Adding games, graphics on handsets

By Ed Plowman
3D Graphics Products Manager
ARM Ltd

Asia's enthusiam for graphic-rich on-the-move gaming is set to stimulate a huge demand for next-generation mobile phones with sophisticated 3D graphics capability. Handset OEMs are identifying which features or applications will attract consumers to replace their mobile phones frequently and sustain this huge market at its current high growth rates. It will take a mixture of features and applications to motivate users to invest in a higher-performance handset.

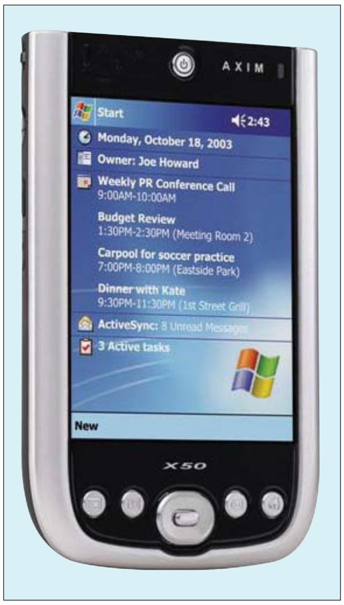
Fortunately, customers' penchant for playing electronic video games has caught the attention of mobile phone vendors. They believe that incorporating advanced 3D graphics facilities into handsets is key to tempting users to indulge in playing more games on their phones.

Only two years ago, very basic 2D bitmap graphics-enabled games on mobile phones were just additional features, like the ability to send text messages. The gaming feature became popular in Asian countries, and consumers were easily attracted to next-generation handsets with color screens. Handset manufacturers have responded with advanced graphics features, culminating in QVGA resolution screens with 16bit-per-pixel color and limited graphics acceleration, together with more complex and interactive games. Growth in higher-performance, graphics-oriented mobile phones in Asia is accelerating faster than anywhere else in the world.

The first step was to introduce 3D graphics facilities through software. This required a significant increase in processor performance, yet still had significant limitations for some operations. Highly complex interactive features are not yet viable. Nonetheless, user imagination had been captured.

In recent years, the first mobile devices with true, hardware-based 3D graphics facilities have emerged with many more versions in the pipeline. Demand in Asia, at least, is expected to explode.

The rest of the world, meanwhile, continues to question why mobile phones need color screens, let alone 2D/3D graphics on what is essentially a voice communication device. Advanced gaming on handsets is likely to catch on in Europe and the United States over the next two years. There is, however, a growing group of advanced users in the West that is interested in features such as connecting to the Internet, wireless networking, GPS services and applications such as built-in camera, multimedia messaging, mobile video and MP3 playback, as well as games. Such smart phones are foreseen to speedily evolve and proliferate.



3D graphics capability is essential to other display-related features such as video processing.

Non-trivial technology

Meanwhile, graphics technology within the dedicated video games sector has also leapt forward. For such devices, particularly console-based machines, there is considerably less pressure on squeezing huge amounts of processing power and dedicated games engines into a small space. Power consumption remains an issue for portable game devices, though they often have the space to incorporate additional hardware, such as sophisticated 3D graphics chipsets and accelerators. Dual screens, touchscreens, interactivity and the ability to handle large numbers of pixels are the challenges facing the likes of Nintendo and Sony.

To incorporate anything close to this performance in a mobile phone requires a complete rethinking on the architecture. A battery life of about six hours for a games console is more or less acceptable. For a mobile phone, even twice as long would be a complete non-starter. Hence, the majority of handset designers are taking the route where additional processing power, operating independently of the core CPU, is essential to minimize power consumption.

Yet, the pressure on space inside a mobile phone is intense and extra functionality has to be added with the minimum of silicon overhead. Even with the first graphics screens, the task of handling and shifting pixels required some sort of co-processor or dedicated DMA engine that understood pixel space. Today, there is a considerable range of DSP-based chips and cores, graphics accelerators, dedicated co-processors, graphics engines and software solutions available. How they are implemented is critical. Architectural, hardware/software partitioning and SoC or chipset decisions are made early in the design process and are crucial. The manufacturing cost can only be amortized against a large volume market. Late design changes can be devastating in terms of cost, while project delays can risk missing significant market (and, therefore, profit) windows.

