The POWER GXT2000P Graphics Accelerator

Introduction

The graphic accelerator industry is cluttered with high-end and entry-level solutions. High-end solutions typically offer the performance and features needed by today's workstation applications at a high cost while entry-level solutions typically lack in performance and/or features. Often, the optimal solution lies somewhere between these two extremes. This leaves system designers with very few choices for their midrange workstation.

The POWER GXT2000P graphics accelerator is targeted at the midrange workstation market where the customer is cost sensitive but still requires the performance of a high end graphics system. GXT2000P is an elegant solution, employing the industry's first single chip 256-bit raster engine and industry standard synchronous graphics dynamic random access memory (SGRAM) to provide the graphic features and performance required by the most demanding applications. Whether your business is mechanical computer-aided design (MCAD), mechanical computer-aided engineering (MCAE), Digital Content Creation (DCC) or Animation and Visualization, the GXT2000P is specifically designed to accelerate your applications at an affordable price. Not only does the GXT2000P accelerate your 3D applications, it is also a superb 2D accelerator. For the power-user on a budget, the GXT2000P is the right choice. When combined with the RS/6000® 43P Model 150, the Power GXT2000P technology sets a new standard for graphics price/performance in the workstation market.

GXT2000P Features Overview

The GXT2000P is a single, half-sized PCI adapter card specifically designed for the entry 3D marketplace. It provides double buffered 24-bit color at display resolutions up to 1920x1200 at 76 Hz. and includes support for monitors which comply with the ISO 9241 Part 3 standard. The GXT2000P supports a number of graphic application programming interfaces (APIs) including X11R6.1, OpenGL™ 1.2 and graPHIGS™.

Hardware features include:

- 24-bit double buffered color
- 24-bit depth (Z) buffer
- 8-bit alpha buffer
- 8-bit overlay buffer
- 4-bit stencil buffer
- 4-bit clip buffer
- 4-bit window identification buffer (WID)

Hardware accelerated functions include:

- Point, line, triangle, rectangle and quadrilaterals commands
- Gouraud shading
- Depth checking
- Anti-aliased lines
- Point, bilinear and trilinear mipmapped texture mapping
- Polygons rendered as points or lines
- Accurate line style support
- Poly-marker support
- Color processing and comparison
- YUV to RGB color space conversion
- Up and down scaling of video images

These features and the GXT2000P innovative implementation are described in the following sections.

GXT2000P Description

General Rendering Features

Probably the most important facet of a graphics accelerator is its rendering pipeline. The GXT2000P offers a fully integrated 2D/3D raster pipe. The only difference between a 2D and 3D rendering command is the number and types of parameters. Rendering a 2D object is accomplished by merely disabling various stages of the raster pipe.

In general, the graphics accelerator provides commands for rendering points, lines, triangles and quadrilaterals. Supplied with each of these commands is a set of data which defines the vertices of primitives (i.e. Point, line, triangle or quadrilateral). Information contained in each vertex along with the state of the raster pipe determines if the primitive will be drawn as a 2D or a 3D object.

2D Features

With a fully integrated 2D/3D raster pipe, most rendering functions are common for both 2D and 3D APIs. This section describes the few functions dedicated to the rendering of 2D objects.

A packed 16-bit integer XY coordinate native to the X11 API is supported on the GXT2000P in addition to a fixed point XY coordinate required by the 3D APIs. Supporting the 2D APIs native format in hardware eliminates the need for software to convert the coordinates as well as reducing the number of data items transferred to the accelerator.

Applying patterns to rectangle and line drawing is a widely used function of the 2D APIs. This function is fully accelerated in hardware by the GXT2000P. Two modes of styled or patterned lines are implemented by the accelerator to support the X11 and OpenGL as well as the graPHIGS APIs: count based and pattern based. The hardware supports a 32x32 pattern array which is applied to rectangles. The stipple pattern is an array of 1-bit values which select between

the foreground and background colors. When rendering a transparent object, the hardware will use the block write feature of the SGRAM for better performance whenever possible.

The GXT2000P contains a unique BLT mode expanding a one-bit-per-pixel image into any one of the pixel types supported. The image can be expanded to one or two colors. When an image is expanded to one color (i.e. Transparent), the hardware attempts to use the block write feature of the SGRAM. This function is used to accelerate the rendering of text.

In addition to the features already mentioned, the GXT2000P contains many of the pixel processing features required by the 2D APIs. These include:

- 256 logical operations
- transparent color key for BLTs
- the ability to selectively update planes of a pixel (i.e. Plane Write Mask).

