1

Introduction

1.1 Mobile 3D Graphics

Mobile devices are leading the second revolution in the computer graphics arena, especially with regard to 3D graphics. The first revolution came with personal computers (PC), and computer graphics have been growing in sophistication since the 1960s. To begin with it was widely used for science and engineering simulations, special effects in movies, and so on, but it was implemented only on specialized graphics workstations. From the late 1980s, as PCs became more widely available, various applications were developed for them and computer graphics moved on to normal PCs – from specific-purpose to normal usage.

Three-dimensional graphics are desirable because they can generate realistic images, create great effects on games, and enable slick effects for user interfaces. So 3D graphics applications have been growing very quickly. Almost all games now use 3D graphics to generate images, and the latest operating systems – such as Windows 7 and OS X – use 3D graphics for attractive user interfaces. This strongly drives the development of 3D graphics hardware. The 3D graphics processing unit (GPU) has been evolving from a fixed-function unit to a massively powerful computing machine and it is becoming a common component of desktop and laptop computers.

A similar revolution is happening right now with mobile devices. The International Telecommunications Union (ITU) reports that 3.3 billion people – half the world's population – used mobile phones in 2008, and Nokia expects that there will be more than 4 billion mobile phone users (more than double the number of personal computers) in the world by 2010 [1]. In addition, mobile devices have been dramatically improved from simple devices to powerful multimedia devices; a typical specification is 24-bit color WVGA (800×480) display screen, more than 1 GOPS (giga-operations per second) computing power, and dedicated multimedia processors including an image signal processor (ISP), video codec and graphics accelerator.

Mobile 3D Graphics SoC: From Algorithm to Chip Jeong-Ho Woo, Ju-Ho Sohn, Byeong-Gyu Nam and Hoi-Jun Yoo

So 3D graphics is no longer a guest on mobile devices. A low-cost software-based implementation is used widely in low-end mobile phones for user interfaces or simple games, while a high-end dedicated GPU-based implementation brings PC games to the mobile device.

Nowadays, 3D graphics are becoming key to the mobile device experience. With the help of 3D graphics, mobile devices have been evolving with fruitful applications ranging from simple personal information management (PIM) systems (managing schedules, writing memos, and sending e-mails or messages), to listening to music, playing back videos, and playing games. Just as with the earlier revolution in the PC arena, 3D graphics can make mobile phone applications richer and more attractive – this is the reason why I have used the phrase "second revolution."

