

Display-Table of Contents

Topic	Page
Table of Contents	1
Introduction	2
Video Cards	2
Color	3
How Data is Represented	4
Monitors	4
CRT Monitor Basics	4
LCD Monitors Basics	5
CRT Monitors	5
Electron Gun and Deflection Yoke	6
The Shadow Mask and Aperture Grill	7
The Advantage of CRT Monitors	8
LCD Monitors	8
LCD Technology	9
Refresh Rate	10
LCD Advantages	10
Video Cards	10
Parts of a Typical Video Card	11
A. GPU	11
B. Video Card Memory	12
C. AGP Connector	12
D. Display Connector	12
How the Video Card Works in 2D	13
Memory Requirements	13-14
3D Rendering	14-15
Immediate Mode Rendering	15
Tile Based-Rendering	16
How it Works	16

Introduction

Before the era of PCs, Computers outputted information through printers, or sequences of LEDs being turned on or off to show the binary values of processed data. As computers became more powerful and user-friendly, monitors started to be used. The first monitors had only two colors (usually green and black). Today, computers can display more colors than our eyes can decipher, but they still have a long way to go before they are even close to perfect.

There are two major types of monitors in use today, CRT and LCD monitors. Both types work much differently than the other, and both types have their own advantages and disadvantages. For example, LCD monitors consume less power and produce crisper images, but they are also much more expensive, and produce far less brighter images than CRT monitors.

Video Cards

This display section will not only teach you about how monitors work, but also how display adapters (commonly known as video cards) work. Display adapters process the information and output it to the monitor to represent the picture you see on the screen. In the last few years, video cards have developed at a very quick pace. They can render life-like images, and give the PC the power it needs to display 3D images at even hundreds of FPS (Frames Per Second), all at a fraction of the price from twenty years ago.



Size does not matter: the largest Video Card in the world, but in no way the best one.

© The Computer Museum History Center, 1996

Although most of the Video Cards produce more or less the same result when rendering images or frames for computer games, the way they work depend on the type of Video Card. Essentially, they all work almost the same way when rendering 2D images, but when it comes to 3D images, which are most commonly found in Professional Animation Programs or Computer games, there are basically two ways in which they work.

How do the different monitors work? How do the different display adapters work? This is all covered in the display section.

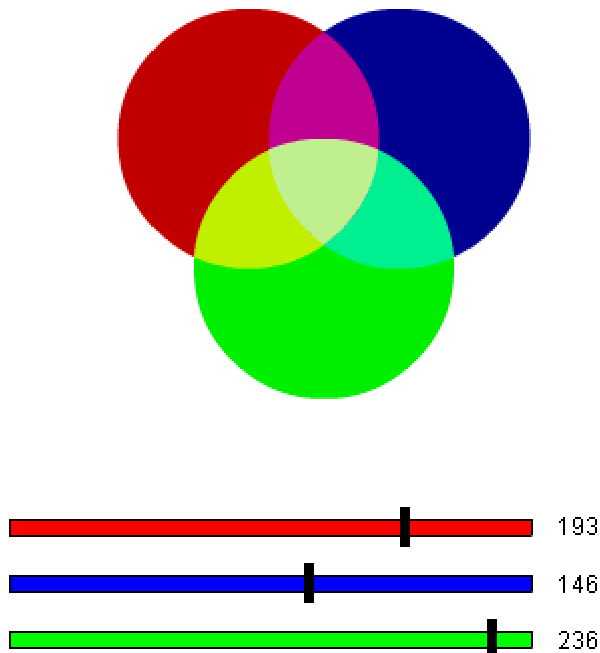
Color

At this time, even the cheapest new monitors can support 2^{24} (16,777,216) colors, which is many more colors than the human eye can decipher. How does the monitor make that many colors?

It is quite simple. The monitor is made of many small dots called pixels. Each pixel is assigned a different color. When many of these pixels are put together, they make an image on the screen that the user can look at. To give you an idea of just how many pixels there are on your monitor, most 15 inch monitors have about 800,000 pixels. Monitors which have a screen size of 17 inches have about 1.3 million pixels, while a typical 19 inch monitor has almost 2 million pixels!