Graphics is not an add-in feature but an integral part of all display-related activities. Integrating and verifying such complex electronics takes a huge amount of engineering effort. The emergence of applications-specific platforms has become essential to developers, facilitating design reuse and saving considerable design time and verification effort. 3D graphics functionality is most successfully implemented as an application subsystem, especially with the use of dedicated accelerators, which perform functions that would be inefficient to implement via the general-purpose processing platform.

Industry foresight

The mobile phone industry needs the full support of game developers to ensure there is a wide choice of affordable games available. One potentially thorny issue is the API required to communicate between the processing hardware, software and graphics engine. With foresight, the mobile phone industry has avoided the early experiences of the PC 3D games sector, where a variety of proprietary APIs created unnecessary competition and conflict, and, some would argue, effectively limited the rate of the overall market growth. The major handset OEMs, chip vendors, graphics engine and software suppliers, game developers and infrastructure companies were quick in identifying the potential problem and have cooperated and collaborated extensively to ensure a common API was agreed on from the start.

The emergence of Open GL-ES as an industry standard API for mobile 3D graphics is set to clear the way for handset manufacturers to compete in a market unrestricted by incompatible games software formats. Open GL-ES is derived from the desktop market and, being particularly easy to scale, has been stripped down to remove legacy and other unnecessary features, so that it can be implemented in a much smaller footprint.

Yet, an understandable cautiousness has crept into the mobile phone sector's thinking. While it is banking on 3D graphics–or, in truth, the games the technology will enable–to drive the future phone-replacement market, it is by no means guaranteed, despite the fact that it was games that drove 3D graphics into the PC and then the laptop. After all, gaming is still only an additional feature of what is, essentially, a voice communication device.

There are several of these 1G mobiles with 3D graphics features now emerging based on many architectures. Some can claim slightly better power consumption (and, therefore, longer battery life); others are promoting better visual quality or specific feature performance. Overall, however, there will not be considerable differences in terms of performance or feature sets among them. Until it is confirmed by consumers that console-quality 3D graphics is not only a differentiating but a driving factor for handset replacement, manufacturers remain wary of overcommitting, while trying to shift consumer perception from a "nice to have" to a "must have" feature. This first generation, though, will identify the minimum technical requirements for entry into this market and firmly establish Open GL-ES as the API standard.

Open GL-ES has widespread industry support not only from handset manufacturers and device and core vendors, but also from graphics engine specialists, games developers and OS suppliers. While Open GL-ES is a low-level API, it is wholly complementary to the higher-level JSR 184 API for J2ME applications in the Java gaming environment.

Evolving standard

The Khronos Group is a member-funded consortium focused on the creation of royalty-free, cross-platform open-standard graphics APIs to enable the authoring and playback of dynamic media on a wide variety of platforms and devices. All Khronos members may contribute to the development of Khronos API specifications, vote at various stages before public deployment and are able to accelerate delivery of cutting-edge 3D platforms and applications through early access to specification drafts and conformance tests.

Khronos API specifications include OpenML for capturing, transporting, processing, displaying and synchronizing digital media–including 2D/3D graphics and A/V streams; OpenVG, which provides a low-level hardware acceleration interface for vector graphics libraries such as Flash and SVG; OpenMAX, which standardizes access to media-processing primitives used extensively in graphics, audio and image libraries and video codecs such as MPEG-4; and Open GL-ES.

Open GL-ES is a royalty-free, cross-platform API for full-function 2D and 3D graphics on embedded systems, including handheld devices, appliances and vehicles. It is a well-defined subset of desktop OpenGL, creating a flexible and powerful low-level interface between software and graphics acceleration. OpenGL-ES 1.0 includes Common and Common-Lite profiles for floating-point and fixed-point systems and the EGL specification for portably binding to native windowing systems. Development of OpenGL-ES 2.0 is underway.

Meanwhile, a working group has been established that will help the standard migrate to other mobile platforms using 3D graphics, especially in the automotive and avionics sectors.

Tech arms race

With the second-generation 3D graphics-capable handsets, the "arms race" will start in earnest. Manufacturers will be far more competitive in terms of pure technological performance, especially as there will be little fragmentation once the API standard is fully established. There is some concern that handset OEMs should hold back from expanding their platforms too fast and creating proprietary extensions. The industry must ensure that the Open GL-ES standard API expands with the market.

In fact, the Open GL-ES road map is already in place, and work has commenced on Open GL-ES 2.0. Currently, the API is based on a fixed-function pipeline where features are enabled or disabled depending on the current render context. It allows OEMs to differentiate in terms of throughput, number of textures and similar features.