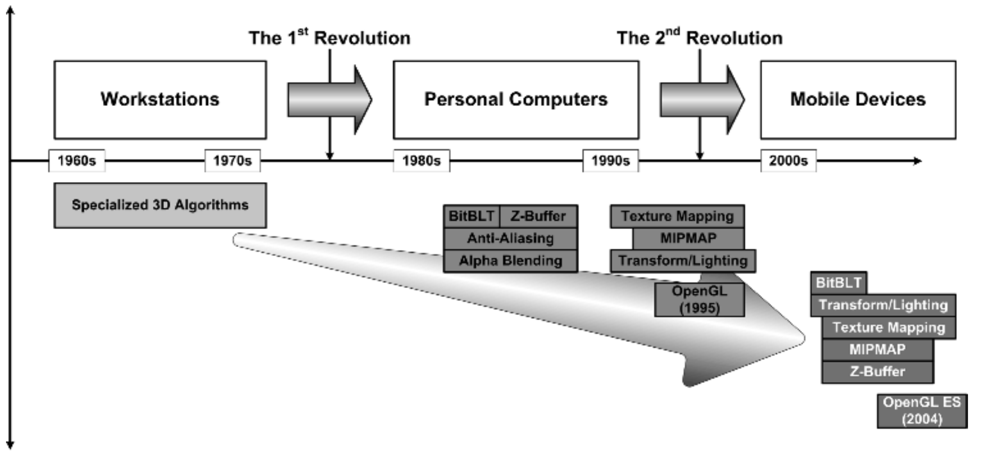


Figure 1.1 A history of 3D graphics

Development of mobile 3D graphics was started basically in the late 1990s (Figure 1.1). Low-power GPU hardware architectures were developed, and the software algorithms of PCs and workstations were modified for mobile devices. Software engines initially drove the market. Among them, two notable solutions – "Fathammer's X Forge" engine and "J-phone's Micro Capsule" – were embedded in Nokia cellular phones and J-phone cellular phones. Those software solutions do provide simple 3D games and avatars, but the graphics performance is limited by the computation power of mobile devices. So new hardware solutions arrived to the market. ATI and nVidia introduced "Imageon" and "GoForce" using their knowledge of the PC market. Besides the traditional GPU vendors like nVidia and ATI, lots of challengers introduced great innovations (Figure 1.2). Imagination Technology's MBX/SGX employs tile-based rendering (discussed in Chapter 5) to reduce data transactions between GPU and memory. Although tile-based rendering is not widely

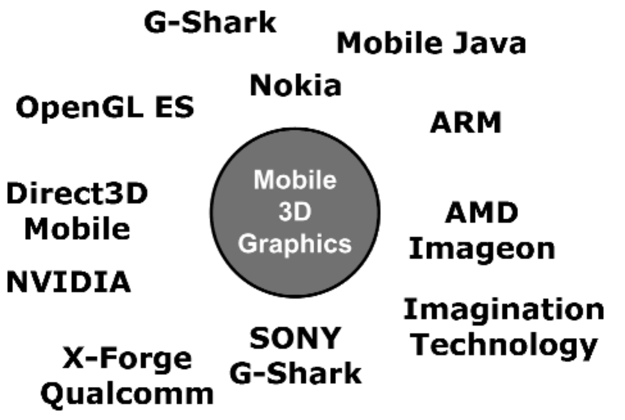


Figure 1.2 Mobile 3D graphics

used on the PC platform, it is very useful in reducing power consumption so that the MBX/SGX has become one of the major mobile GPUs on the market. FalanX and Bitboys developed their own architectures – FalanX Mali and Bitboys Acceleon – and they provided good graphics performance with low power consumption. Although those companies merged into ARM and AMD, respectively, their architectures are still used to develop mobile GPUs in ARM and AMD.

1.2 Mobile Devices and Design Challenges

As mentioned in the previous section, mobile devices have evolved at a rapid pace. To satisfy various user requirements there are lots of types of mobile device, such as personal digital assistant (PDA), mobile navigator, personal multimedia player (PMP), and cellular phone. According to their physical dimensions or multimedia functionality, these various devices can be categorized into several groups, but their system configurations are very similar. Figure 1.3 shows two leading-edge mobile devices and their system block diagram. Recent high-performance mobile devices consist of host processor, system memories (DRAM and Flash memory), an application processor for multimedia processing, and display control. Low-end devices do not have a dedicated application processor, to reduce hardware cost. Evolution of the embedded processor and display devices has led to recent exciting mobile computing.

1.2.1 Mobile Computing Power

In line with Moore's law [2], the embedded processors of mobile devices have been developing from simple microcontroller to multi-core processors and the computing

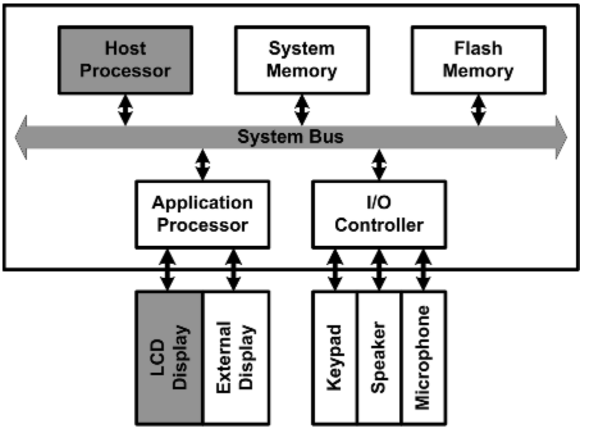


Figure 1.3 Mobile devices and their configuration

power has kept increasing roughly 50% per year. To reduce power consumption, an embedded processor employs RISC (Reduced Instruction Set Computer) architecture, and the computing power already exceeds that of the early Intel Pentium processors.

