

Video Hardware

Display Adapters and Monitors

Although the monitor (or video display) is as vital to a PC's user interface as the mouse and keyboard, it is actually a latecomer to computing. Before CRT (cathode ray tube) monitors came into general use, the teletypewriter was the standard computer interface—a large, loud device that printed the input and output characters on a roll of paper. Early personal computers often used nothing more than a panel of blinking light-emitting diodes (LEDs) for a display.

The first monitors used on computers displayed only text in a single color (usually green), but to users at the time they were a great improvement, allowing real-time display of input and output data. Over time, color displays were introduced, screen sizes increased, and LCD technologies moved from the portable computer to the desktop. The latest trends reflect the increasing convergence of entertainment and computer technologies.

Although modern video hardware is much more sophisticated than that of the past, you should still be careful when selecting video hardware for your computer. A poor display can cause eyestrain or otherwise significantly diminish the experience of using your PC.

The video subsystem of a PC consists of two main components:

- **Monitor (or video display)**—The monitor is a display device usually based on an LCD or LED backlit panel, but it may also use CRT, plasma, or DLP technology.
- **Display adapter (also called the video card, graphics adapter, or graphics processing unit [GPU])**—Although this often refers to an adapter card plugged into a slot, the video adapter circuitry might also be built into the motherboard or included as part of the motherboard's chipset. Although it sounds strange, the circuitry is still called an *adapter* or *card* even if it is fully integrated into the motherboard, chipset, or processor.

This chapter explores the range of PC video display adapters on the market today and the monitors that work with them.

Note

The term *video*, as it is used in this context, does not necessarily imply the existence of a moving image, such as on a television screen. Any circuitry that feeds signals to a monitor or display is a video display adapter, regardless of whether it is used with applications that display moving images, such as multimedia or videoconferencing software.

For this reason, video cards are sometimes referred to as *graphics cards* or *display adapters*.

Video Display Adapters

A video display adapter (aka video card) provides the interface between your computer and your monitor and transmits the signals that appear as images on the display. Throughout the history of the PC, there has been a succession of standards for video hardware that represents a steady increase in screen resolution, color depth, and performance. The following list of standards can serve as an abbreviated history of PC video-display technology:

- MDA (Monochrome Display Adapter)
- HGC (Hercules Graphics Card)
- CGA (Color Graphics Adapter)
- EGA (Enhanced Graphics Adapter)
- VGA (Video Graphics Array)
- SVGA (Super VGA)
- XGA (Extended Graphics Array) and beyond

IBM pioneered most of these standards, but other manufacturers of compatible PCs adopted and enhanced them as well. Today, IBM no longer sets standards for the PC business (it sold its PC business to Lenovo in 2005), and most of these standards are obsolete.

Today's VGA and later video adapters can also handle most older software written for CGA, EGA, and other obsolete graphics standards. This enables you to use many, if not most, older graphics software (such as games and educational programs) on your current system.

Video Adapter Types

A monitor requires a source of input. The signals that run to your monitor come from one or more video display adapters in the system.

There are four basic types of video display adapters:

- **Discrete plug-in video cards**—These cards require the use of an expansion slot but provide the highest possible level of features and performance.
- **Discrete video on the motherboard**—The same discrete circuitry that can be found on a video card can also be directly built in or mounted on the motherboard. This is how high-end video is installed in modern laptops and some older desktop systems; however, modern desktops normally use either discrete video on a plug-in card or motherboard chipset or processor integrated video.
- **Motherboard chipset integrated video**—This is video circuitry integrated into the motherboard chipset. Because it shares the system random access memory (RAM) and other components, it provides an economical solution that also uses less power (ideal for laptops).

- **Processor integrated video**—This is video circuitry integrated into the processor, either as a separate die within the processor package or directly in the processor die. This form of integrated video also shares the system RAM and requires a motherboard with specific chipsets and video interface connectors on-board. Because chipset or processor-based integrated video shares the system RAM and other components, it has the lowest cost of any video solution. Performance is generally lower than that of a separate card-based adapter, especially for 3D gaming or other graphics-intensive applications. Resolution and color-depth options are also more limited than those available with add-on video cards. Because it is economical on power, integrated video is used in many laptops for improved battery life. Many desktop systems with integrated video also allow the installation of a discrete video plug-in card as an upgrade.

The term *video adapter* applies to either discrete or integrated video circuitry. The term *graphics adapter* is essentially interchangeable with *video adapter* because all video cards developed except the original IBM monochrome display adapter (MDA) can display graphics as well as text.

Note

AMD no longer uses the ATI brand name for either chipsets or video chips. Instead, AMD uses the RADEON brand name for consumer level video products, and Fire GL for professional level workstation video products.

Integrated Video/Motherboard Chipsets

Although built-in video has been a staple of low-cost computing for a number of years, until the late 1990s, most motherboard-based video simply mounted discrete video components on the motherboard. The performance and features of discrete video are essentially the same whether it is soldered into the motherboard or plugged in via an expansion card. In most cases the built-in discrete video can be upgraded by adding a video card. Some older motherboard-based discrete video implementations also had provisions for memory upgrades.

However, in recent years, the move toward increasing integration on the motherboard has led to the development of motherboard chipsets that include video support as part of the chipset design. In effect, the motherboard chipset takes the place of most of the discrete video card components and uses a portion of main system memory as video memory. The use of main system memory for video memory is often referred to as *unified memory architecture* (UMA), and although this memory-sharing method was also used by some built-in video that used its own chipset, it has become much more common with the rise of integrated motherboard chipsets.

Silicon Integrated Systems (SiS) pioneered chipsets with integrated video in 1996 and 1997 with its SiS5510 and SiS5596 chipsets for laptop and desktop systems, respectively. In 1997, Cyrix Semiconductor (now owned by VIA Technologies) introduced the MediaGX, which was the first to build both graphics and chipset functions into a PC-compatible CPU. National Semiconductor and later AMD developed improved versions of the MediaGX known as the Geode GX series.

Intel introduced motherboard chipsets with integrated graphics in 1999, starting with its 810 chipset for the Pentium III and Celeron processors. The 810 (code-named Whitney) heralded the beginning of widespread industry support for this design and the beginning of Intel's dominance in the graphics market. Intel later followed the release of the 810 series (810 and 810E) with the 815 series for the Pentium III and Celeron, most of which also feature integrated video.

Since then, Intel has offered versions of both desktop and mobile chipsets with integrated graphics and has become the world's largest supplier of graphics chips. This may sound strange because most people think of NVIDIA and AMD (thanks to its ownership of ATI) when it comes to graphics. Although NVIDIA and AMD may dominate in the high-end discrete graphics chip market, the market for lower-cost desktop and laptop systems with integrated graphics is larger than that for discrete graphics. Table 12.1 shows graphics chip market share data from JPR (Jon Peddie Research).

Table 12.1 Graphics Chip Market Share

Vendor	Q4 2010	Q4 2009	Q4 2008	Q4 2007
Intel	52.5%	51.1%	47.8%	43.5%
NVIDIA	22.5%	26.5%	30.7%	33.6%
AMD	24.2%	21.7%	19.3%	18.3%
SiS	0.0%	0.0%	1.1%	2.0%
VIA/S3	0.8%	0.7%	1.0%	2.5%
Matrox	0.1%	0.0%	0.1%	0.1%
Total	100.0%	100.0%	100.0%	100.0%

Table 12.2 shows the types and features for the integrated graphics available in Intel motherboard chipsets over the years.

Table 12.2 Intel Chipset Integrated Video

Processor Family	Chipset	Video Type	Features
PII, PIII, Celeron	810, 815	3D with Direct AGP	Basic 2D and 3D acceleration.
PII, PIII, Celeron	845	Extreme Graphics	Supports alpha blending, fog, anisotropic filtering, hardware motion compensation, and advanced textures.
PII, PIII, Celeron	865	Extreme Graphics 2	Enhanced version of Extreme Graphics, adding better memory management, zone rendering, and faster pixel and texturing rendering.
P4, Celeron	910, 915	GMA 900	An enhanced version of Extreme Graphics 2, adding support for most DirectX 9 features (does not support vertex shaders), optional dual-display support (requires ADD2 card), and support for widescreen LCD panels.
P4, PD, Celeron	945	GMA 950	A faster version of GMA 900 with support for DirectX 9.0c (Windows Aero capable), Shader Model 2.0, and Open GL 1.4 + Extensions.
Core 2, PD, P4	946, Q963, Q965	GMA 3000	Based on GMA 950. Adds execution units that can perform vertex and pixel shading for 3D graphics or video processing for video playback.
Core 2, P-dual	G31, G33, Q33, Q35	GMA 3100	Based on GMA 3000, but with reduced clock rates for somewhat lower performance and power consumption.
Core 2, P-dual, PD	G965	GMA X3000	Substantial redesign of GMA 3000 despite having a similar name. Includes support for DirectX 9.0c, Shader Model 3.0, hardware T&L (Transform & Lighting), OpenGL 1.5, hardware-based VC-1 decode, and Clear Video Technology for better video playback.
Core 2, P-dual, PD	GL960, GM965	GMA X3100	Enhanced version of GMA X3000 for mobile applications with upgraded support for DirectX 10 and Shader Model 4.0.

Table 12.2 Continued

Processor Family	Chipset	Video Type	Features
Core 2, P-dual, PD	G35	GMA X3500	Enhanced version of GMA X3000 for desktop applications with upgraded support for DirectX 10, Shader Model 4.0, and OpenGL 2.0.
Core2 Duo/Quad, P-dual	G41, G43, Q43, Q45	GMA X4500	High-performance version of GMA X3500 with added support for DisplayPort.
Core2 Duo/Quad, P-dual	G45, GM45	GMA X4500HD	Same as GMA X4500 but includes hardware accelerated VC-1 and AVC decode, allowing 1080p high-definition video playback (Blu-ray disc [BD] movies).

*PII – Pentium II**P4 – Pentium 4**PD – Pentium D**PIII – Pentium III**P-dual – Dual-Core Pentium**GMA = Graphics Media Accelerator*

Most other chipset vendors supporting Intel processors have also made versions with integrated graphics. Table 12.3 provides a reference to the most popular third-party integrated graphics chipsets for Intel Pentium 4 and later processors.