Now back to the colors. Each pixel's color is determined by the intensity of three sources of light. These sources of light are three different colors. They include red, green and blue. Each of the three lights can have 2^8 or 256 different intensities of the color, going from black to the purest form of each of the colors (for the red light source, an intensity of 0 would be black, while intensity of 255 would be a pure red).

Since each of the three lights can produce 256 different colors, and there are three light sources, the monitor can produce 256^3 or 16,777,216 different colors.



As you can see, there are three light sources or masks. They overlap to make the final color. The color in the middle of all three masks represents the color that the user would see for the specific pixel. The colors in between any two masks represent the color that would be created if only these two masks were mixed together. Note that if you leave all the bars at 0, black is made, while all the bars at 255 would make white.

How Data is represented

Now that we know how the color is made by the monitor, it is time to learn how the monitor knows which colors to use. We know that each of the colors can have 256 different intensities (0 to 255). Since eight bits or one byte can also have 256 combinations, we can use one byte to represent the intensity of one of the three masks or sources of light. Since there are three lights, we need three bits or 24 bytes to give the monitor information as to which color each pixel would be. For example, if the monitor got data for a specific pixel which was 00000000 00000000 00000000, it would mean that the red, green and blue masks should have an intensity of 0, making the pixel a pure black. On the other hand, if the monitor got data for a specific pixel which was 11111111 11111111 11111111, it would mean that the red, green and blue masks should have an intensity of 255, making the pixel a pure white.

Monitors

Monitors are a very important part of today's PC. Can you think of anything that you could do with your computer without a monitor?. If you cannot think of anything, you are not the only one. Yet, believe it or not, early computers did not have one. Instead of displaying information on the screen, there would be a sequence of LEDs (Light Emitting Diodes). This is basically just a light that could be turned on or off. When the LED is on, this would symbolize a binary "one", while off, this would symbolize a binary "zero". The other way to get processed data from the computer would be through a printer. This method was faster than the LED method, but you would have to print an enormous amount of information, instead of reading it all from a monitor.

The LED and printer method of output limited what people could do with a computer. It made the computer very hard to operate and understand. Also, operations such as word processing, or even playing games, were impossible.

CRT Monitor Basics

When monitors came, it made it much easier for typical people to use computers. The most common type of monitor was one which included a CRT tube. It was very heavy, took up space, and consumed a lot of electricity, while only being able to produce two colors for us to see. Slowly, the 2-color limit changed into a 16-color limit, then to a 256-color limit. Now monitors produce more colors than our eye can decipher. Also, instead of focusing on the amount of colors the monitor can display, monitor producers are focusing on other features, such as cost, efficiency, image quality, monitor display size, flatness of the monitor display (the flatter the better, most CRTs are convex), weight, as well as volume.

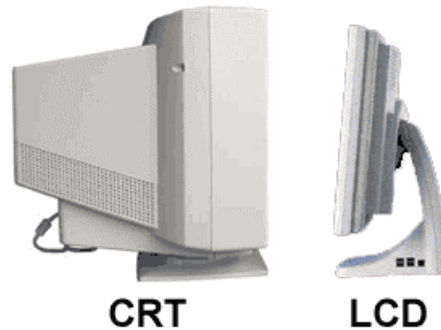


Although TVs work similarly to CRT Monitors, they produce less detail, and irritate the eyes more.

© www.sony.com, 2001

LCD Monitor Basics

While CRT monitors can be developed to have some of the features named above, there are new technologies that have fully taken care of some of these problems. One of these new technologies is the LCD Monitor. Some of its advantages are that it is very light, has a much smaller area than CRTs and is extremely efficient.



As you can see, LCD monitors are much smaller, and lighter.
© www.iiyama.com, 2001

LCD Monitors are not perfect though. They also have their own disadvantages, including the fact that they are 3-4 times more expensive than CRT Monitors. Both the CRT and the LCD Monitors are covered in depth later in the Display section, but first we need to know how both of the monitors produce the different colors for the user to see.

