in board	304
Table 48. AIB comparison.....	305
Table 49. Vulcan test suite.....	306
Table 50. DirectX 12 test suite	307
Table 51. Synthetic benchmark test suite	308
Table 52. Pmark comparison variables.....	309
Table 53. Ray-tracing synthetic benchmarks.....	310
Table 54. Nvidia RTX 4000 series specifications	314

INTRODUCTION

Welcome to our annual summary of GPU and associated developments for the year.

Jon Peddie Research (JPR) has been tracking, reporting, testing, and forecasting the GPU industry since 2001, and before there were GPUs, Jon Peddie Associates (JPA) tracked PC and workstation graphics from 1984 and published the first add-in board report in 1987.

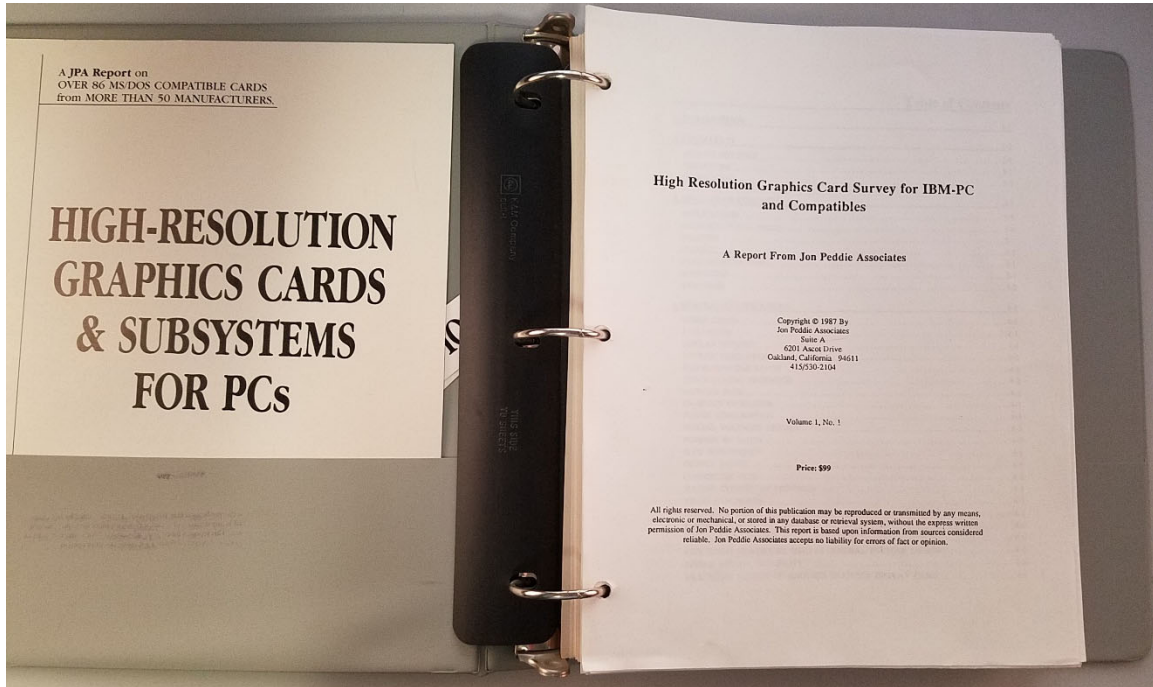


Figure 1. The first-ever PC AIB report—circa 1987

That body of work, comprising 35 years of research, has given us a unique perspective, understanding, and insight into the inner workings of GPUs and their capabilities.

This seventh edition of *GPU Developments* is a compendium of reports from our client service *TechWatch*, our testing service Mt. Tiburon Testing Labs, our quarterly market report on GPUs, *Market Watch*, our semi-annual report *Workstations*, and our quarterly report *Add-in Boards*, as well as excerpts from our semi-annual report on *PC Gaming Hardware*, *Mobile Devices*, *TV Gaming*, *XR Gaming*, and our *Digital Content Creation* reports.

There were quite a few surprises, a couple of disappointments, and a lot of business as usual. In 2022, we saw the introduction of two new GPU suppliers, taking us up to 11.

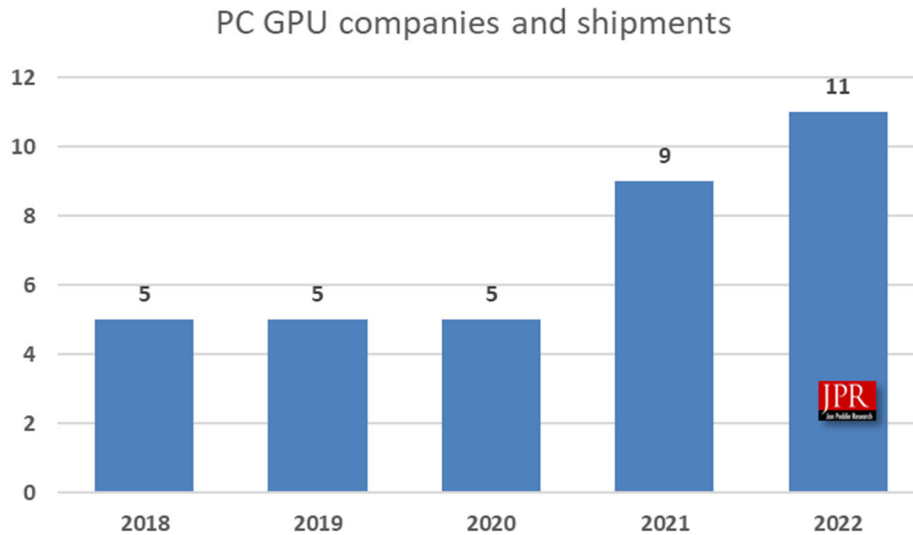


Figure 2. PC suppliers not including intellectual property providers

Intellectual property (IP) suppliers that serve the PC, console, and mobile markets increased to five in 2010 and have remained at that level since. In addition, there are two mobile SoC suppliers, Apple and Qualcomm, that design and manufacture GPUs; all other mobile SoC suppliers used IP from other companies like Arm and Imagination Technologies.

PC	IP	SoC
AMD	Arm	Apple
Bolt	DMP	Qualcomm
Innosilicon	IMG	
Intel	Think Silicon	
Jingia	Verisilicon	
MetaX	Xi-Silicon	
Moore Threads		
Nvidia		
SiArt		
Xiangdixian		
Zhaoxin		

Table 1. GPU suppliers

Counting PC, mobile, and IP, there are 18 companies worldwide designing and producing GPUs. Some of the companies make GPUs for various categories. For example, AMD and Intel make GPUs that are integrated with a CPU, AMD makes GPUs and provides IP, Nvidia makes GPUs for the datacenter, PCs, and automobile, Qualcomm makes SoC used in VR, smartphones, cars, etc.

GPUs, cards, and AIBs

It has become common, although incorrect, to refer to an add-in board (AIB) as a GPU. People might say “the new Nvidia GeForce GPU,” using GPU instead of AIB. So, the terms have