Table 12.3 Third-Party Chipsets with Integrated Video for Intel Processors

Processor Family	Vendor	Chipset	Video Type	Features
P4	ATI	IGP 330, IGP 340	RADEON VE (based on RADEON 7000)	DirectX 7, dualhead, DVD Playback
P4	ATI	9100 IGP, 9100 Pro IGP, 9000 Pro IGP	RADEON 9200 with pipelines reduced to two	DirectX 8.1, dualhead
P4	ATI	RS400, RC410	RADEON X300	DirectX 8.0, Open GL 2.0; RS400 also supports PCIe x16 slot
Core 2, PD, P4, Celeron D	NVIDIA	nForce 610i	GeForce 7050	DirectX 9.0SM3, DVI with HDCP
Core 2, PD, P4, Celeron D	NVIDIA	nForce 630i	GeForce 7100, 7150	DirectX 9.0SM3, dualhead, HDMI, DVI with HDCP
P4	VIA	P4M266	S3 Graphics ProSavage8 3D	DirectX 7 without hardware T&L
Core2, P4	VIA	PM800, P4M800 Pro	S3 UniChrome Pro	DirectX 7 without hardware T&L

*PII – Pentium II**P-dual – Dual-Core Pentium**PIII – Pentium III**PD – Pentium D**P4 – Pentium 4*

Thanks in part to AMD's takeover of ATI in late 2006 and AMD's encouragement of third-party chipset support for its 32-bit and 64-bit processors, there are many AMD and third-party chipsets with integrated video for AMD processors, including the most recent models. Table 12.4 provides an overview of the video features of chipsets from AMD and NVIDIA, both of which currently produce chipsets for AMD processors.

Table 12.4 Chipsets with Integrated Video for 64-bit AMD Processors

Processor Family	Vendor	Chipset	Video Type	Features
Athlon 64, Phenom, Sempron	AMD	690G	RADEON X1250	DirectX 9 SM2 without hardware vertex processing; HDMI and VGA
Athlon 64, Phenom, Sempron	AMD	690V	RADEON X1200	Slower version of X1250; lacks HDMI support
Athlon 64, Phenom, Sempron	AMD	740G	RADEON 2100	Faster version of RADEON X1250
Athlon 64, Phenom, Sempron	AMD	780V	RADEON HD3100	DirectX 10.1, OpenGL 3.2
Athlon 64, Phenom, Sempron ory; Hybrid	AMD	780G	RADEON HD3200	DirectX 10.1, OpenGL 3.2; faster version of HD3100; sideport mem-
Athlon 64, Phenom, Sempron	AMD	790GX	RADEON HD3300	Faster version of HD3200 with HDCP; Hybrid
Athlon 64, Phenom, Sempron	AMD	785G	RADEON HD4200	DirectX 10.1, OpenGL 3.2; sideport memory; Hybrid; Blu-ray playback dual DVI, HDMI
Phenom II, Athlon II, Sempron	AMD	880G	RADEON HD4250	Faster version of HD4200
Phenom II, Athlon II, Sempron	AMD	890GX	RADEON HD4290	Faster version of HD4250
Athlon 64 X2, Athlon 64, Sempron	NVIDIA	nForce 405, 410, 430	GeForce 6100	DX9 SM3, VGA
Athlon 64 X2, Athlon 64, Sempron	NVIDIA	nForce 430	GeForce 6150	Faster version of 6100 with DVI, HDTV playback
Athlon 64 X2, Athlon 64, Sempron	NVIDIA	nForce 630a	GeForce 7025	DX9 SM3, HDMI, HDCP, PureVideo, TV encoder
Athlon 64 X2, Athlon 64, Sempron	NVIDIA	nForce 630a	GeForce 7050PV	DX9 SM3, HDMI, HDCP
Phenom, Athlon 64 X2, FX, Athlon 64	NVIDIA	GeForce 8100	GeForce 8100	DX10, HDMI, HDCP
Phenom, Athlon 64 X2, Athlon 64	NVIDIA	GeForce 8200	GeForce 8200	DX10, OpenGL 3.2, HDMI, HDCP, SLI Hybrid, PureVideo
Phenom, Athlon 64 X2, Athlon 64	NVIDIA	GeForce 8300	GeForce 8300	Faster version of 8200

Newer integrated chipsets support digital outputs (such as Digital Visual Interface [DVI], High Definition Multimedia Interface [HDMI], or DisplayPort) for use with digital LCD panels and home theater components. Figure 12.2, later in this chapter, illustrates how you can differentiate these ports.

Although a serious 3D gamer will not be satisfied with the performance of integrated graphics, business, home/office, and casual gamers will find that integrated chipset-based video on recent platforms offers satisfactory performance and provides significant cost savings compared to a separate video card. If you decide to buy a motherboard with an integrated chipset, I recommend that you select one that also includes a Peripheral Component Interconnect Express (PCIe) x16 video expansion slot. This enables you to add a faster video card in the future if you decide you need it.

CPUs with Integrated Video

As I have noted elsewhere in this book, the design of personal computers started out with the use of many discrete chip-level components on the motherboard. As Chapter 4 discusses in detail, individual chips have been replaced by chipsets, several of which reduce the former functions of North and South Bridge, Super I/O, and video to a single chip. The next step, placing video and 3D rendering within the processor itself, is now a reality on both Intel and AMD platforms.

By moving video functions within the processor itself, improvements in processor speed, memory speed, and bus speed directly affect video and 3D speeds. Intel was the first to place video components inside the processors with its Core i-series processors, but AMD now also offers this feature in its new Fusion series of processors for netbooks, notebooks, tablets, and desktops. Table 12.5 compares the video graphics features found in Intel and AMD processors.

Table 12.5 Intel and AMD Processors with Integrated Video

Vendor	Processor	Video Type	Features
Intel	1st Gen Core i3/i5/i7	HD Graphics	Similar to GMA X4500HD but with two additional execution units (for a total of 12) and higher clock rates. Implemented as a separate 45nm die within the processor package.
Intel	2nd Gen Core i3/i5/i7	HD Graphics 2000	Similar to HD Graphics but with higher dynamically adjustable clock rates and only six execution units. Directly integrated into the processor die, which allows access to the CPU cache.
Intel	2nd Gen Core i3/i5/i7	HD Graphics 3000	HD Graphics 2000 but with 12 execution units and higher maximum clock rates for higher performance.
AMD	Fusion	RADEON HD 6000 series	DirectX 11, DirectCompute, OpenCL,

DirectCompute – Microsoft DirectX 11 (also available in DX10) API supporting general-purpose computing on GPUs
OpenCL – Open Computing Language; enables any application to access the GPU for non-graphical processing; standard managed by Khronos Group (www.khronos.org)

Video Adapter Components

Video display adapters contain certain basic components, usually including the following:

- Video basic input/output system (BIOS). Provides the firmware controlling the video card.
- Video processor/video accelerator. Places text or graphics information on-screen, and provides 3D acceleration.
- Video memory. Holds the information.
- Digital-to-analog converter (DAC). Formerly a separate chip, the DAC is usually incorporated into the video processor/accelerator chip. The DAC is not necessary on a purely digital subsystem (digital video card and display); however, most display subsystems still include analog VGA support.
- Bus connector. Connects a discrete video card to the motherboard.
- Video driver. Software that communicates with the OS to control the card.

On mid-range and high-performance video cards, such as the card shown in Figure 12.1, most of the components are underneath the cooling system. This card uses a combination of a fan and a heatsink to cool its GPU and memory modules.

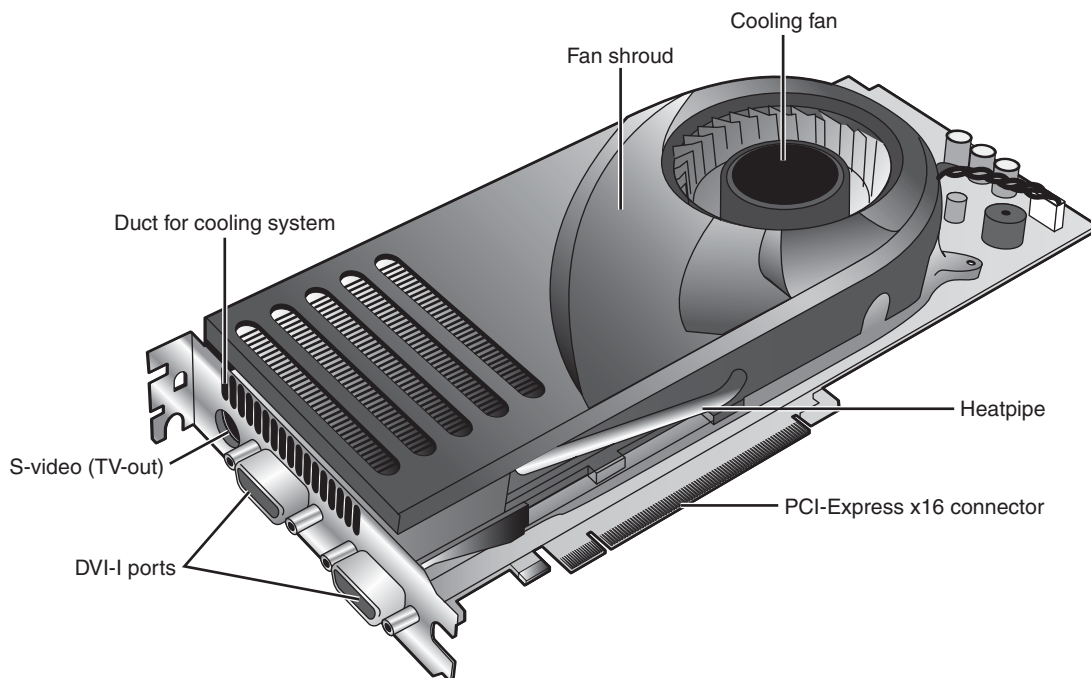


Figure 12.1 A typical example of a mid-range video card optimized for dual-GPU gaming (NVIDIA SLI) and HDTV support.

Virtually all video adapters on the market today use chipsets that include 3D acceleration features. The following sections examine the video BIOS and processor in greater detail.

The Video BIOS

Video adapters include a BIOS that is separate from the main system BIOS. If you turn on your monitor first and look quickly, you might see an identification banner for your adapter's video BIOS at the beginning of the system startup process.