CRT Monitors

CRT Monitors are very common in the computer world. They are the most popular type, and the least expensive, while still delivering great performance and detail. CRT stands for Cathode Ray Tube. This is the most important part of the monitor, because it is what makes the image. The Cathode Ray Tube has three major parts inside it. These include the Electron Gun, the Deflection Yoke, as well as either the Shadow Mask or the Aperture Grill (depending on the type of monitor you have).



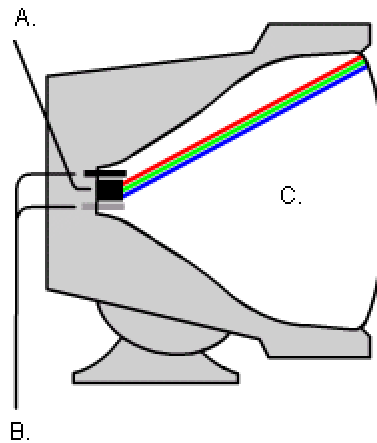
Viewsonic's E790 19" Monitor
© www.viewsonic.com, 2001
Modified by the TQ Team "Gates of Creation"

The Electron Gun and Deflection Yoke

Just like most TVs, CRT Monitors use an Electron Gun and Deflection Yoke. The Electron Gun has three electron streams used to represent three colors, which are red, green and blue. On a Monitor with a 24-bit depth, each of the streams can have 256 different intensities, which allows for more than 16 million color combinations. These three electron streams have to be fired at every single pixel on the monitor, one pixel at a time. In order to refresh the whole screen on a typical 19" monitor, the electron gun has to fire at almost two million pixels.

This is where the Deflection Yoke comes in. It is really an electromagnet that surrounds the Electron Gun. By activating a certain part of an electromagnet, the electron streams move closer to that part of the electromagnet, which lets them hit a different pixel. On a monitor with a resolution of 1600 X 1200, there are 1200 rows, and 1600 columns. The Deflection Yoke has to draw the electron streams to hit all 1600 pixels in one row, go to the next row down, and repeat the process. Once the whole screen refreshes, this has to be done again. The operation is done so fast that the human eye cannot even see the monitor refresh.

A typical CRT monitor refreshes at 60-75 Hz, which means that the whole screen refreshes 60-75 times in one second. Here is a demonstration which shows how a CRT monitor refreshes the screen using the Electron Gun and Deflection Yoke.



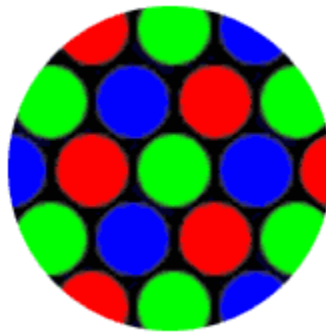
A. Electron Gun
B. Deflection Yoke
C. CRT Tube

The streams are constantly changing their intensities while refreshing the picture. If this did not happen, the whole screen would be one color. The top of the Deflection Yoke lowers its power to let the electron stream go down, and then the lower part of the Deflection Yoke intensifies its power to draw the electron stream to the bottom pixels. Now that we know how these two parts work, it is time to learn about the part of the CRT Monitor that actually lets us see the colors.

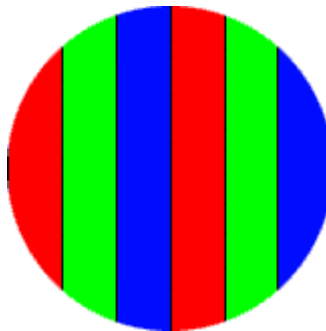
Shadow Mask and Aperture Grill

Right behind the glass on the front of a CRT monitor is a metal plate as well as a thin coat of phosphorus. After the Electron Gun has been aimed and fired, it passes through holes in the metal plate, and hits the phosphorus, causing it to glow. There are two variations of the metal plate commonly used today. The first is the Shadow Mask, which is filled with circular holes to let the electrons through. The second is the Aperture Grill, which consists of a series of parallel vertical slots to let the electrons through.