With Open GL-ES 2.0, certain elements of the graphics pipeline can be programmable, allowing the content developer to define precisely how the vertex or pixel is processed. This enables the use of procedural algorithms so that code performance will be dependent on implementation, essentially giving more scope for innovation and greater differences between vendors in terms of feature sets and performance–particularly, with respect to visual quality and effects, but still maintaining a common platform for the developer.

Graphics subsystem developers will have to start thinking differently for the mobile phone market, compared to the desktop or console market, where larger screens are used. The pixel density of the mobile phone's 2.2-inch screen is much higher than that of a laptop. Emerging 2.5-inch phone displays are likely to have 400 pixels per inch, doubling current densities. Thus, the conventional route to high-resolution graphics using high volumes of polygons and fast polygon rates is inappropriate. Smart designers will be looking at higher-order techniques for lighting and texturing, but based on coarser models. The differentiators will be pixel quality and the number of high-end processes that can be applied to those pixels, measured as operations per vertex per second and operations per pixel per second. Through this route, developers aim to provide the same scene detail effects that previously could only be achieved through millions of polygons, yet they must remain within the constraints of processor performance, memory capacity and power consumption. You need to look no further than existing titles like Farcry on the PC to see these techniques in action.

Already, the industry's best minds are working on Open GL-ES 2.0. These are engineers that understand the market and the limitations of the hardware, and they will not underestimate the resource required to develop both hardware and software for future applications.

2G handsets could move into the design phase later this year, with the first products launching just over a year from now, though much will depend on market acceptance of the first-generation models. By the time these devices become widely available, it is anticipated that Open GL-ES 2.0 will publish, ready for development of 3G mobile devices with 3D graphics capability.

Compact solutions

One of the most popular 3D graphics solutions available today for mobile device platforms is the PowerVR MBX graphics accelerator range available from ARM. Jointly developed with Imagination Technologies, the core is used in an application processor to provide PC- and games-console-quality 3D graphics in mobile phones, wireless gaming and other embedded devices.

Available in two versions, MBX R-S and MBX HR-S, the graphics accelerators offer QVGA and SVGA graphics performance, respectively. The MBX R-S version is a smaller die, primarily aimed at smaller wireless devices. The MBX HR-S is a higher-performance core. Both cores have an option to add a vertex geometry processor to offload geometry tasks from the main CPU and enable advanced lighting features.

Supporting a full 2D/3D feature set and offering Open GL-ES compatibility, the PowerVR MBX solutions use tile-based rendering to keep all bandwidth-intensive pixel processing on-chip. Memory bandwidth requirements are minimized by eliminating all Z-buffer traffic between the 3D core and external memory, and by ensuring that only visible pixels are ever written to the frame buffer. Tile-based rendering allows high-bit precision rendering and compositing even on systems with 16bit frame buffers, and also enables deferred texturing, which eliminates all redundant texturing operations. This approach facilitates the use of a unified memory architecture in which the graphics subsystem shares system SDRAM with the CPU, thus enabling a low-cost, space-saving SoC implementation.

MBX cores are easily integrated with the ARM926EJ-S and ARM1136J-S cores as well as the ARM PrimeXsys Platforms, enabling fast time-to-market. The core architecture is compact, saving space and silicon cost. Power consumption has been minimized using sophisticated power-management techniques, including module and register-level clock gating.

Additional integration support is provided through the availability of ARM's AMBA advanced high-performance bus interface and a range of ARM PrimeCell peripherals. These preverified, silicon-proven soft IP macrocells are easily integrated with the 3D graphics acceleration solution via the PrimeCell Peripheral multiport memory controller interface, optimized for use with both MBX R-S and HR-S.

Three-dimensional graphics capability is most likely to be one of the key driving forces behind stimulating handset replacement over the next five years. While gaming will be the principal application, 3D graphics capability will become essential to other display-related features, such as video processing. With the emergence of open standards, game developers will be encouraged to provide a rich supply of affordable games on the mobile phone platform.

The planned evolution of Open GL-ES provides a clear roadmap for the entire mobile 3D graphics sector, enabling steady growth and yet providing a healthy, competitive market. Within a few years, mobile 3D graphics technology will spread to other platforms and sectors, including automotive and avionics, especially as the open standards evolve. □