Similar to the system BIOS, the video adapter's BIOS takes the form of a read-only memory (ROM) chip containing basic instructions that provide an interface between the video adapter hardware and the software running on your system. The software that makes calls to the video BIOS can be a stand-alone application, an operating system (OS), or the main system BIOS. The programming in the BIOS chip enables your system to display information on the monitor during the system POST and boot sequences, before any other software drivers have been loaded from disk.

◀ See "BIOS Basics," p. 251 (Chapter 5, "BIOS").

In a few cases (primarily with older video hardware), the video BIOS also can be upgraded, just like a system BIOS. The video BIOS normally use a rewritable chip called an electrically erasable programmable read-only memory (EEPROM). On old cards, you might be able to completely replace the chip with a new one—again, if supplied by the manufacturer and if the manufacturer did not hard-solder the BIOS to the printed circuit board. Most video cards use a surface-mounted BIOS chip rather than a socketed chip. A BIOS you can upgrade using software is referred to as a *flash* BIOS, and most video cards that offer BIOS upgrades use this method. However, because the video BIOS is only used during startup for VGA emulation, such upgrades are rarely necessary, and most vendors fix problems by issuing updated drivers rather than BIOS updates.

Note

Video BIOS upgrades are sometimes referred to as *firmware upgrades*.

The Video Processor

The video processor (also known as the video chipset, video graphics processor, or GPU) is the heart of any video adapter and essentially defines the card's functions and performance levels. Two video adapters built using the same chipset will have the same basic capabilities. However, cards built using the same chipset can vary in the clock speeds at which they run the chipset, memory, and other components, as well as in the amount and type of memory installed. Therefore, performance can vary. The software drivers that OSs and applications use to address the video adapter hardware are written primarily with the chipset in mind. You can normally use a driver intended for an adapter with a particular chipset on any other adapter using the same chipset, or the same chipset families.

Identifying the Video and System Chipsets

Before you purchase a system or a video card, you should find out which chipset the video card or video circuit uses. For systems with integrated chipset video, you need to find out which integrated chipset the system uses. This allows you to have the following:

- A better comparison of the card or system to others
- Access to technical specifications
- Access to reviews and opinions
- The ability to make a better buying decision
- The choice of card manufacturer or chipset manufacturer support and drivers

Because video card performance and features are critical to enjoyment and productivity, find out as much as you can before you buy the system or video card by using the chipset or video card manufacturer's website and third-party reviews. Poorly written or buggy drivers can cause several types of problems, so be sure to check periodically for video driver updates and install any that become available. With video cards, support after the sale can be important, particularly for 3D gaming. Therefore, you should check the manufacturer's website to see whether drivers for your video card have problems rendering your favorite games or working with your favorite software and how often updates are available.

Note

Sometimes, newer drivers for a particular video card or chipset render newer 3D games and programs better than older games and programs. If you are using the latest driver and older games or programs aren't playing as well as they should, you should consider downgrading to an older version of the driver.

Note that although NVIDIA and AMD are the leading suppliers of discrete graphics processors, they do not make video cards. Instead, they create video card reference designs, which the various card manufacturers use to develop their own specific cards. Because each card manufacturer can customize or modify the designs as it chooses, two cards that use the same graphics chipset may differ in features as well as in actual performance. This means a variety of video cards use the same chipset; it also means you are likely to find variations in card performance, software bundles, warranties, and other features between cards using the same chipset.

Although vendors might make modifications to the reference designs provided by NVIDIA and AMD, you can use the drivers provided by the GPU manufacturers for your GPU if the card or motherboard vendor does not provide updated drivers on a timely basis.

Video RAM

Most discrete video adapters rely on their own onboard memory that they use to store video images while processing them. Systems with integrated video use the universal memory architecture (UMA) feature to share the main system memory. In any case, the memory on the video card or the memory borrowed from the system performs the same tasks.

The amount of video memory determines the maximum screen resolution and color depth the device can support, among other features. You often can select how much memory you want on a particular video adapter; for example, cards with anywhere from 512MB to 4GB of video memory are common choices today. Although having more video memory is not guaranteed to speed up your video adapter, it can increase the speed if it enables a wider bus (for example, from 128 bits wide to 256 bits wide) or provides non-display memory as a cache for commonly displayed objects. It also enables the card to generate more colors and higher resolutions and allows 3D textures to be stored and processed on the card, rather than in slower main memory.

Many types of memory have been used with video adapters. These memory types are summarized in Table 12.6.

Table 12.6 Memory Types Used in Video Display Adapters

Memory Type	Definition
FPM DRAM	Fast Page-Mode RAM
VRAM	Video RAM
WRAM	Window RAM
EDO DRAM	Extended Data Out DRAM
SDRAM	Synchronous DRAM
MDRAM	Multibank DRAM
SGRAM	Synchronous Graphics RAM
DDR SGRAM	Double-Data Rate SGRAM
GDDR2 SGRAM	Graphics DDR version 2 SGRAM
GDDR3 SGRAM	GDDR version 3 SGRAM
GDDR4 SGRAM	Graphics DDR version 4 SGRAM
GDDR5 SGRAM	GDDR version 5 SGRAM

Some of these, including FPM DRAM, EDO DRAM, and SDRAM, were also used for main memory in PCs. All of the others were specifically designed for use in graphics subsystems.

◀ For more information about FPM DRAM, EDO DRAM, and SDRAM, **see** Chapter 6, “Memory,” **p. 315**.

VRAM and WRAM

VRAM and WRAM are dual-ported memory types that can read from one port and write data through the other port. This improves performance by reducing wait times for accessing the VRAM compared to FPM DRAM and EDO DRAM.

SGRAM

SGRAM was designed to be a high-end solution for fast video adapter designs. SGRAM is similar to SDRAM in its capability to be synchronized to high-speed buses up to 200MHz, but it differs from SDRAM by including circuitry to perform block-writes to increase the speed of graphics fill and 3D Z-buffer operations.

DDR SGRAM

DDR SGRAM is designed to transfer data at speeds twice that of conventional SGRAM by transferring data on both the rising and falling parts of the processing clock cycle.

GDDR2 SGRAM

There have been several variations of what has been called GDDR2. The first was based on standard 2.5V DDR SDRAM with some enhancements, whereas the second was actually based on 1.8V DDR2 SDRAM, and with much higher performance and cooler operation.

GDDR3 SGRAM

GDDR3 SGRAM is based on DDR2 memory, but with two major differences:

- GDDR3 separates reads and writes with a single-ended unidirectional strobe, whereas DDR2 uses differential bidirectional strobes. This method enables higher data transfer rates.
- GDDR3 uses an interface technique known as pseudo-open drain, which uses voltage instead of current. This method makes GDDR3 memory compatible with GPUs designed to use DDR, GDDR2, or DDR2 memory. To determine the type of memory used on a particular video card, check the video card manufacturer's specification sheet.

GDDR4 SGRAM

Some of the newer cards use GDDR4 SGRAM. Compared to GDDR3, GDDR4 memory has the following features:

- Higher bandwidth. GDDR4 running at half the speed of GDDR3 provides comparable bandwidth to its predecessor.
- Greater memory density, enabling fewer chips to be needed to reach a particular memory size.

GDDR5 SGRAM

GDDR5 SGRAM is based on the previous GDDR standards with several modifications to allow increased performance. The main differences include the following:

- Signal optimization using data/address bit inversion, adjustable driver strength, adjustable voltage, and adjustable termination
- Adaptive interface timing using data training that is scalable per bit or byte
- Error compensation, including real-time error detection on both read/write and fast resending

GDDR5 is also designed for extreme power management such that power is used only when necessary. This allows higher clock speeds with cooler operation. Current GDDR5 parts are rated up to 7Gbps per chip, allowing 28GBps total bandwidth.

Video RAM Speed

Video RAM speed is typically measured in MHz, GHz, or by bandwidth in Mb/Gb or MB/GB per second. Faster memory and faster GPUs produce better gaming performance, but at a higher cost. However, if you are primarily concerned about business or productivity application performance, you can save money by using a video card with a slower GPU and slower memory.

Unless you dig deeply into the technical details of a particular graphics card, determining what type of memory a particular card uses can be difficult. Because none of today's video cards feature user-upgradeable memory, I recommend that you look at the performance of a given card and choose the card with the performance, features, and price that's right for you.

RAM Calculations

The amount of memory a video adapter needs to display a particular resolution and color depth is based on a mathematical equation. A location must be present in the adapter's memory array to display every pixel on the screen, and the resolution determines the number of total pixels. For example, a screen resolution of 1920×1080 requires a total of 2,073,600 pixels.

If you were to display that resolution with only two colors, you would need only 1 bit of memory space to represent each pixel. If the bit has a value of 0, the dot is black, and if its value is 1, the dot is white. If you use 32 bits of memory space to control each pixel, you can display more than 4 billion colors because 4,294,967,296 combinations are possible with a 32-digit binary number ($2^{32}=4,294,967,296$). If you multiply the number of pixels necessary for the screen resolution by the number of bits required to represent each pixel, you have the amount of memory the adapter needs to display that resolution. Here is how the calculation works:

$$\begin{aligned} 1920 \times 1080 &= 2,073,600 \text{ pixels} \times 32 \text{ bits per pixel} \\ &= 66,355,200 \text{ bits} \\ &= 8,294,400 \text{ bytes} \\ &= 7.91 \text{ MiB} \end{aligned}$$

As you can see, displaying 32-bit color (4,294,967,296 colors) at 1920×1080 resolution requires just short of 8MiB of RAM on the video adapter. Because most modern video cards have at least 512MiB or more, you can see that two-dimensional images don't require much memory.

3D video cards require more memory for a given resolution and color depth because the video memory must be used for three buffers: the front buffer, back buffer, and Z-buffer. The amount of video memory required for a particular operation varies according to the settings used for the color depth and Z-buffer. Triple buffering allocates more memory for 3D textures than double buffering but can slow down the performance of some games. You can usually adjust the buffering mode used by a given 3D video card through its properties sheet.

Begin Note

As we discuss 3D in this and following sections, we are referring to the rendering of 3D graphics on your display. 3D display technologies (which involve special glasses and 120MHz or greater refresh rates on LCD or LED backlit displays) is a different issue.

