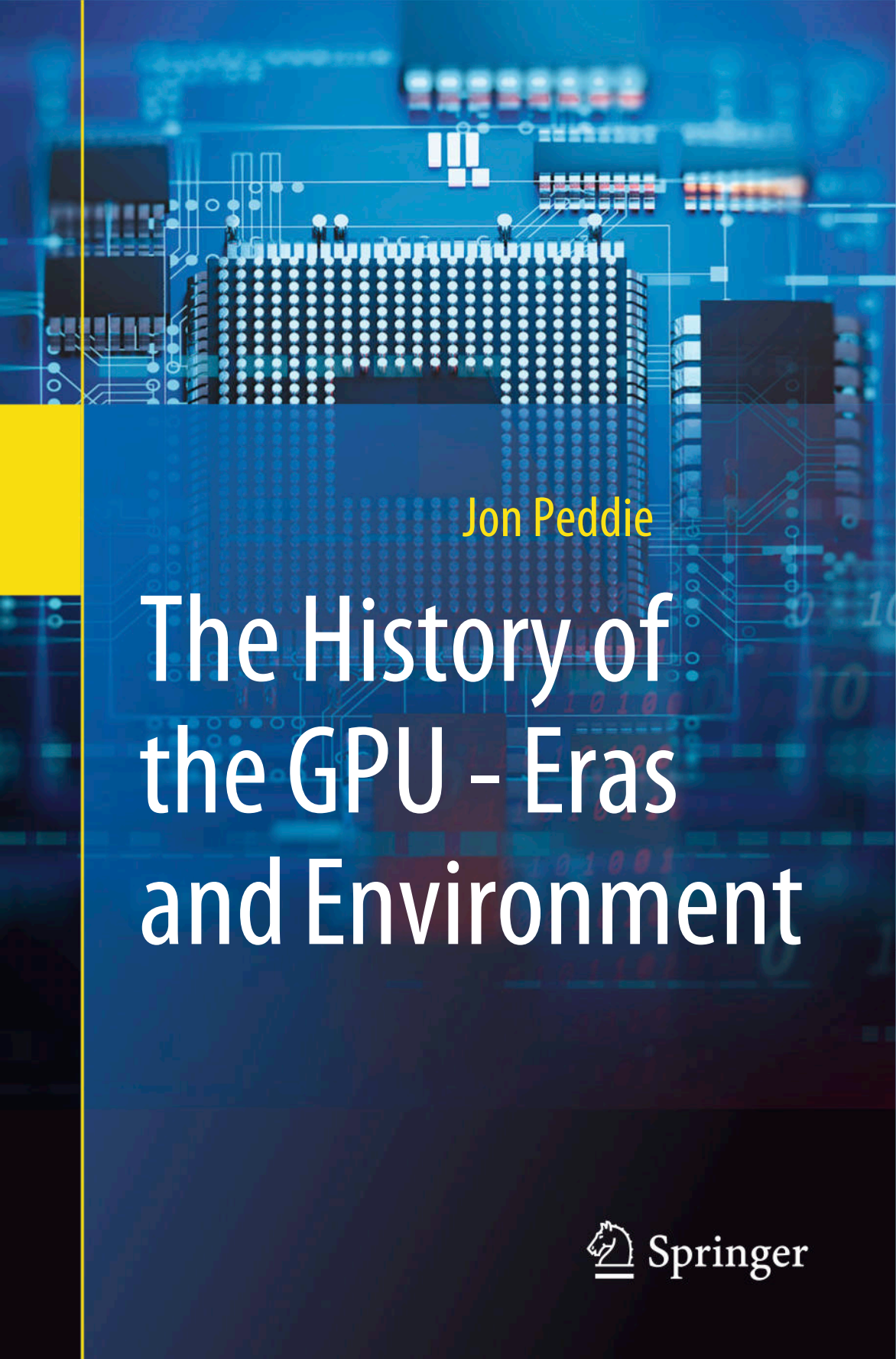
Jon Peddie
The History of the GPU - Eras and Environment
Springer