The Shadow Mask is the more popular type. It has three holes for each pixel on the screen, because of the three streams of electrons. Here is an illustration of what a monitor with a Shadow mask would look like up close:



The other common type is the Aperture Grill. As stated earlier, it features a series of parallel vertical slots to let the electrons through. It was developed by Sony, and is used by any monitors which use Sony's Trinitron technology, such as Mitsubishi, NEC, as well as Sony itself. One major advantage is that the Aperture Grill allows for more tightly packed pixels than the Shadow mask. This is evident when comparing the illustration of the Aperture Grill to the Shadow Mask. Notice that there is almost no spacing between the slots in the Aperture Grill, while there is much more spacing in the Shadow Mask:



The Advantages of CRT Monitors

There are many advantages to having a CRT Monitor. The first, and most obvious one, is the price. A CRT Monitor can be 3-4 times less expensive than a LCD Monitor. CRT Monitors are also much brighter than LCD Monitors, which lets the user see the screen more easily from farther distances and bigger angles than LCD Monitors. Lastly, they have an excellent refresh rate. Although this is not very obvious when working in Microsoft Word, or surfing the net, it is very obvious when watching high quality movies, as well as playing 3D games on the computer.

LCD Monitors

Liquid Crystal Display Monitors are slowly entering the world of Personal Computers. Although right now businesses account for most of the sales of LCD Monitors, this will slowly be shifting to home users as prices come down. For now, they are still much more expensive than CRT Monitors. Many people have been wondering whether they are really worth the price. To help you decide, we will show you how LCD technology works, and talk about some of the benefits over CRT Monitors.



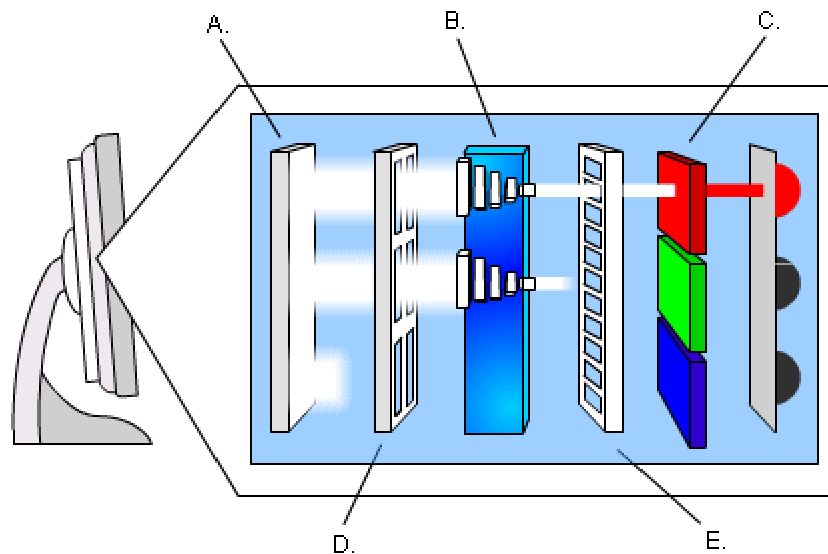
One of Viewsonic's 19" LCD Monitors
© www.viewsonic.com, 2001
Modified by the TQ Team "Gates of Creation"

LCD Technology

Liquid Crystal Displays are a bit more complicated than CRT Monitors. On the back of the screen is a fluorescent light which constantly lets out light. Right in front of it is a HPF (Horizontal Polarizing Filter). This filter only allows horizontal frequencies of light to pass through (if light is moving at a different angle from the horizontal position, it will not be allowed through the HPF). Next, the light that does pass through goes through a layer of liquid crystals. Voltage is applied to the liquid crystals, and depending on how much is applied, the liquid crystals twist the light anywhere from 0 to 90 degrees. The more the light is twisted, the higher the intensity of the light.

Next, the light passes through a Vertical Polarizing Filter, which only lets vertical light through. The more the light was twisted in the Liquid Crystals, the more light is let through the Vertical Polarizing Filter to be used in the front of the monitor as a pixel.