3D Features

Three dimensional (3D) objects are basically two dimensional (2D) objects with various rendering techniques such as Gouraud shading, depth checking and texture mapping applied to give the object a sense of depth. The GXT2000P supports many of these techniques in hardware.

Gouraud shading is a technique of linearly interpolating the color across the surface of a primitive. Shading smoothes the transitions between the primitives, such as triangles, which are used to create the surfaces of a 3D object. Without shading, 3D objects would appear to have discontinuities in the surface at primitive boundaries. Software algorithms also utilize smooth shading to illuminate 3D objects in a scene.

Depth checking is a technique used to determine the visible pixels of 3D objects. The GXT2000P supports a 24-bit Z-buffer and the necessary pixel processing to eliminate the non-visible surfaces.

Texture mapping is a technique of adding surface detail to a 3D object. The GXT2000P supports point, bilinear, and trilinear mipmapped filtering of one and two dimensional textures. True color, luminance-alpha (4- or 8-bits per band) and 8-bit luminance textures are supported in hardware. The GXT2000P supports a wide range of texture blending functions as well as the addition of a specular color component for producing hi-lites. The standard OpenGL clamp and wrap modes are supported by the GXT2000P for textures with borders.

In addition to the features listed above, the GXT2000P provides anti-aliased line and rendering of polygons as points or lines to accelerate wire frame applications. The GXT2000P also provides unique hardware functions for improving the performance of graPHIGS applications. For example, the hardware supports the graPHIGS definition of an “accurate” line style, poly-markers, special color processing and comparison functions.

With all these features supported in hardware, the GXT2000P is capable of accelerating the rendering of realistic 3D scenes.

Video Features

In addition to all the features for rendering graphics, the GXT2000P also incorporates a video engine for supporting video applications. The video engine supports up and down scaling of RGB and YUV images. This function improves the GXT2000P's ability to support video applications such as video conferencing and collaboration.

System Management Features

In addition to accelerating graphics primitives, the GXT2000P contains many features help manage the graphics system. The section describes some of these features. The primary benefits of these system management features are to enhance the user interactivity of the graphics system and to improve the performance of the API.

Most three dimensional (3D) applications run under a two dimensional (2D) graphical user interface (GUI). The GXT2000P graphics accelerator contains special instructions for saving and restoring the adapter's state to allow for rapid swaps between these two environments. These instructions move the state information between the graphic accelerator and system memory using Direct Memory Access (DMA). Providing such features improves the interactivity of the graphic system.

DMA is also exploited by the APIs to send lists of commands (or Command Queues) to the GXT2000P graphics accelerator and to move pixel or video data between system memory and the accelerator. The Command Queues de-couple the graphics rendering from applications. This allows the application to prepare work for the graphics sub-system while the accelerator is processing the previous instructions. DMA is also used to inform the system software of an event by writing status to a pre-specfied address. Events can range from the completion of a Command Queue operation to a change in the state of the accelerator.

Many 3D applications require double buffering. Double buffering is a technique for creating the effect of smooth motion by drawing the next frame into a back buffer (off screen) and swapping the display to this buffer when the frame is complete. To improve the performance of buffer swaps, the GXT2000P provides a synchronization command which notifies the system software when a scene is completely rendered. Once notified of the completed frame, software can schedule a buffer swap with the accelerator. The GXT2000P also provides a mechanism to perform the buffer swaps during the vertical blank interval of the monitor.

Window Management Features

Window management is something many adapters overlook since it usually does not affect the performance benchmarks. The GXT2000P provides a number of features specifically for managing windows on the display surface. For example, four sets of rectangular clippers are provided in hardware to assist the 2D GUI in managing overlapping windows. In addition, the accelerator provides per pixel clipping for non-rectangular windows and for window geometries which can not be described using the hardware clippers. In addition, a window offset register is provided to allow applications to render in a coordinate system relative to their window on the

screen. The window offset register is maintained by the window manager and defines the position of the application's window on the screen.

The GXT2000P graphics accelerator supports a 4-bit Window Identification (WID) buffer. The WID buffer is stored in the unified frame buffer memory along with the other pixel information provides the 2D GUI with Window Identification (WID) planes. These WIDs provide the support necessary to allow multiple pixel types to exist on the screen simultaneously. Some attributes which make-up a pixel's type are: the source buffer (i.e. front or back), pixel format (i.e. 8-bit, 24-bit etc.) and color palette.

Management of the screen often involves filling large areas of memory with constant data. To improve the performance of such operations, the GXT2000P graphics accelerator implements the block write feature of the SGRAM. Using this feature allows an area of the screen to be filled up to four times faster than when using the normal write operation.