Typically, recent mobile devices have one or two processors as shown in Figure 1.4. Low-end devices have a single processor so that multimedia applications are implemented in software, while high-end devices have two processors, one for real-time operations and the other for dedicated multimedia operations. The host processor performs fundamental operations such as running the operating system, and personal information management (PIM). Meanwhile the application processor is in charge of

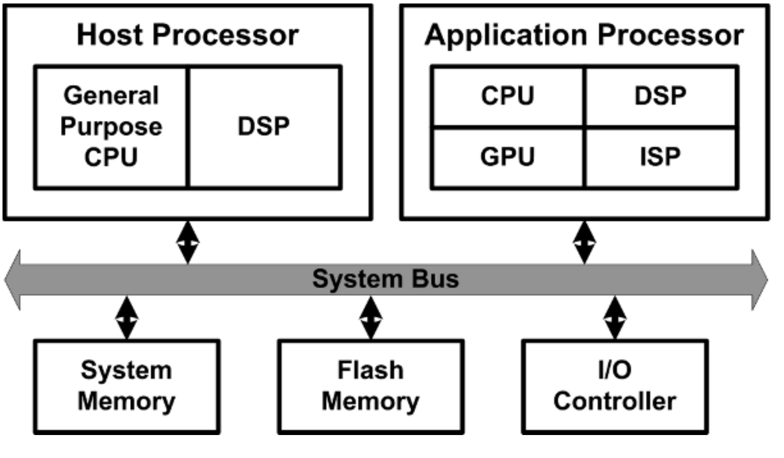


Figure 1.4 Embedded processors and system architecture

high-performance multimedia operations such as MPEG4/H.264 video encoding or real-time 3D graphics. To increase computing power, the newest processors employ multi-core architecture. Some high-performance processors contain both a general-purpose CPU and DSP together, and some application processors consist of more than four processing elements to handle various multimedia operations such as video decoding and 3D graphics processing.

1.2.2 Mobile Display Devices

It is safe to say that evolution of mobile display devices leads the revolution of mobile devices, especially the multimedia type. The first mobile devices had a tiny monotone display that could cope with several numbers or characters. Recent mobile devices support up to VGA (640×480) 24-bit true-color display. The material of the display device is also changing from liquid crystal to AMOLED (Active Mode Organic Light Emitting Diode). The notable advantages of AMPLED are fast response time (about 100 times faster than LCD), and low power consumption. Since it does not require back-lighting like the LCD, the power consumption and weight are reduced, and the thickness is roughly one-third of the LCD. Of course the functionality of the display device is improved too, so that nowadays we can use touch-screens on mobile devices.

1.2.3 Design Challenges

Although the funtionality of mobile devices is greatly improved, there are many design challenges in component design. In short, there are three major challenges.

Physical dimension – The main limitation of mobile devices is definitely their physical size. For portability the principal physical dimension is limited to about 5 inches (12.5 cm), and the latest high-end cellular phones do not exceed 4 inches. That means there is limited footprint on the system board, and components should be designed with small footprint.

Power consumption – Since the mobile device runs on a battery, the power consumption decides the available operating time. As the performance increases it consumes more power owing to the faster clock frequency or richer hardware blocks. Therefore, increasing operating time by reducing power consumption is as important as increasing computing power.

System resources – Mobile devices cannot have rich system resources owing to the physical dimension and power consumption. They cannot utilize a wide-width system bus and cannot use high-performance memory such as DDR2 or DDR3. Despite this, mobile devices provide quite high performance to satisfy user requirements.

To meet these design challenges, many mobile components are designed as SoC (System-on-a-Chip). Since the SoC includes various functional blocks such as processor, memory, and dedicated functional blocks in a single die, we can achieve high performance with low power consumption and small area.