Although current chipset-integrated graphics solutions feature 3D support, the performance they offer is limited by being based on older, less powerful 3D GPUs and by the narrow data bus they use to access memory. Because integrated graphics solutions share video memory with the processor, they use the same data bus as the processor. In a single-channel-based system, this restricts the data bus to

64 bits. A dual-channel system has a 128-bit data bus, but today's fastest 3D video cards feature a 512-bit or wider data bus. The wider the data bus, the more quickly graphics data can be transferred.

For these reasons, you are likely to be disappointed (and lose a lot of games!) if you play 3D games using integrated graphics. To enjoy 3D games, opt for a mid-range to high-end 3D video card based on a current AMD or NVIDIA chipset with 1GB of RAM or more. If you demand peak performance and your budget permits, you might also consider using a multicard solution that allows you to use two or more PCIe video cards to increase your graphics processing performance.

►► **See** "Dual-GPU Scene Rendering," **p. 637** (this chapter).

Note

If your system uses integrated graphics and you have less than 2GB of RAM, you might be able to increase your available graphics memory by upgrading system memory (system memory is used by the integrated chipset). Some Intel chipsets with integrated graphics automatically detect additional system memory and adjust the size of graphics memory automatically.

Video Memory Bus Width

Another issue with respect to the memory on the video adapter is the width of the bus connecting the graphics chipset and memory on the adapter. The chipset is usually a single large chip on the card that contains virtually all the adapter's functions. It is wired directly to the memory on the adapter through a local bus on the card. Most of the high-end adapters use an internal memory bus that is up to 512 bits wide (or more in some cases). This jargon can be confusing because video adapters that take the form of separate expansion cards also plug into the main system bus, which has its own speed rating. When you read about a 256-bit, 384-bit or 512-bit video adapter, you must understand that this refers to the memory connection on the card, not the connection to the motherboard. In two cards with otherwise similar GPU, memory type, and memory size specifications, the card with the wider memory bus is preferable because a wider memory bus boosts performance.

◄◄ **See** "System Bus Types, Functions, and Features," **p. 233** (Chapter 4, "Motherboards and Buses").

The DAC

The DAC (or RAMDAC) does exactly what its name describes. The RAMDAC is responsible for converting the RAM-based digital images your computer generates into signals for analog monitor connections. The speed of the RAMDAC is measured in MHz; the faster the conversion process, the higher the adapter's vertical refresh rate. The speeds of the RAMDACs used in today's high-performance video adapters range from 300MHz to 400MHz. Most of today's video card chipsets include the RAMDAC function inside the 3D accelerator chip, but some dual-display-capable video cards use a separate RAMDAC chip to allow the second display to work at different refresh rates than the primary display. Systems that use integrated graphics include the RAMDAC function in the North Bridge or GMCH chip portion of the motherboard chipset.

The benefits of increasing the RAMDAC speed include higher vertical refresh rates, which allow higher resolutions with flicker-free refresh rates (72Hz–85Hz or above) on CRT displays. Typically, cards with RAMDAC speeds of 300MHz or above display flicker-free (75Hz or above) at all resolutions up to 1920×1200 on CRT displays. Of course, as discussed earlier in this chapter, you must ensure that any resolution you want to use is supported by both your monitor and video card.

Video Display Interfaces

Video display adapters connect a PC to a display and therefore must work through two main interfaces. The first is the system interface, meaning the connection between the video adapter and the PC, and the second is the display interface, meaning the connection between the video adapter and the display. By using standardized versions of these interfaces, we end up having video adapters and displays that are both compatible and easily interchangeable. This section discusses the available system and display interfaces as well as the differences between them.

The System Interface

Older video adapters were designed for use with earlier bus standards, such as the IBM MicroChannel Architecture (MCA), Interactive Multimedia Association (ISA), Enhanced Interactive Multimedia Association (EISA), and VESA local bus (VL-Bus). Because of their relatively slow performance by today's standards, all are now obsolete. Current video display adapters use PCIe interface standards to connect to a system, although older systems based on Accelerated Graphics Port (AGP) are still relatively common. Peripheral Component Interconnect (PCI)-based adapters are still found in some old systems or AGP-based systems that utilize a second adapter.

- ◀◀ **See** "The PCI Bus," p. 232 (Chapter 4).
- ◀◀ **See** "PCI Express," p. 235 (Chapter 4).
- ◀◀ **See** "Accelerated Graphics Port," p. 237 (Chapter 4).

AGP

The AGP, an Intel-designed dedicated video bus introduced in 1997, delivers a maximum bandwidth up to 16 times greater than that of a comparable PCI bus. AGP was the mainstream high-speed video-to-system interface for several years but has been replaced by the more versatile and faster PCIe standard.

The AGP slot is essentially an enhancement to the existing PCI bus; however, it's intended for use only with video adapters and provides them with high-speed access to the main system memory array. This enables the adapter to process certain 3D video elements, such as texture maps, directly from system memory rather than having to copy the data to the adapter memory before the processing can begin. This saves time and eliminates the need to upgrade the video adapter memory to better support 3D functions. Although AGP version 3.0 provides for two AGP slots, this feature has never been implemented in practice. Systems with AGP have only one AGP slot.

Note

Although the earliest AGP cards had relatively small amounts of onboard RAM, most later implementations use large amounts of on-card memory and use a memory aperture (a dedicated memory address space above the area used for physical memory) to transfer data more quickly to and from the video card's own memory. Integrated chipsets featuring built-in AGP use system memory for all operations, including texture maps.

Windows 98 and later versions support AGP's Direct Memory Execute (DIME) feature. DIME uses the main memory instead of the video adapter's memory for certain tasks to lessen the traffic to and from the adapter. However, with the large amounts of memory found on current AGP video cards, this feature is seldom implemented.

Four speeds of AGP are available: 1x, 2x, 4x, and 8x. (See Table 12.7 for details.) Later AGP video cards support AGP 8x and can fall back to AGP 4x or 2x on systems that don't support AGP 8x.

Table 12.7 AGP Speeds and Technical Specifications

AGP Speed	AGP Specification	Clock Speed	Transfer Rate	Slot Voltages
1x	1.0	66MHz	266MBps	3.3V
2x	1.0	133MHz	533MBps	3.3V, 1.5V ¹
4x	2.0	266MHz	1066MBps	1.5V
8x	3.0	533MHz	2133MBps	1.5V ²

1. Varies with card implementation.

2. Uses 0.8V internal signaling.

AGP 3.0 was announced in 2000, but support for the standard required the development of motherboard chipsets that were not introduced until mid-2002. Almost all motherboard chipsets with AGP support released after that time featured AGP 8x support.

Although some systems with AGP 4x or 8x slots use a universal slot design that can handle 3.3V or 1.5V AGP cards, others do not. If a card designed for 3.3V (2x mode) is plugged into a motherboard that supports only 1.5V (4x mode) signaling, the motherboard may be damaged.

◀ **See** “Accelerated Graphics Port,” **p. 237** (Chapter 4).

Caution

Be sure to check AGP compatibility before you insert an older (AGP 1x/2x) card into a system designed for AGP 4x or 8x. Even if you can physically insert the card, a mismatch between the card’s required voltage and the AGP slot’s voltage output can damage the motherboard. Check the motherboard manual for the card types and voltage levels the AGP slot supports.

Some AGP cards can use either 3.3V or 1.5V voltage levels, adjusted via an onboard jumper. These cards typically use an AGP connector that is notched for use with either AGP 2x or AGP 4x slots, as pictured in Chapter 4, “Motherboards and Buses.” Be sure to set these cards to use 1.5V before using them in motherboards that support only 1.5V signaling.

PCIe

PCIe began to show up in systems in mid-2004 and has filtered down to all systems that use discrete video cards or have integrated video that can be upgraded. Despite the name, PCIe uses a high-speed bidirectional serial data transfer method, and PCIe channels (also known as *lanes*) can be combined to create wider and faster expansion slots. (Each lane provides 250MBps, 500MBps, or 1,000MBps data rate in each direction.) Because PCIe is technically not a bus, unlike PCI, the slots do not compete with each other for bandwidth. PCIe graphics cards use up to 16 lanes (x16) to enable speeds of 4GBps, 8GBps, or 16GBps in each direction, as seen in Table 12.8.

Table 12.8 PCIe Video Card Bandwidth

PCIe Architecture	Bandwidth per Lane	Bandwidth for x16 Link
PCIe 1.x	250MBps	4GBps
PCIe 2.x	500MBps	8GBps
PCIe 3.x	1,000MBps	16GBps

Most PCIe implementations include one x16 slot for video and two or more x1 slots for other add-on cards, as well as legacy PCI slots. Systems that support NVIDIA's SLI or AMD CrossFire X dual PCIe video card technologies have up to three or four PCIe video slots running at x8 or x16 speed.

The Display Interface

The display interface connects the video display adapter to the monitor or display. Over the years, several methods of connecting monitors have been available. Some of these interfaces have been analog, and others have been digital. The first interfaces used in the 1980s for early PCs included CGA, EGA, MDA, and HCG. These were all digital interfaces.

Note

For more information on these early display standards, see *Upgrading and Repairing PCs, 10th Anniversary edition*, now out of print but may be available at some bookstores or libraries.

Unlike earlier digital video standards, VGA is an analog system. VGA came out in 1987 and began a shift from digital to analog that lasted for more than 20 years. Only recently has there been a shift back to digital. Why go from digital to analog and then back to digital? The simple answer is that analog was the least expensive way at the time to design a CRT-based system that supported a reasonable resolution with a reasonable number of colors. Now that technology has advanced and LCD panels have largely replaced CRTs, going back to digital interfaces makes sense.

The video interfaces (and connectors) you are likely to encounter in PCs dating from the late 1980s to the present include the following:

- VGA
- DVI
- HDMI
- DisplayPort

VGA is an analog connection, whereas the others are digital. The connectors for these interfaces are shown in Figure 12.2 and are discussed in further detail in the following sections.

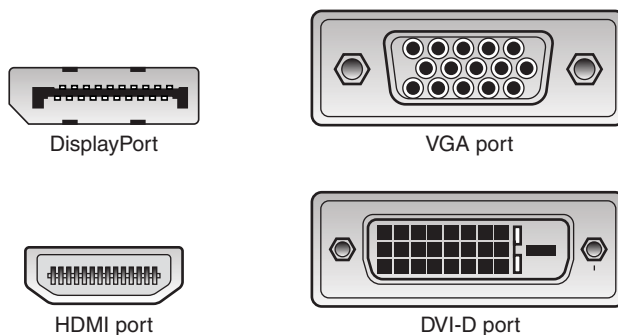


Figure 12.2 Video interface connectors used in PCs from the late 1980s to the present.