The History of the GPU - Eras and Environment

Jon Peddie

The History of the GPU - Eras and Environment

Springer

Jon Peddie
Jon Peddie Research
Tiburon, CA, USA

ISBN 978-3-031-13580-4 ISBN 978-3-031-13581-1 (eBook)
https://doi.org/10.1007/978-3-031-13581-1

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

Foreword

Computer graphics has attracted the world's leading computer scientists and the most prominent computer companies. Computer graphics is attractive for its grand computer science challenge, visceral beauty, and broad market impact by sitting at the intersection of computing and art, the simulation of light, physics, and virtual worlds.

The industry was chaotic and fast-changing, with countless companies innovating. In less than three decades, 3D graphics became standard in computers, evolved from a fixed function accelerator into a programmable shading GPU, became a technology juggernaut, and real-time ray tracing became a reality. The GPU became general purpose, and GPU computing was born, democratizing scientific computing, and enabling deep learning and the modern AI era. No one would have believed a chip design that started out running Quake would evolve to become the engine for conversational AI, self-driving cars, climate simulation, and countless applications that impact most modern life.

As the longest running historian of the graphics industry, Jon masterfully lays out the lineage and broad "family tree" of the GPU and the markets and industries it has come to serve. Its adoption, the rapid rate of its improvements, and breadth and depth of its application make it worthy of having its "origins story" explored in the detail and level of completeness for which Jon's books are known.

As you read Jon's book, you will benefit from his insight and ability to "paint the picture" of how we got here, what it means, and where we might go. The reader will see the awe-inspiring amount of innovation, brilliance, determination, guts, and effort from casts of thousands over decades that have made the GPU worthy of such a treatise. Enjoy!!

Nvidia, California, US

Jenson Huang
Curtis Priem

Preface

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

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

This second book in the series covers the developments that lead up to the integrated GPU, from the early 1990s to the late 1990s.

The book has two main sections, the PC platform, and other platforms. Other platforms include workstations, game machines, and others, which include 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 in 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.

The History of the GPU - Eras and Environment

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 45. In addition, there have been over 30 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 within 48 h. 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—Falanx
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—Falanx
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 Guttag—TI
Karthikeyan (Karu) Sankaralingam—University of Wisconsin-Madison
Kathleen Maher—JPA & JPR
Ken Potashner—S3 & SonicBlue
Kristen Ray—Arm
Lee Hirsch—Nvidia
Luke Kenneth Casson Leighton—Libre-GPU
Mark Kilgard—Nvidia (Iris GL)
Mary Whitton—Iknoas
Megan Zea—PCI SIG
Melissa Scuse—Arm
Mike Diehl—HP
Mike Mantor—AND
Mikko Nurmi—Bitboys
Mikko Alho—Siru
Neal Leavitt—Editing
Neil Trevett—3Dlabs & Khronos
Nick England—Iknoas
Pedro Duarte—Universities of Coimbra
Peter McGuinness—SGS Thompson
Peter.L.Segal—AT&T
Petri Norlund—Bitboys
Phil Roges—ATI
Richard Huddy—ATI
Richard Selvaggi—Tseng Labs
Rick Bergman—ATI/AMD
Robert Dow—JPR
Ross Smith—3dfx
Ruchika Saini—Editing
Sasa Marinkovic—ATI & AMD
Simon Fenny—Video Logic & Imagination Technologies
Steve Brightfield—SiliconArts
Steve Edelson—Edson Labs
Stefan Demetrescu—Stanford
Stephen Morein—Stellar
Tatsuo Yamamoto—Sega/DMP
Tim Leland—Qualcomm
Timothy Miller—Traversal Technology
Tom Forsyth—3Dlabs
Tony Tamasi—3dfx & Nvidia
Trevor Wing—Video Logic

Tiburon, USA Jon Peddie

Contents

List of Figures

List of Tables

Chapter 1
Introduction

By the mid-nineties, the forerunners of the fully integrated single-chip GPU were appearing with a range of functionality. What companies built often reflected their origin of their fonders, and target markets.