1.3 Introduction to SoC Design

System-on-a-Chip has replaced key roles of VLSI (Very Large Scale Integration) and ULSI (Ultra Large Scale Integration) in mobile devices. The change of the name is a reflection of the shift of the main point from “chip” to “system.” You may wonder what “system” means and what the difference is compared with “chip.”

Before SoC, the hardware developer considered how to enhance the performance of the components. At that time, the hardware developer, the system developer and the software developer were separated and made their own domains. In the SoC era, those domains are merging. Engineers, be they a hardware engineer or a software engineer, have to consider both hardware issues and software issues and provide a system solution to the target problem with the end application in mind.

Of course, there are many different definitions of SoC according to the viewpoint, but in this book the *system* means “a set of components connected together to achieve a goal as a whole for the satisfaction of the user.” To satisfy end-user requirements, the engineer should cover various domains. With regard to the software aspect, the engineer should consider the software interface such as API or device driver, specific algorithms, and compatibility. With regard to the hardware aspect, the engineer should consider functional blocks, communication architecture to supply enough bandwidth to each functional block, memory architecture, and interface logics. Moreover, since such a complicated entity can be handled only by CAD (Computer Aided Design) tools, the engineer should have knowledge of CAD, which covers automatic synthesis of the physical layouts.

Therefore, the discipline of SoC design is intrinsically complicated and covers a variety of areas such as marketing, software, computing system and semiconductor IC design as described in Figure 1.5. SoC development requires expertise in IC technology, CAD, software, and algorithms, as well as management of extended teams and project and customer research.

Initially, the concept of SoC came from the PC bus system. By adopting the same bus architectures as those used in the PC, the processing of embedded applications was to be implemented on a single chip by assembling dedicated hard-wired logic and existing general-purpose processors. As the scale of integration and design complexity increased, the concepts of “design reuse” and “platform-based design” were born. The well-designed functional blocks could be reused in the later SoC.

However, such pre-designed functional blocks, called Intellectual Property (IP), are difficult to reuse with SoC because they were optimally developed for specific purposes, not for general-purpose utilization. In addition, since conventional buses were not suitable for the on-chip environment, there was a need to develop new