VGA

IBM introduced the VGA interface and display standard on April 2, 1987, along with a family of systems it called PS/2. VGA originally included the display adapter, the monitor, and the connection between them. Since that time, the display adapters and monitors have evolved, but the VGA 15-pin analog connection went on to become the most popular video interface in history and is still used today in PC video adapters and displays.

VGA is an analog design. Analog uses a separate signal for each CRT color gun, but each signal can be sent at varying levels of intensity—64 levels, in the case of the original VGA standard. This provides 262,144 possible colors (64^3), of which 256 could be simultaneously displayed in the original design. For realistic computer graphics, color depth is often more important than high resolution because the human eye perceives a picture that has more colors as being more realistic.

VGA was designed to be addressed through the VGA BIOS interface, a software interface that forced programs to talk to the BIOS-based driver rather than directly to the hardware. This allowed programs to call a consistent set of commands and functions that would work on different hardware, as long as a compatible VGA BIOS interface was present. The original VGA cards had the BIOS on the video card directly, in the form of a ROM chip containing from 16KB to 32KB worth of code. Modern video cards and laptop graphics processors still have this 32KB onboard BIOS. Typically, the only time the ROM-based drivers are used is during boot, when running legacy DOS-based applications or games, or when you run Windows in Safe Mode.

VGA also describes a 15-pin analog interface connection that can support a variety of modes. The connection is analog because VGA was primarily designed to drive CRT displays, which are analog by nature. When a display is connected via VGA, the digital signals inside the PC are converted to analog signals by the DAC chip in the display adapter and are then sent to the display via the analog VGA connection. The VGA connector is shown in Figure 12.3; the pinouts are shown in Table 12.9.

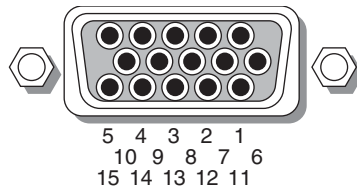


Figure 12.3 The standard 15-pin analog VGA connector.

Table 12.9 15-Pin Analog VGA Connector Pinout

Pin	Function	Direction
1	Red Analog Video	Out
2	Green Analog Video	Out
3	Blue Analog Video	Out
4	Monitor ID 2	In
5	TTL GND (monitor self-test)	—
6	Red Analog GND	—
7	Green Analog GND	—
8	Blue Analog GND	—
9	Key (unused)	—

Table 12.9 Continued

Pin	Function	Direction
10	Sync GND	—
11	Monitor ID 0	In
12	Monitor ID 1	In
13	Horizontal Sync	Out
14	Vertical Sync	Out
15	Monitor ID 3	In

The mating VGA cable connector that plugs into this connector normally has pin 9 missing. This was designed such that the mating hole in the connector on the video card could be plugged, but it is usually open (and merely unused) instead. The connector is keyed by virtue of the D-shape shell and pin alignment, so it is difficult to plug in backward even without the key pin. Pin 5 is used only for testing purposes, and pin 15 is rarely used; they are often missing as well. To identify the type of monitor connected to the system, some early VGA cards used the presence or absence of the monitor ID pins in various combinations.

In addition to the connector and electrical interface, the original VGA standard also defined a number of text and graphics display modes with various resolutions and colors. The original VGA modes allowed for a maximum graphics resolution of 640×480 in only 16 (4-bit) colors. This was the maximum that could be supported by the original 256KB of RAM included on the card.

IBM introduced higher-resolution versions of VGA called XGA and XGA-2 in the early 1990s, but most of the development of VGA standards has come from the third-party video card industry and its trade group, the Video Electronic Standards Association (VESA; www.vesa.org). When VGA originated in 1987, it had low resolution and color capability by today's standards. Since then, VGA has evolved to support higher resolution modes with many more colors. Even the least-expensive video adapters on the market today can work with modes well beyond the original VGA standard.

The minimum resolution and color depth (number of colors) for use with current software is 1024×768 (originally introduced as part of the XGA standard) and 32-bit color depth for 3D adapters.

Note

For more information about early VGA and VESA SVGA, and XGA standards, see "SVGA and XGA" in Chapter 12, "Video Hardware" in *Upgrading and Repairing PCs*, 19th edition, available in its entirety on the disc packaged with this book.

Digital Display Interfaces

The analog VGA interface works well for CRTs, which are inherently analog devices, but VGA does not work well for LCD, plasma, or other types of flat-panel displays that are inherently digital. Video data starts out digitally inside the PC and is converted to analog when the VGA interface is used. When you are running a digital display such as an LCD over an analog interface such as VGA, the signal must then be converted back to digital before it can be displayed, resulting in a double conversion that causes screen artifacts, blurred text, color shifting, and other kinds of problems.

Using a digital interface eliminates the double conversion, allowing the video information to remain as digital data from the PC all the way to the screen. Therefore, a trend back to using digital video interfaces has occurred, especially for inherently digital displays such as LCD flat panels.

Laptop computers have avoided this problem by using an internal digital connection called FPD-Link (Flat Panel Display-Link), which National Semiconductor originally developed in 1995. Unfortunately, this standard was not designed for external connections requiring longer cable lengths or extremely high resolutions. What was needed was an industry standard digital connection for external displays.

To facilitate a digital video connection between PCs and external displays, several digital video signal standards and specifications have been available:

- Plug and Display (P&D)
- Digital Flat Panel (DFP)
- DVI
- HDMI
- DisplayPort

VESA released the P&D and DFP standards in June 1997 and February 1999, respectively. Both were based on the PanelLink TMDS (Transition Minimized Differential Signaling) protocol that Silicon Image developed. Unfortunately, both of these interfaces had relatively low-resolution support (1280×1024 maximum) and were only implemented in a handful of video cards and monitors. As such, they never really caught on in the mass market and were overshadowed by the DVI, which became the first truly popular digital display interface standard.

DVI

DVI was introduced on April 2, 1999 by the Digital Display Working Group (DDWG). The DDWG was formed in 1998 by Intel, Silicon Image, Compaq, Fujitsu, Hewlett-Packard, IBM, and NEC to address the need for a universal digital interface standard between a host system and a display. Unlike the P&D and DFP interfaces that came before it, DVI gained immediate widespread industry support, with 150 DVI products being shown at the Intel Developer Forum in August 1999, only four months after DVI was released. Since then, DVI has become the most popular interface for digital video connections. DVI also allows for both digital and VGA analog connections using the same basic connector.

DVI uses Transition Minimized Differential Signaling (TMDS), which was developed by Silicon Image (www.siliconimage.com) and trademarked under the name *PanelLink*. TMDS takes 24-bit parallel digital data from the video controller and transmits it serially over balanced lines at a high speed to a receiver. A single-link TMDS connection uses four separate differential data pairs, with three for color data (one each for red, green, and blue data) and the fourth pair for clock and control data. Each twisted pair uses differential signaling with a low 0.5V swing over balanced lines for reliable, low-power, high-speed data transmission. A low-speed VESA Display Data Channel (DDC) pair is also used to transmit identification and configuration information, such as supported resolution and color-depth information, between the graphics controller and display.

TMDS is designed to support cables up to 10 meters (32.8 feet) in length, although the limits may be shorter or longer depending on cable quality. Several companies make products that can amplify or re-drive the signals, allowing for greater lengths. Figure 12.4 shows a block diagram of a single-link TMDS connection.

Using TMDS, each color channel (red/green/blue) transmits 8 bits of data (encoded as a 10-bit character) serially at up to 165MHz. This results in a raw throughput of 1.65Gbps per channel. There are three color channels per link, resulting in a maximum raw bandwidth of 4.95Gbps per link. Because the data is sent using 8b/10b encoding, only 8 bits of every 10 are actual data, resulting in a maximum true video data throughput of 3.96Gbps. This enables a single-link DVI connection to easily handle computer video resolutions as high as WUXGA (1920×1200) as well as 1080p HDTV (1920×1080 with progressive scan).

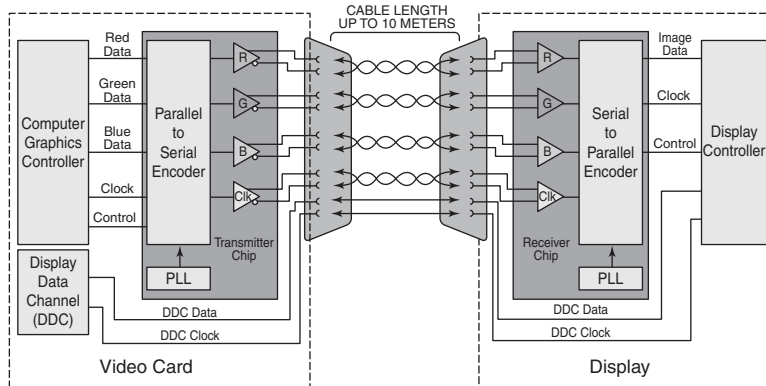


Figure 12.4 A single-link TMDS connection.

If more bandwidth is necessary, the DVI standard supports a second TMDS link in the same cable and connector. This uses three additional TMDS signal pairs (one for each color) and shares the same clock and DDC signals as the primary link. This is called *dual-link DVI*, and it increases the maximum raw bandwidth to 9.9Gbps and the true data bandwidth to 7.92Gbps, which will handle computer resolutions as high as WQUXGA (3840×2400). Normally only 30-inch or larger flat-panel displays use resolutions high enough to require dual-link DVI. Even higher resolution displays can be supported with dual DVI ports, each with a dual-link connection.

TMDS links include support for DDC, a low-speed, bidirectional standard for communication between PCs and monitors, created by VESA. DDC defines the physical connection and signaling method, whereas the communications and data protocol is defined under the VESA Extended Display Identification Data (EDID) standard. DDC and EDID allow the graphics controller to identify the capabilities of the display so the controller can automatically configure itself to match the display's capabilities.

DVI uses Molex MicroCross connectors with several variations. The DVI standard was primarily designed to support digital devices; however, for backward compatibility, it can also support analog devices. The DVI-D (digital) connector supports only digital devices, whereas the DVI-I (integrated) connector supports both digital and analog devices via the addition of extra pins. Figure 12.5 and Table 12.10 show the DVI-I (integrated) connector and pinout and pinout.