LSI graphics accelerator chips appeared in the early 1980s as 2D drawing engines, with CAD being the primary application. In 1981, SGI introduced the first 3D transform engine, the Geometry Engine [1]. The first company to offer a 3D AIB was Matrox in 1987 with the SM 640 that used the SGI Geometry Engine [2].

Semiconductor companies saw the need to decouple the geometry processing from the CPU and make it more tightly coupled with the pixel pipeline. Several companies developed floating-point processors (FPUs) to do the job.

Weitek introduced its floating-point processor (FFP) in the early 1980s and was successful with it as a coprocessor to the CPU. Some groups tried using it for geometry processing. The Pixel Planes system used it and it was used in the Sun workstation, and even by SGI. However, it never was used in any PC graphics AIBs. It was not a geometry processor per se but a general-purpose floating-point coprocessor. In the early 1990s, the company introduced a VGA clone, the Power 9100. The company did not do well, and in late 1996, Rockwell's Semiconductor Systems bought the assets.

In May 1994, 3Dlabs announced its first consumer 3D graphics chip, Gigi (which stood for *Game Glint*). It was a scaled-down version of the GLINT 300TX. Creative Labs built an AIB with it, the 3D Blaster VLB, in early 1995. However, Windows did not have an API that exposed T&L functions in the graphics hardware. Therefore, 3Dlabs chips came with 3Dlabs drivers for Windows 95/NT (accelerated 2D), accelerated D3D drivers, and accelerated OpenGL drivers. The 3D drivers would use whatever acceleration was enabled by the API and available on the current hardware.

The first consumer grade Geometry Engine chip was the Fujitsu (MB86242) Pinolite FXG-1, revealed at the Hot Chips conference at Stanford University in 1995 [3] and formally announced in July 1997 [4]. Rendition was the first company to use (and helped Hercules develop) an AIB with it and Rendition's 3D chip in 1998. However, the AIB's release got canceled because 3Dfx came out with a faster product.

J. Peddie, *The History of the GPU – Eras and Environment*,
https://doi.org/10.1007/978-3-031-13581-1_1

Also, DirectX did not expose the geometry engine to applications so custom drivers had to be written.

Also, in 1997, 3Dlabs announced their Glint Gamma (G1) stand-alone geometry processor for the professional graphics market. Then they adapted it to a Creative Labs consumer AIB in 1988.

Being first is always a tricky thing to identify. First to announce or first to ship? Or, if you could find out, first start a design? However, the one thing that can be said is that 3D consumer chips appeared in 1995.

Those early 3D chips did not include hardware transform and lighting (T&L) capabilities. The CPU or a separate coprocessor was used for those operations. But Moore's law was on the march, and it would be only a few years until the first fully integrated single-chip graphics processing unit (GPU) was introduced, marking the beginning of the GPU era.

1.1 Nvidia's NV10—First Integrated PC GPU (September 1999)

The term *GPU* has been in use since at least the 1980s. Nvidia popularized it in 1999 by marketing the GeForce 256 AIB as "the world's first GPU." It offered integrated transform, lighting, triangle setup/clipping, and rendering engines as a single-chip processor.

It had all the features for a truly integrated GPU.

Very large-scale integrated circuitry (VLSI) started taking hold in the early 1990s. As the number of transistors that engineers could incorporate on a single chip increased almost exponentially, the number of functions in the CPU and the graphics processor increased. Among the biggest consumers of the CPU were graphics transformation compute elements into graphics processors. Architects from various graphics chip companies decided that transform and lighting (T&L) was a function that should be in the graphics processor. A T&L engine is a vertex shader and a geometry translator—many names for the little FFP.

In 1997, 3Dlabs (in the UK) developed its Glint Gamma processor, the first programmable transform and lighting engine, as part of its Glint workstation graphics chips. They introduced the term *GPU*—geometry processor unit. 3Dlabs' GPU was a separate chip named Delta and was known as the DMX. 3Dlabs' GMX was a coprocessor to the Glint rasterizer.