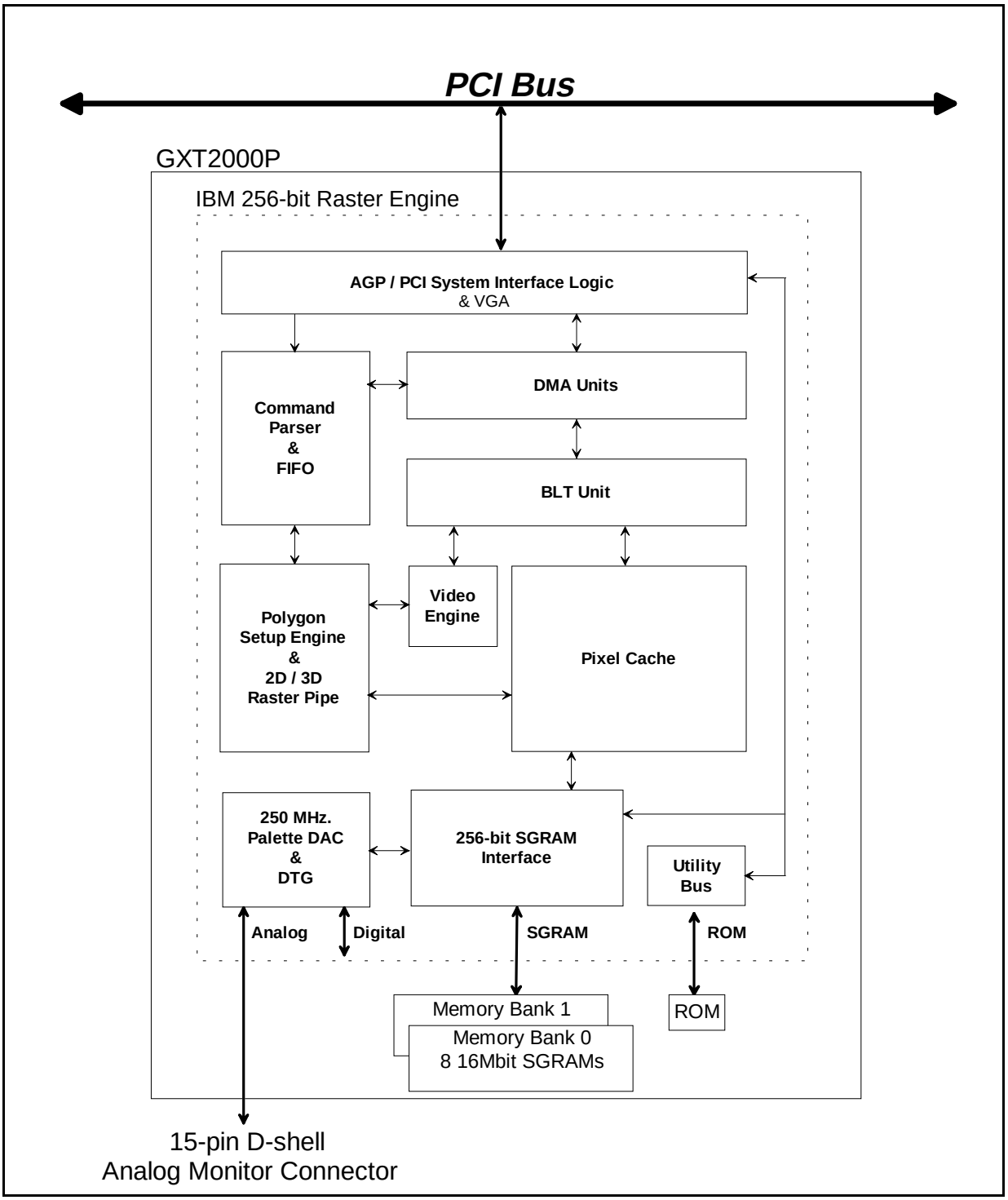
GXT2000P Hardware

The GXT2000P graphics accelerator contains the IBM® 256-bit Graphics Rasterizer and 32Mbytes of unified frame buffer memory. The accelerator is a 32-bit, 66MHz. PCI bus revision 2.1 compliant accelerator with a DDC-2B compliant monitor interface. The design of the GXT2000P was specifically targeted at the OpenGL and graPHIGS APIs.

The Raster Engine

The IBM 256-bit Graphics Rasterizer is the industry's first 256-bit raster engine. The raster engine design takes advantage of IBM's leadership in advanced integrated circuits technology to deliver a single chip raster engine with exceptional performance. The raster engine integrates the following units onto a single piece of silicon.