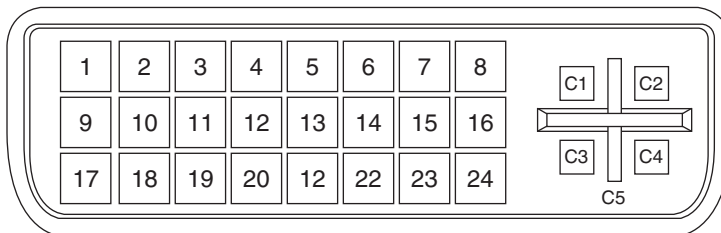


Figure 12.5 The DVI-I connector.

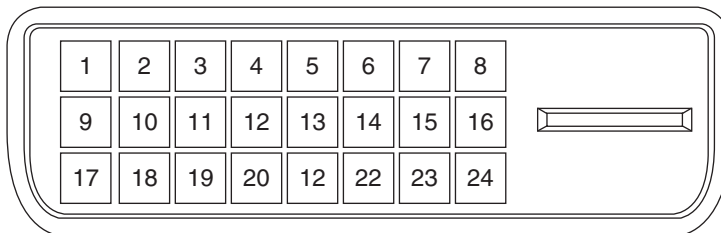
Table 12.10 DVI-I Connector Pinout

Pin	Signal	Pin	Signal
1	TMDS Data2-	13	TMDS Data3+
2	TMDS Data2+	14	+5V Power
3	TMDS Data2/4 Shield	15	GND
4	TMDS Data4-	16	Hot Plug Detect
5	TMDS Data4+	17	TMDS Data0-
6	DDC Clock	18	TMDS Data0+
7	DDC Data	19	TMDS Data0/5 Shield
8	Analog Vert. Sync	20	TMDS Data5-
9	TMDS Data1-	21	TMDS Data5+
10	TMDS Data1+	22	TMDS Clock Shield
11	TMDS Data1/3 Shield	23	TMDS Clock+
12	TMDS Data3-	24	TMDS Clock-
—	—	—	—
C1	Analog Red	C3	Analog Blue
C2	Analog Green	C4	Analog Horiz. Sync
—	—	C5	Analog GND

TMDS = Transition Minimized Differential Signaling

Note: The DVI-D (digital only) connector lacks the analog C1–C4 pins.

The DVI-D connector is the same as the DVI-I connector, except that it lacks the analog connections. By virtue of the unique MicroCross connector design, a digital-only device can connect only to receptacles with digital support, and an analog-only device can plug in only to receptacles with analog support. This design feature ensures that an analog-only device cannot be connected to a digital-only receptacle, and vice versa. Figure 12.6 shows the DVI-D connector. The pinout is the same as the DVI-I connector, except for the missing analog signals. The DVI-D connector is widely used on laptop port replicators and docking stations that provide DVI support.

**Figure 12.6** The DVI-D connector.

The DVI-I connector shown in Figure 12.5 can be converted into a VGA port for use with CRTs or with analog LCD panels via a simple adapter. Often new graphics cards purchased at retail that support only DVI come with just such an adapter that allows you to connect a traditional VGA connector from the display to the adapter.

Unfortunately, the Digital Display Working Group (DDWG) that created DVI has disbanded, leaving DVI frozen in time at the DVI 1.0 specification level. This means that DVI will not be updated in the future. Although it has enjoyed tremendous popularity as the first widely used digital display interface, the PC industry as a whole is slowly moving to DisplayPort as the replacement for DVI. Many cards feature both DVI and DisplayPort ports, or HDMI and DisplayPort ports.

HDMI

HDMI was designed by a group of multimedia companies (Hitachi, Panasonic, Philips, Silicon Image, Sony, Thomson, and Toshiba) as a way to provide a single-cable connection for transporting digital video and audio signals between consumer electronics hardware such as big-screen TVs, video games, DVD players, and home theater systems. HDMI was introduced in December 2002, version 1.3a was introduced in November 2006, version 1.4 was introduced in June 2009, and version 1.4a was introduced in March 2010.

HDMI is basically a superset of DVI and uses the same TMDS (Transition Minimized Differential Signaling) as does DVI. Unlike DVI, however, each color channel also carries multiplexed audio data. HDMI 1.2a and earlier supports a maximum data clock rate of 165MHz, sending 10 bits per cycle, or 1.65Gbps per channel. There are three channels per link, resulting in a maximum raw bandwidth of 4.95Gbps. Because the data is sent using 8b/10b encoding, only 8 bits of every 10 are actual data, resulting in a true data throughput of 3.96Gbps. This enables a single-link HDMI 1.2a or earlier connection to easily handle computer video resolutions as high as WUXGA (1920×1200) as well as 1080p HDTV (192×1080 with progressive scan) plus audio data.

HDMI 1.3 increases the maximum clock rate to 340MHz, resulting in 10.2Gbps raw bandwidth, or a true data throughput of 8.16Gbps. This increase allows a single-link HDMI connection to have slightly more throughput than a dual-link DVI connection, which handles computer resolutions as high as WQUXGA (3840×2400) plus audio data.

HDMI 1.4 adds support for Ethernet connections through the HDMI cable, audio return, 3D Blu-ray video playback, all current as well as future resolutions up to 4096×2160, better color spaces support, a micro-HDMI connector, new cable standards, and support for automotive connectors. HDMI 1.4 also introduces cable speed and feature categories: Standard (up to 1080i), High Speed (1080p), Standard HDMI with Ethernet, High Speed HDMI with Ethernet.

HDMI 1.4a adds support for new side-by-side and top-and-bottom 3D broadcast standards.

HDMI can also carry up to eight channels of uncompressed digital audio at 24-bit/192KHz along with Dolby Digital, DTS, Dolby TrueHD, and DTS-HD Master Audio compressed audio formats. Because it uses a single cable for both audio and video signals, HDMI provides an excellent way to reduce the cabling tangle present in home theater systems that use conventional analog audio and video cables. For home theater users who subscribe to HDTV satellite or cable services, HDMI is ideal because it supports high-bandwidth digital content protection (HDCP), which these services use to protect content from piracy while still assuring high-quality viewing and listening. To avoid reduced-quality playback of protected content, all devices (including the DVD player or set-top box, AV receiver, and display) must support HDCP.

Besides transmitting high-quality audio and video between devices, HDMI carries additional signals. HDMI uses the DDC to identify the capabilities of an HDMI display, such as resolutions, color depth, and audio. DDC enables optimal playback quality on different devices. HDMI also supports the optional consumer electronic control (CEC) feature, which enables one-button control of all CEC-enabled devices for one-touch play or record or other features.

Table 12.11 compares the HDMI versions.

Table 12.11 HDMI Versions

HDMI Version	Release Date	Notes
1.0	December 2002	Original release.
1.1	May 2004	Added DVD audio support.
1.2	August 2005	Added Super Audio CD support, PC support for Type A connector and native sRGB color-space, and requirement for displays to support low-voltage sources.
1.2a	December 2005	Added specifications for CEC features.
1.3	June 2006	Increased single-link bandwidth to 10.2Gbps raw (8.16Gbps actual data). Added support for 30/36/48-bit color, Dolby TrueHD and DT-HD Master Audio support, auto lip sync, and compact Type C connector.
1.3a	November 2006	Made minor adjustments to CEC and other specifications. Added new compliance test.
1.4	June 2009	10/100 Ethernet support (requires new cable); audio return channel; 3D input/output protocols; support for future 4K resolutions; new color spaces: sYCC601, Adobe RGB, and Adobe YCC601; HDMI Micro connector (also known as Type D); cable speed and feature standards.
1.4a	March 2010	Support for additional Side-by-Side and Top-and-Bottom broadcast 3D standards.

HDMI versions 1.3b (March 2007), 1.3b1 (November 2007), and 1.3c (August 2008) added HDMI compliance testing revisions, and they have no effect on HDMI features, functions, or performance. The latest specification for HDMI adopters is still 1.3a.

Because HDMI is essentially a superset of DVI, it is backward-compatible with DVI. This means that using a simple and inexpensive adapter, you can connect an HDMI source to a DVI display and connect a DVI source to an HDMI display. However, unless both the source and monitor support HDCP, you might not be able to play premium HDTV content, or the resolution might be reduced. Although some graphics cards claimed HDCP support as early as the first part of 2006, changes in the HDCP standard may prevent early cards from working properly. You should contact your graphics card and monitor vendor to determine whether a particular device supports HDCP.

Current HDMI cables correspond to HDMI Type A, Type C or Type D (introduced with HDMI 1.4). Type A is a 19-pin connector. Type C is a smaller version of Type A, originally designed for use in DV camcorders or other portable devices, but also used on some video cards. It uses the same pinout, and Type A-to-Type C adapters are available from various vendors. HDMI version 1.0 also defined a 29-pin Type B dual-link cable that has not been used in products. Type D is also a 19-pin connector that is about half the height of Type A and less than half the width of Type A, and is also shorter and narrower than Type C connectors.

Figure 12.7 illustrates a typical HDMI Type A cable and the location of pin 1 on the cable and connector.

The pinout for HDMI Type A and Type C cables is shown in Table 12.12. Type D uses the same pinout as Type A cables.

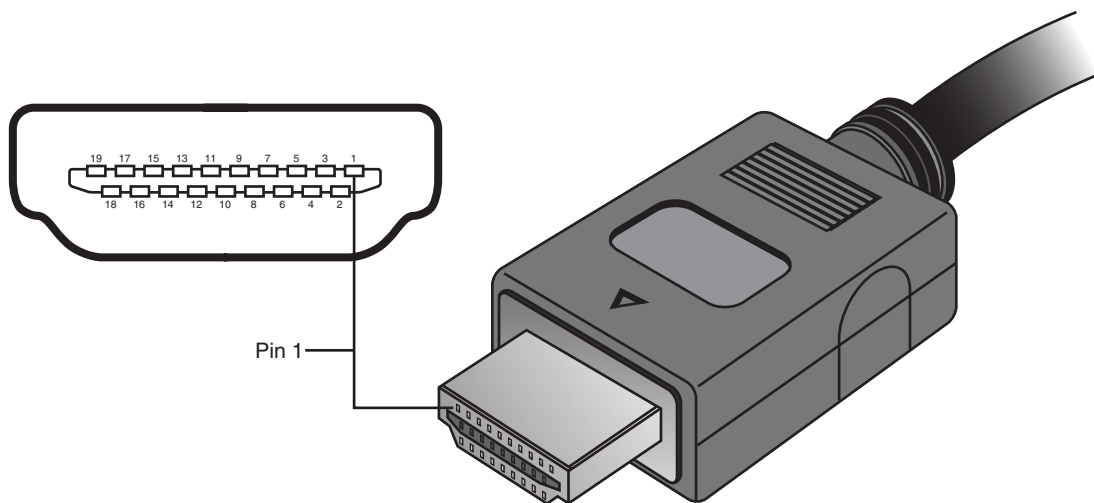


Figure 12.7 HDMI Type A cable and socket use a two-row 19-pin interface.