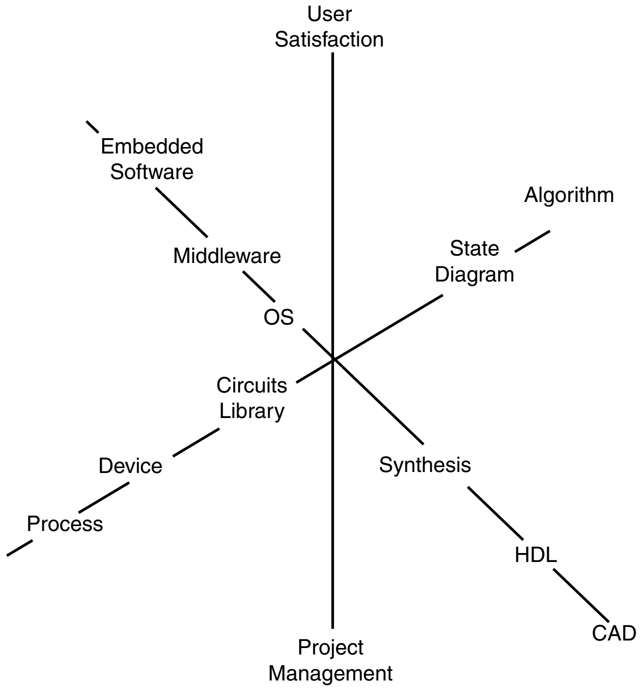


Figure 1.5 Disciplines required for the design of SoC

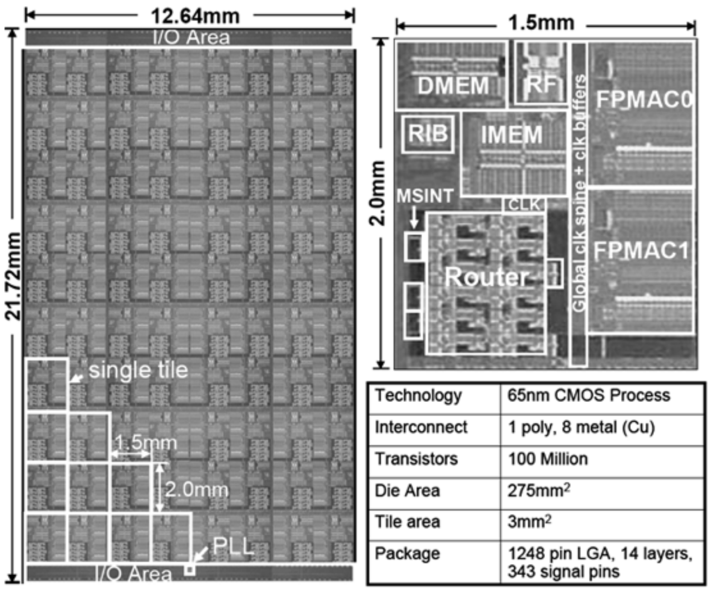
communication architecture with specific characteristics – such as wide bit width, low power, higher clock frequency, and a tailored interface. The details of design reuse and platform-based SoC design are discussed in Chapter 2.

Figure 1.6 shows an example of SoC. Intel's research chip [3] has 80 CPUs inside.

1.4 About this Book

This book describes design issues in mobile 3D graphics hardware. PC graphics hardware architecture with its shortcomings in the mobile environment is described, and several low-power techniques for mobile GPU and its real implementation are discussed.

Chapter 1 introduces the current mobile devices and mobile 3D graphics compared with desktop or arcade-type solutions. Chapter 2 discusses the general chip implementation issue, such as how to design the SoC, and includes an explanation of SoC platforms. The SoC design paradigm, system architecture, and low-power SoC design are addressed in detail. Chapter 3 deals with basic 3D graphics, the fixed-function 3D graphics pipeline, the application-geometry rendering procedure, and the programmable 3D graphics pipeline. In Chapter 4 we articulate the differences between conventional and mobile 3D graphics, and introduce the principles of mobile 3D graphics and standard mobile 3D graphics APIs.



Technology	65nm CMOS Process
Interconnect	1 poly, 8 metal (Cu)
Transistors	100 Million
Die Area	275mm^2
Tile area	3mm^2
Package	1248 pin LGA, 14 layers, 343 signal pins

Figure 1.6 Example of an SoC implementation: Intel's 80-core processor and its unit CPU. The Intel logo is a registered trademark of Intel Corporation

The design of 3D graphic processors is discussed in Chapters 5–7. Chapter 5 explains the hardware design techniques for mobile 3D graphics, such as low-power rasterizer, low-power texture unit, and several hardware schemes for low-power shaders. Chapter 6 covers the real chip implementation of mobile 3D graphics hardware. For academic architecture, KAIST RAMP architecture is introduced and the industrial architectures, SONY PSP and Imagination Technology SGX, are also described. Chapter 7 has a detailed explanation of the low-power rasterization unit with RTL code. In this chapter, readers can grasp the basic concept of how to design low-power 3D graphics processors. The future of mobile 3D graphics is very promising because people will carry more and more portable equipment in the future with high-performance displays. Finally, Chapter 8 looks at the future of mobile 3D graphics.

We also include appendices to introduce to chip design by verilog HDL The reader can run the verilog file to check the algorithms explained in the earlier chapters and get a taste of real 3D graphics chip design.

References

1 Tolga Capin, et, al., "The State of the Art in Mobile Graphics Research", IEEE Computer Graphics and Applications, Vol. 28, Issue 4, 2008, pp. 74–84.
2 Gordon E. Moore, "Cramming more components onto integrated circuits", Electronics, vol. 38, no. 8, 1965.
3 J. Held, et al, "From a Few Cores to Many: A Tera-scale Computing Research Overview," white paper, Intel Corporation, www.intel.com.