- AGP/PCI System Interface
- Command Parser
- 2 DMA Units
- BLT Unit
- Polygon Setup Engine
- 2D/3D Raster Pipe / Texture Engine
- VGA Unit
- Video Engine
- 250 MHz. / 30-bit Palette DAC
- Display Timing Generator
- 256-bit memory interface



The system interface combines standard AGP and PCI logic core functions available from IBM Microelectronics Division with custom logic to provide a 32-bit, 66MHz. interface and an AGP 2X version 2.0 compliant interface. The GXT2000P uses the PCI interface for connecting to the RS/6000 system.

The Command Parser is responsible for analyzing the command data stream supplied by the host system and notifying the other units of the function to perform. The data stream may be sent to the Command Parser directly or through Command Queues which are fetched by the DMA unit. In addition, the Command Parser saves and restores the state of the raster engine during context switches.

The two DMA units provide the raster engine with Command Queue data, context information, BLT data and Texture data. The units allow the raster engine to fetch Commands and BLT or Texture data without the need of arbitration or internal buffer flushing. The DMA units are designed to stream data between the accelerator and system memory. Both DMA units are designed to allow a scatter-gather technique for accessing noncontiguous memory locations.

The BLT unit converts between the linear system memory and the (X,Y) coordinate system of the screen. Part of the conversion is to remove unneeded data when extracting a smaller image from within a larger, linearly mapped image in system memory.

The Polygon Setup Engine accepts all the drawing commands from the parser and decomposes the commands into points, lines and triangles for the Raster Pipe. The Polygon Setup Process also performs all the calculations necessary for the Raster Pipe to render the primitive.

The Raster Pipe contains the interpolation logic and fragment processing for rasterizing the drawing primitives. The Raster Pipe is compatible with the OpenGL 1.2, graPHIGS and XWindows programming interfaces with virtually all the rasterization being performed by hardware. The Raster Pipe is optimized to run at the maximum frame buffer bandwidth. In addition, the Raster Pipe contains a Texture Engine which supports 1D and 2D texture mapping with point, linear, bilinear and trilinear mipmap filtering. The textures are stored in the unified frame buffer.

The Video Engine provides support for point and bilinear scaling of BLT data. The Video Engine supports both up and down scaling of RGB, YUV 4:2:0 and YUV 4:2:2 pixels. This engine is used to support applications such as video conferencing.

The Palette DAC contains a 1024 entry color lookup table and a separate 256 entry, 30-bit gama correction table. The palette can be used for mapping the color buffer data and the overlay data. The Palette DAC supports a wide range of pixel formats. These include: 8-bit Indexed, Direct and True Color; 16-bit Direct and True Color; 24-bit Direct and True Color. Selection of the pixel format is through a 16 entry Window Attribute Table (WAT). The WAT entry is selectable per pixel by the 4-bit Window ID (WID) stored in the frame buffer. The Palette DAC directly connects the monitor to the raster engine with three 10-bit monotonic Digital to Analog Converters (DACs) that are capable of up to 250 MHz operation. The Palette DAC also provides a cross-hair and a 64x64 sprite hardware cursor. In addition to the analog connection, the raster engine provides for a digital connection to a monitor. The GXT2000P graphics accelerator does not provide support for digital monitors.

The Display Timing Generator (DTG) provides the necessary controls for synchronizing the output of the DAC to the raster scan of the monitor. The DTG is compatible with the VESA™ monitor timing parameters.

The memory interface supports industry standard 16 Mbit synchronous graphics dynamic random access memory (SGRAM). The memory interface is 256-bit wide and operates at 100 MHz. The memory interface provides all the controls for reading and writing the frame buffer as well as keeping the memory refreshed. The memory interface supports the block write and write-per-bit features of the SGRAM. The memory interface also provides the Palette DAC with a continuous stream of WID, color and overlay buffer data.

The 256-bit data path to the frame buffer provided by the raster engine gives the GXT2000P the needed bandwidth to the frame buffer to support display resolution of up to 1920x1200 at 76Hz without crippling the rendering performance. The unified frame buffer design allows the memory to be dynamically allocated between the color buffers, Z-buffer and Texture memory which helps prevent wasting memory.

The Frame Buffer

The unified frame buffer consist of two banks of eight 16Mbit SGRAM memory for a total of 32 Mbytes. The peak bandwidth of the frame buffer is 3.2 Gbytes. The frame buffer is typically allocated to the following buffers:

- 24-bit Color, double buffered
- 8-bit Alpha
- 8-bit Overlay (4-bit double buffered)
- 24-bit Z-buffer
- 4-bit stencil
- 4-bit clip
- 4-bit Window ID

For a 1280x1024 screen, all the buffers listed above consume close to 16 Mbytes which leaves approximately 16 Mbytes for on card texture storage. For larger screen sizes, less texture memory will be available.