Table 12.12 HDMI Type A/Type C Connector Pinout

Pin Number	Description	Pin Number	Description
1	TMDS Data2 +	11	TMDS Clock Shield
2	TMDS Data2 Shield	12	TMDS Clock -
3	TMDS Data2 -	13	CEC
4	TMDS Data1 +	14	Reserved
5	TMDS Data1 Shield	15	SCL
6	TMDS Data1 -	16	SDA
7	TMDS Data0 +	17	DDC/CEC ground
8	TMDS Data0 Shield	18	+5V power
9	TMDS Data0 -	19	Hot Plug Detect
10	TMDS Clock +	—	—

Note: +5V power is 50mA max.

Figure 12.8 illustrates a typical HDMI-DVI adapter cable.

Note

The adapter cable shown in Figure 12.8 is not designed to work with graphics cards and drivers that do not support HDTV resolutions and timings. You may need to upgrade your graphics card driver before using an HDMI-DVI cable. Although some set-top boxes include DVI ports, this type of adapter cable is only intended for PC-HDTV connections.

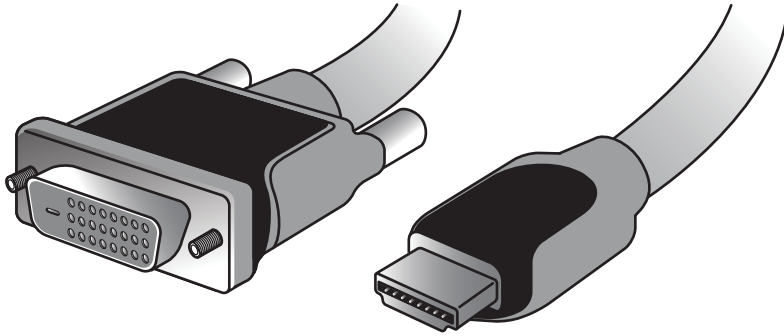


Figure 12.8 HDMI-DVI adapter cable.

Starting in late 2006, some vendors began to release PCIe cards including HDMI ports. Some provide HDMI input and output for use with HDV camcorders, whereas others using ATI or NVIDIA chipsets are graphics cards that also include HDMI output. Although HDMI is a royalty-based interface, requiring an annual license fee of \$10,000 plus a payment of 4 cents per device, and the newer DisplayPort interface is royalty-free, HDMI continues to be used on a wider range of current video cards and systems with integrated graphics than the obsolescent DVI or emerging DisplayPort standards. Some recent cards, such as the one shown in Figure 12.1, use the mini-HDMI (HDMI Type C) connector to save space on the card bracket.

For more information about HDMI, see the HDMI Founders website at www.hdmi.org.

DisplayPort

DisplayPort is the latest digital display interface standard. It is designed to replace VGA, DVI, and HDMI for use in PCs and to coexist with HDMI in consumer electronics devices. Dell originated the design in 2003 and then turned it over to VESA in August 2005. In May 2006, VESA published it as an open industry standard.

DisplayPort is designed to replace all the previous digital and analog interfaces, including DVI, HDMI, and even VGA. In addition, it is a royalty-free interface and does not incur the licensing fees of HDMI or the implementation patent fees of DVI. Also, DisplayPort is designed both as an internal and an external interface, meaning it can replace the FPD-Link (Flat Panel Display-Link) interface used internally in most laptops as well as providing connections to external displays. In short, DisplayPort is designed to be the ultimate universal display interface for PCs now as well as in the future.

Previous digital display interfaces such as DVI and HDMI used TMDS, which requires extra logic on both the source and display ends, logic that must usually be licensed from Silicon Image. DisplayPort instead uses a packetized (network-like) interface that can easily be implemented in chipsets without the extra cost logic required for DVI or HDMI. DisplayPort is kind of like a high-speed Ethernet for video, and the network-like design allows for features such as multiple video streams over a single connection, which means you can connect multiple displays to a single port.

Because it is a license-free, royalty-free design, DisplayPort has seen rapid adoption throughout the industry. In fact, all new chipsets and GPUs since 2008 from Intel, NVIDIA, and AMT/ATI already have integrated DisplayPort support. In 2008, major manufacturers including Dell, HP/Compaq, Lenovo, and Apple introduced products with DisplayPort and endorsed DisplayPort as the successor to DVI and HDMI for most digital display connections. Despite this endorsement, the introduction of DisplayPort-enabled products has proceeded slowly. However, DisplayPort is currently available on an increasing number of laptops, 23 inch or larger LCD and backlit LED displays, computers, and graphics cards for desktop and technical workstation use.

On the technical side, DisplayPort is a high-speed serial interface with up to four main data lanes (differential signal pairs) carrying multiplexed video and audio data, each of which supports a raw data rate of 1.62Gbps, 2.7Gbps, or 5.4Gbps (DisplayPort 1.2 or later only). Using all four lanes results in a maximum raw bandwidth of 6.48Gbps, 10.8Gbps, or 21.6Gbps, respectively. Because 8b/10b encoding is used, only 8 bits of every 10 are data, resulting in maximum true data throughputs of 5.184Gbps, 8.64Gbps, or 17.28Gbps, respectively.

Audio is optional, with support for up to eight channels of 16- or 24-bit linear PCM data at a 48KHz, 96KHz, or 192KHz sampling rate, with an uncompressed maximum audio bandwidth of 6.144Mbps.

DisplayPort 1.1 includes the following features:

- Small external connectors (slightly larger than universal serial bus [USB] size) with optional latching. Four display connectors can fit on a single PCIe expansion card bracket and fit easily in laptops.
- Cable lengths of up to 15 meters (49 feet), which allows for remote displays or projectors.
- Micro-packet network architecture over one to four lanes. Connections can use only as many lanes as necessary for reduced wire counts.
- High performance. A true data bandwidth of 8.64Gbps (four lanes at 2.16Gbps per lane) allows WQXGA 2560×1600 resolution.
- Support for internal (embedded) as well as external LCD connections. This allows a universal interface for both desktop and laptop systems.
- Optional audio that supports displays with built-in speakers.
- Optional HDCP to allow playing protected media.
- Interoperability with DVI and HDMI over a DisplayPort connector. You can connect to DVI or HDMI with simple and inexpensive adapters.
- An auxiliary 1Mbps channel, which allows for two-way communication for integrated cameras, microphones, and so on.
- A powered connector, which powers some LCD displays directly.
- An optional latching connector that uses a simple thumb-press release design with no bolts or jackscrews.

DisplayPort 1.2 is fully backward compatible with 1.1 and adds the following features:

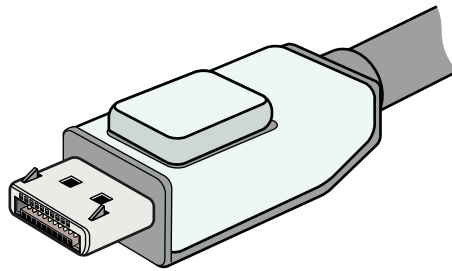
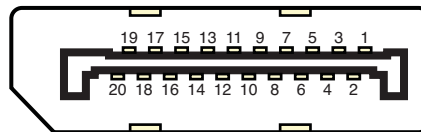
- Double the performance. DisplayPort 1.2 offers 21.6Gbps raw (17.28Gbps true) bandwidth, which is more than twice that of HDMI 1.3a and nearly triple that of DVI.
- Multiple data streams, which allows support for two WQXGA 2560×1600 or four WUXGA 1920×1200 monitors daisy-chained over a single cable.
- An auxiliary channel speed increase to 480Mbps. This allows USB 2.0 speed connections for cameras, microphones, or other devices.
- The Mini DisplayPort connector. This connector is approximately half the size yet provides full functionality for laptops or other devices where space is at a premium.

Table 12.13 compares the versions of DisplayPort.

Table 12.13 DisplayPort Versions

Version	Release Date	Notes
DisplayPort 1.0	May 2006	Initial release
DisplayPort 1.1	March 2007	Introduced “hybrid device” class, added support for HDCP content protection, and increased connector power requirements
DisplayPort 1.1a	January 2008	Corrected errata and clarified minor items
DisplayPort 1.2	Late-2009	Doubled performance, multiple data streams, a high-speed auxiliary channel, and the Mini DisplayPort connector

The DisplayPort connector has 20 pins and is only slightly larger than USB size (15.9mm vs. 12mm wide). The pins consist of four data lanes (differential pairs), an auxiliary channel (differential pair), plus configuration and power pins. Figures 12.9 and 12.10 show the DisplayPort cable/plug and socket.

**Figure 12.9** DisplayPort cable with latching plug connector (Belkin).**Figure 12.10** DisplayPort socket and pin configuration.

Apple introduced the Mini DisplayPort connector in October 2008, which was subsequently included as part of the official DisplayPort standard in 1.2 and later releases. The Mini DisplayPort connector has the same full complement of 20 pins as the standard DisplayPort connector, but it's about half the size (at only 7.4mm wide). Figures 12.11 and 12.12 show the Mini DisplayPort cable/plug and socket. Table 12.14 shows the DisplayPort socket connector pinout.

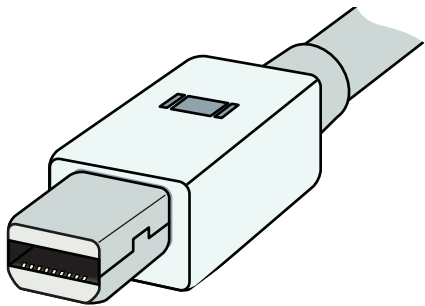


Figure 12.11 Mini DisplayPort cable and plug (Apple).