Three individual streams of light are used to create one pixel on the LCD Monitor. The next step, after the light passes through the Vertical Polarizing Filter, is to let it pass through a red, green or blue color filter. Once this is done, three different colored streams of light are turned into one pixel and displayed on the screen. To help you understand this more, we have created an illustration to show you what goes on in the LCD Monitor.



- A. Fluorescent Light
- B. Liquid Crystal
- C. Color Filter
- D. Vertical Polarizing Filter
- E. Horizontal Polarizing Filter

Refresh Rate

LCD Monitors also have a refresh rate, but it is not measured in the same way that it would be on a CRT Monitor. It is a measurement of how fast (in milliseconds) a pixel can change color. For an LCD Monitor, 25 milliseconds is very good. This means that it takes 25 milliseconds to refresh the whole screen. To measure how many times this monitor would be able to refresh the screen in 1 second, we divide 1 second by 25 milliseconds. $1000 \text{ milliseconds (which is 1 second)} / 25 \text{ milliseconds} = 40$. This monitor would be the equivalent of 40Hz on a CRT Monitor. As you can see, a typical CRT Monitor (60-75Hz) can refresh almost twice as fast as an LCD Monitor. Knowing this and the fact that our eyes can see 60 frames per seconds, we can see a very big disadvantage in this area. Playing games or watching DVD movies is therefore not as good on an LCD Monitor as it is on a CRT Monitor. With this in mind, lets take a look at some of the advantages of an LCD Monitor.

LCD Advantages

LCD Monitors have some very big advantages, especially for small and large businesses. Although the initial cost is 3 to 4 times as much, in the long run it saves money. LCD Monitors use up to three times less energy than CRT Monitors. Another advantage is the size and weight. Since no large CRT Tube is needed, they are not more than 1-2" thick, and weigh much less than CRT Monitors.

Video Cards

The Video Card is the fastest-evolving component of the PC. In fact, nVidia and ATI, leaders in the Video Card business, roll out more advanced video cards roughly every six months. If you compare this to CPUs, which evolve very quickly, you would realize the rapid evolution of Video Cards. Ever since the introduction of the third dimension to PCs, the race has been on to produce the fastest Video Card. Slowly, this has changed, and companies are starting to focus on image quality as much as the speed.



nVidia's Geforce3 delivers great performance as well as flawless image quality.
Screenshot taken from the game "Evolve"

Video Cards can be referred to with many different names. They include Video Boards, Graphics Boards, Display Adapters, as well as Video Cards. The names all mean the same thing. On this web site, we will most commonly refer to them as Video Cards.

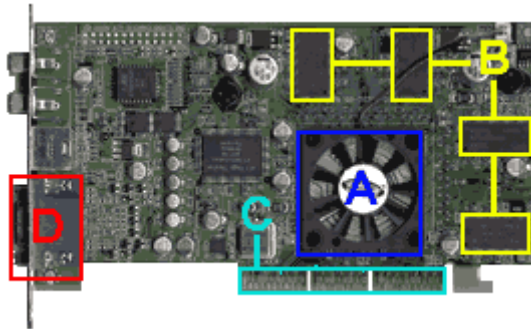
The Video Card segment of the Display Section will focus on the most important parts of today's video cards. This includes the parts of today's typical Video Card,

how the Video Card works in 2D, as well as the two contrasting methods of rendering in 3D. These two methods are called Immediate-Based Rendering, and Tile-Based Rendering.

Parts of a Typical Video Card

Although there are hundreds of parts to a Video Card, we will only cover the main ones. The ones we cover will give you an idea of how the Video Card can be regarded to as a mini-system, which has many of the components that the computer has itself. This is one of the reasons that it can usually be the one of the more expensive parts of a computer.

The image below shows ATI's Raedon Video Card, with the main components highlighted. Below the image, there will be descriptions of the individual parts to help you understand the Video Card better.



A. Graphics Processor Unit

Similar to the computer's CPU, the GPU processes the information that the Video Card needs in order to display the image on the monitor. The GPU has three typical functions. They are:

- Co-processor designed to take the load off the CPU. With today's complex images, a processor created especially for video is needed to help the computer run faster, and to create more life-like images.
- Used as a Graphics Accelerator. This helps the video card to render the graphics faster, based on the commands sent from the CPU.
- Frame Buffer: Controls the memory on the Video Card. It controls when the information will be sent to the Monitor.