Technology

The IBM 256-bit Raster Engine is manufactured by IBM Mircoelectronics Division in their 5SE (0.27 micron) application specific integrated circuit (ASIC) technology. The raster engine design utilizes more than 6 million transistors. The silicon is packaged in an IBM flip-chip ceramic ball grid array (CBGA).

Development Process

The GXT2000P was developed by IBM's Visual Systems Group in Austin, Texas with assistance from groups around the world. The requirements for the graphics accelerator were created from

customers, end-users and API specialists. The main goal of the GXT2000P design was to meet the requirements of the X11, OpenGL and graPHIGS Graphical Display Interface (GDI) while keeping the cost in line with an entry level graphics solution.

The design group in Austin, Texas applied an innovative design methodology using a combination of industry standard and IBM proprietary tools. The design methodology has evolved over the past 12 years to incorporate both software and hardware emulation and an industry standard hardware description language (VHDL). Emulation allows the co-developement of software and hardware which has improved the development cycle of a graphic accelerator ASIC of this complexity.

Performance

While most graphics adapters are measured in terms of their fill rate and texture mapping performance, these are very poor metrics for the actual productivity which will be realized by the user. The best metrics available for performance comparisons are the API benchmarks and application performance measurements. The tables below illustrate the success of the GXT2000P. Systems with similar configurations and price were chosen from each vendor for comparision. The configuration and list price were taken from the Graphics Performance Characterization (GPC) Committee web page. The selected systems may not be the highest performing systems the vendor offers.

Unix System Benchmark Comparisons:

Benchmark[1]	System	IBM RS/6000 43P-260	IBM RS/6000 43P-150	HP B180L	HP C240	SGI Octane™ 250 MHz	Sun Ultra 60 model 1360
		Multi-processor : Uni-processor	Uni-processor	Uni-processor	Uni-processor	Uni-processor	Uni-processor
	Graphics	GXT2000P	GXT2000P	VISUALIZE fx2	VISUALIZE fx4	Maximum Impact	Elite3D m3
PLBwire93		731.7 : 448.1	258.8	289.3	319.6	n/a	285.8
PLBsurf93		801.9 : 561.5	477.2	424.1	578.6	n/a	496.4
ProCDRS-01		~17.6[2] : 17.6	11.1	5.98	17.8	10.4	n/a
CDRS-04		~188.9[2] : 188.9	103.7	n/a	n/a	74.1	76.72

1: Notes for Benchmark data are shown at the end of the paper.
2: Results are similar to uni-processor system. Exact multi-processor numbers may differ slightly.
n/a: Results are not available.

Summary

The POWER GXT2000P graphics accelerator provides IBM RS/6000 customers with a new level of graphics price/performance. One reason for the outstanding price/performance is the level of

integration provided by the IBM 256-bit Raster Engine. Integrating a full graphic adapter onto a single piece of silicon and utilizing industry standard SGRAM yields a half sized PCI adapter with the features and performance of many full sized and muti-card solutions.

The GXT2000P is designed to meet the performance and functional requirements of applications in the MCAD, MCAE, DCC and Animation and Visualization market segments.

In general, the POWER GXT2000P graphics accelerator delivers a superb balance of 2D and 3D graphics acceleration for the most demanding of applications. When combined with the RS/6000 43P Model 150, the GXT2000P offers the performance and features of high end graphics system at an affordable price.

Notes:

IBM performance figures were measured in a development-level system environment and are presented for illustrative purposes only. While these values should be indicative of the performance of generally available systems, this cannot be guaranteed. Other performance figures were taken from the following references:

- PLB benchmarks are from http://www.spec.org/gpc/plb/plb.summary.html on December1, 1998. These benchmarks are geometric means of literal and optimized Picture Level Benchmark (PLB) tests for 3D wireframe and 3D surface tests. The benchmark and tests were developed by the Graphics performance Characterization (GPC) Committee. The results shown used the graPHIGS API. Larger values indicate better performance.
- CDRS-04 benchmark is from http://www.spec.org/gpc/opc/opc.cdrs.summary.html on December 1, 1998. This benchmark is the weighted geometric mean of individual viewset metrics. The viewsets were developed by ISVs (Independent Software Vendors) with the assistance of OPC (OpenGL Performance Characterization) member companies. Larger values indicate better performance.

Biography

Charles Johns is a member of the POWER GXT2000P design team. The design team is part of the IBM Server Group, Austin, Texas.