Then in September 1999, Nvidia introduced the NV10 GPU with an integrated T&L engine for their consumer graphics chip. ATI quickly followed with their Radeon graphics chip and called it a *VPU*—visual processing unit. But Nvidia popularized the term GPU. It has forever since been associated with the GPU and credited with inventing it (Fig. 1.1).

Built on TSMC's 220 nm process, the 120 MHz NV10 had 23 million (incorrectly listed as 17 on some websites) transistors in a 139 mm^2 die and used DirectX 7.0.

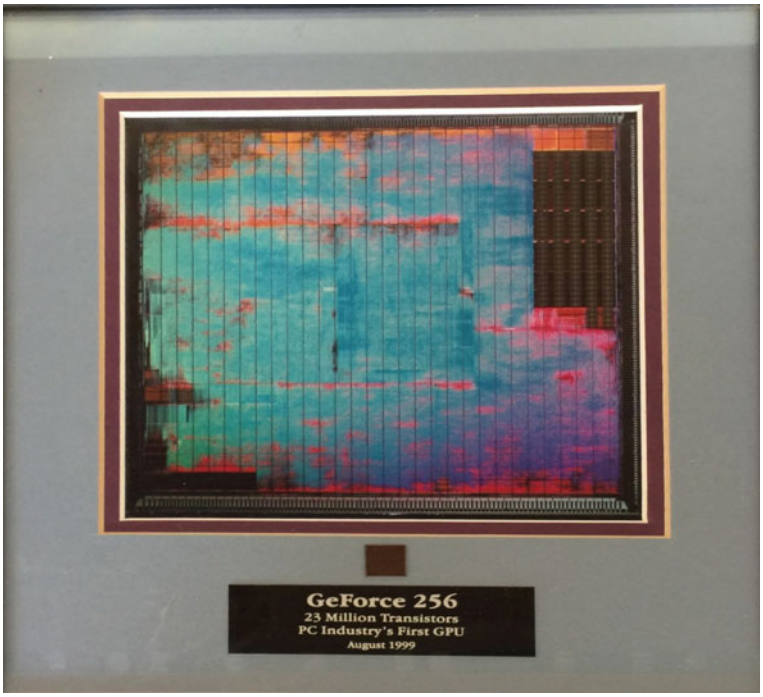

Fig. 1.1 Die shot of Nvidia's first GPU, the NV10. Reproduced with permission from Curtis Priem

The GeForce 256 AIB employed the NV10 with SDR memory. Refer to the block diagram in Fig. 1.2.

The chip had many advanced features, including four independent pipelined engines that ran at 120 MHz. That allowed the GPU to produce a 480 Mpix/s fill rate.

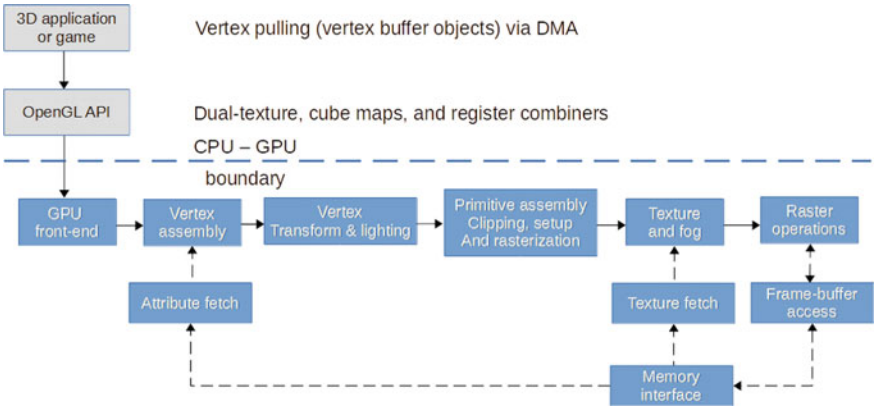

Fig. 1.2 Nvidia's GeForce 256 (NV10) block diagram with OpenGL