The fastest GPU to date, nVidia's Geforce3 is a step ahead of the competition.

© nVidia Corp., 2001

B. Video Card Memory

The Memory is a very important part of a Video Card. In 2D, it stores the information needed to render images, such as the color of every individual pixel. In 3D programs and games, it has to hold a 3D image. This means that it has to hold depth, textures as well as lighting, which takes up much more memory. On most Video Cards, it even has to hold parts of the 3D images that are hidden, usually because they are covered by another object in front of it.



Video Card Memory

Memory is keeping the Video Card development back, because memory technology has not advanced as fast as other technologies used by the Video Card, such as the GPU. A new way of image rendering has been developed to limit the performance impact of the memory bottleneck. This topic is covered in the Tile Rendering page of the Display section.

C. AGP Connector

The Accelerated Graphics Port Connector connects the Video Card to the rest of the computer. AGP allows the Video card to access the system memory without having to go through the CPU, which improves video performance on computers. It allows the CPU to perform other tasks while the Video Card takes information from the system memory.



D. Display Connector

The Display Connector connects the Video Card to the Monitor. All the data that has finished being processed by the GPU and is ready to be sent to the monitor as an image is sent through here. Note: CRT Monitors and LCD Monitors use different inputs. Most Video Cards work with only CRT Monitors. Some work with LCD Monitors, but they are not very powerful.

applies if you are a casual computer user, not if you are a Professional or a Video Game fanatic. For those people, read the following pages in the Display section to get an in-depth look at Video Cards for 3D.

3D Rendering

When you view a 3D scene or play a 3D game on your PC, all you actually see is a 2-dimensional image. The image appears to be 3D, but the monitor is only able to display a 2D image because of the fact that it is flat. For many years, programs were designed, modeled and drawn in 2D. They looked much like a painting or a board game, everything looked flat against the screen.

A few years ago, programmers and game designers realized that although the monitor can only display 2D images, they can still be designed to give the appearance of 3 dimensions to give the images more life. The problem with this was that it took a lot of power from the computer. Even though the image on the monitor was only 2 dimensional, the computer still had to render the whole image in 3D. This meant that everything, even objects covered by other objects in front of them, had to be rendered. Following this, the Video Card had to decide which parts can be seen and which can't. For example, if a box was covering a part of the image behind it, the Video Card would have to recognize that. Also other parts, such as the back of objects, would have to be invisible in the final image sent to the monitor.



Thanks to unperfected depth perception by the Video Card, a player in this online multiplayer game seems to be on top of a rifle, three times bigger than him, which another player is holding.

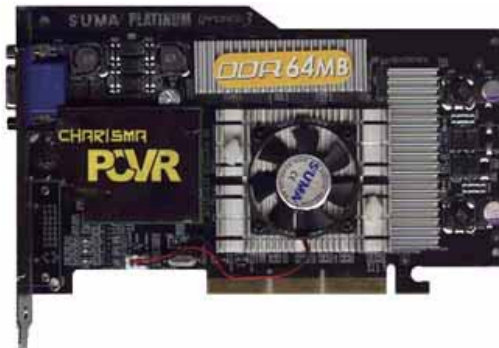
Screenshot taken from the game "Counter-Strike"

There is an analogy that can be used to explain what the Video Card does to make a 3D image. This analogy is taking a picture of a car. First, the whole car has to be built from scratch. Next, the picture is taken from the front. Although the picture is 2D, it seems 3D because it was taken of a real car. To create this 3D-realism time had to be spent to make the car from scratch. The same thing would have to happen inside the Video Card. It would have to build a 3D object, and then it would only use a fraction of the whole object in the picture.

The hardest problem is to overcome the depth perception. A 2D scene has height and width, while a 3D image also has depth. The computer has to decide which object should be visible. This is determined on the object's depth, distance away from the camera, as well as which objects are blocking any parts of it. If flawless

precision is not maintained, people would see problems with the images, such as seeing through walls, or objects not having the right perspective.

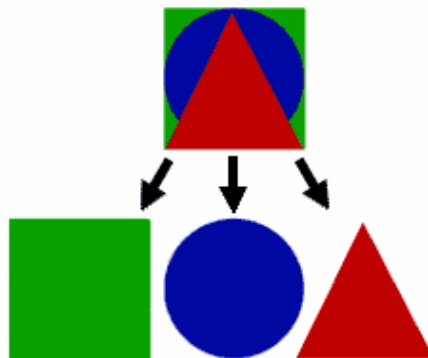
Since this kind of precision takes a lot of power, and the CPU could not handle it alone, a processor similar to the CPU, but specialized to handle 3D image rendering, was developed. It is basically used to render everything, and then decide what is and what is not seen.



This Video-Card features the most powerful 3D Accelerator GPU, the Geforce3
© www.anandtech.com, 2001

Immediate Mode Rendering

Immediate Mode Rendering is today's industry standard for 3D rendering of images. Using the example with the car, we can see that this is a very inefficient way to render images. In the typical Video Card, the whole car is built as a wire-frame model. Then, the whole car has textures applied to it. This is the real problem, because it is the most time-wasting thing the Video Card has to do. Following the application of textures to the car model, the Video Card needs to decide what is seen. The parts which are seen are sent to the monitor, while the rest of the data is useless. Immediate Mode Rendering exploits the memory speed problem further by doing things the way it does. Since 2/3 of the 3D image are not seen, the Video Card uses only uses 1/3 of the data it processes.



This example shows you that to render the top image, all three shapes on the bottom have to be created and textured. Is there any more efficient approach to 3D rendering? Slowly, a new approach is becoming more widely accepted. This approach is Tile-Based Rendering.

Tile-Based Rendering

Immediate Mode Rendering is not the most efficient type of rendering available, and you would think Video Card makers would migrate to a more efficient way of rendering in 3D. Well, for the most part they have not. Other modes of rendering have not received widespread support. One has gotten limited support though. It is called Tile-Based Rendering.



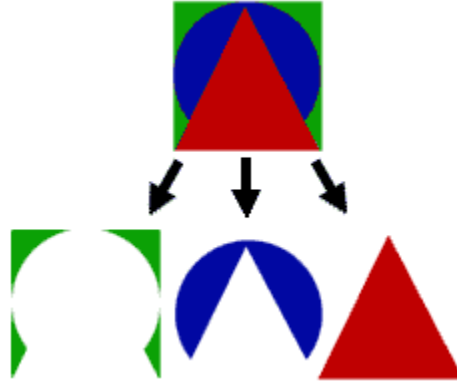
The KyroII Chipset has gained good support, thanks to the Video Card Giant Hercules' endorsement
© www.hercules.com, 2001

How It Works

Tile-Based rendering is quite a good technique for rendering 3D images faster. The makers of the Kyro2 GPU knew that more than 2/3 of everything that is rendered by standard IMR (Image Mode Rendering) Video Cards does not actually get used. They also knew that the whole 3D image had to be wire framed, just like it is in IMR, to keep the right perspective. Texturing wire frames which will not be used is where most of the performance gets lost, so they found a way to cut down on this factor to increase the efficiency tremendously.

How did they do it? Well, the standard IMR Video Card wire frames everything, and then textures everything. Finally, when it finishes texturing, it does a depth perspective test to figure out what is displayed, and what is not. The KyroII GPU does the same thing, except in a different order.

It first wire frames everything, just like IMR Video Cards. Next, instead of applying textures, it does the depth perspective test to figure out what will be displayed and what will not. Finally, it textures only the parts which will be displayed, saving time, and being more efficient. Tile-Based Rendering also reduces the problem with the memory speed bottleneck, allowing for faster innovation, and greater Video Card speeds.



As you can see from this illustration, to render the image on the top, much less work has to be done with Tile-Based rendering than Immediate Mode Rendering. Currently, Hercules is one of the only Video Card producers to use the KryoII GPU, but this will probably change, especially when the Kryo3 (a more advanced Tile-Based Rendering GPU) is released.