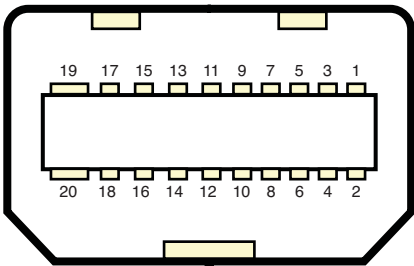


Figure 12.12 Mini DisplayPort socket and pin configuration.

Table 12.14 DisplayPort Socket Connector Pinout

Top Row		Bottom Row	
Pin Number	Description	Pin Number	Description
Pin 1	Lane 0+	Pin 2	Ground
Pin 3	Lane 0-	Pin 4	Lane 1+
Pin 5	Ground	Pin 6	Lane 1-
Pin 7	Lane 2+	Pin 8	Ground
Pin 9	Lane 2-	Pin 10	Lane 3+
Pin 11	Ground	Pin 12	Lane 3-
Pin 13	Config1	Pin 14	Config2
Pin 15	Aux. CH+	Pin 16	Ground
Pin 17	Aux. CH-	Pin 18	Hot Plug Detect
Pin 19	Power Return	Pin 20	+3.3V Power

This pinout is for source-device connectors. Sink-device connectors will have lanes 0–3 reversed in order (that is, lane 3 will be on pins 1 and 3, lane 0 will be on pins 10 and 12, and so on).
+3.3V power at 500mA max.

VESA has created several icons and logos associated with DisplayPort. The basic DisplayPort icon is used to label products incorporating DisplayPort technology, whereas the DisplayPort Certification Compliance logo is used on product marketing material to indicate devices that have been tested to

ensure they are fully interoperable with other DisplayPort devices. Figure 12.13 shows the DisplayPort Certification Compliance logo. VESA maintains a list of certified devices on the www.displayport.org website.

The DisplayPort Multimode icon adds two “+” symbols to indicate a port or device that is fully backward compatible with both DVI and HDMI technology (via inexpensive cable adapters). Figure 12.14 shows the DisplayPort Multimode icon. Figure 12.15 shows an inexpensive DisplayPort to DVI adapter.



Figure 12.13 DisplayPort Certification Compliance logo (left).



Figure 12.14 Icon indicating a DisplayPort with Multimode (DVI and HDMI) support (right).

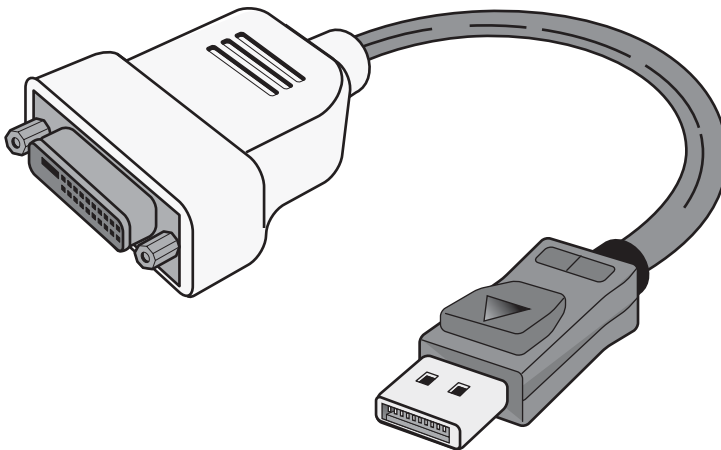


Figure 12.15 DisplayPort to DVI adapter, which works on MultiMode DisplayPort connectors.

When DisplayPort was released, many people wondered why we needed another digital display interface when we already had DVI and HDMI. Unfortunately, those interfaces carry both technical and licensing limitations that were preventing their universal adoption. DisplayPort is designed to overcome not only the technical limitations, but especially the licensing constraints and fees that the other interfaces brought along as baggage. The advanced technical capabilities of DisplayPort, including its ability to work with the new Thunderbolt I/O technology, combined with the elimination of licensing and its backward compatibility with existing interfaces, are likely to ensure its rapid adoption throughout the PC marketplace.

TV Display Interfaces

When video technology was introduced, it was based on television. However, a difference exists between the signals used by a television and those used by a computer monitor. In the United States, the National Television System Committee (NTSC) established color TV standards in 1953. Some other countries, such as Japan, followed this standard. Many countries in Europe, though, developed more sophisticated standards, including Phase Alternate Line (PAL) and Sequential Couleur Avec Mémoire (SECAM). Table 12.15 shows the differences among these standards.

Table 12.15 Television Versus Computer Monitors

Standard	Year Est.	Country	Lines	Rate
<i>Television</i>				
NTSC	1953 (color) 1941 (B&W)	U.S., Japan	525	60 fields/sec
PAL	1941	Europe ¹	625	50 fields/sec
SECAM	1962	France	625	25 fields/sec
HDTV ²		Various countries	1080p 720p	25 fields/sec 60 frames/sec
NTSC	1953 (color) 1941 (B&W)	U.S., Japan	525	60 fields/sec
PAL	1941	Europe ¹	625	50 fields/sec
SECAM	1962	France	625	25 fields/sec
HDTV ²		Various countries	1080i 1080/720p	25 fields/sec 60 frames/sec
<i>Computer</i>				
VGA	1987	U.S.	640×480 ³	72Hz

Field = 1/2 (0.5 frame)

1. England, Holland, and West Germany.
2. Various analog and digital HDTV standards have been introduced around the world, starting in France in 1948 (SECAM 755i, a now-discontinued analog version); Japan's analog MUSE 1035i is the oldest system (1979–present) in use. The U.S. standards (both digital) are listed above; other countries use various analog and digital standards.
3. VGA is based on more lines and uses pixels (480) versus lines; genlocking is used to lock pixels into lines and synchronize computers with TV standards.

To display computer screens on standard TV or to record them to videotape, you can use the TV-out connector on your video card or integrated motherboard video. TV-out is generally S-video or composite video. The resolution that these devices display on a TV set or record on videotape often is limited to VGA (640×480) or SVGA (800×600).

To connect your PC to an HDTV monitor, it is preferable to use a digital signal via a DVI, HDMI, or DisplayPort connection. If your current video adapter only has analog VGA output, you'll want to upgrade to a video adapter with a DVI, HDMI, or DisplayPort digital output. Because most HDTVs use HDMI, if your video card has DVI or DisplayPort, you can use a DVI-to-HDMI or DisplayPort-to-HDMI adapter if necessary. If you need HDCP support for watching HD premium content, make sure your display and card support HDCP. Otherwise, you may not be able to watch the program or it may be displayed at reduced resolution.

3D Graphics Accelerators

Since the late 1990s, 3D acceleration—once limited to exotic add-on cards designed for hardcore game players—has become commonplace in the PC world. With the introduction of the Aero desktop in Windows Vista and later, 3D imaging is even utilized in the user interface, joining other full-motion 3D uses such as sports, first-person shooters, team combat, driving, and many other types of PC gaming. Because even low-cost integrated chipsets offer some 3D support, virtually any user of a recent-model computer can enjoy 3D lighting, perspective, texture, and shading effects.

Note

At a minimum, enabling the Windows Aero interface in Vista and later requires graphics hardware that supports DirectX 7 3D graphics; however, for maximum functionality, graphics hardware that supports DirectX 9 or greater is required. Games are now being released that require DirectX 11, which is not available for Windows XP and earlier versions.

How 3D Accelerators Work

To construct an animated 3D sequence, a computer can mathematically animate the sequences between keyframes. A keyframe identifies a specific point. A bouncing ball, for example, can have three keyframes: up, down, and up. Using these frames as reference points, the computer can create all the interim images between the top and bottom. This creates the effect of a smoothly bouncing ball.

After it has created the basic sequence, the system can then refine the appearance of the images by filling them in with color. The most primitive and least effective fill method is called *flat shading*, in which a shape is simply filled with a solid color. *Gouraud shading*, a slightly more effective technique, involves the assignment of colors to specific points on a shape. The points are then joined using a smooth gradient between the colors.

A more processor-intensive (and much more effective) type of fill is called *texture mapping*. The 3D application includes patterns—or textures—in the form of small bitmaps that it tiles onto the shapes in the image, just as you can tile a small bitmap to form the wallpaper for your Windows desktop. The primary difference is that the 3D application can modify the appearance of each tile by applying perspective and shading to achieve 3D effects. When lighting effects that simulate fog, glare, directional shadows, and others are added, the 3D animation comes close indeed to matching reality.

Until the late 1990s, 3D applications had to rely on support from software routines to convert these abstractions into live images. This placed a heavy burden on the system processor in the PC, which has a significant impact on the performance not only of the visual display, but also of any other applications the computer might be running. Starting in the period from 1996 to 1997, chipsets on most video adapters began to take on many of the tasks involved in rendering 3D images, greatly lessening the load on the system processor and boosting overall system performance.

There have been roughly 11 generations of 3D graphics hardware on PCs, a process that has lasted more than a decade, as detailed in Table 12.16.

Table 12.16 Brief History of 3D Acceleration

Generation	Dates	Technologies	Example Product/ Chipset
1st	1996–1997	3D PCI card with passthrough to 2D graphics card; OpenGL and GLIDE APIs	3dfx Voodoo
2nd	1997–1998	2D/3D PCI card	ATI Rage, NVIDIA RIVA 128
3rd	1999	2D/3D AGP 1x/2x card	3dfx Voodoo 3, ATI Rage Pro, NVIDIA TnT2
4th	1999–2000	DirectX 7 API, AGP 4x	NVIDIA GeForce 256, ATI Radeon
5th	2001	DirectX 8 API, programmable vertex and pixel shaders	NVIDIA GeForce 3, NVIDIA GeForce 4 Ti
6th	2001–2002	DirectX 8.1 API	ATI Radeon 8500, ATI Radeon 9000
7th	2002–2003	DirectX 9 API, AGP 8x	ATI Radeon 9700, NVIDIA GeForce FX 5900
8th	2004–2005	PCIe, DirectX 9.0c	ATI X800, NVIDIA GeForce 6800
9th	2004–2006	Dual-GPU rendering with PCIe x8, x16	ATI X1900, NVIDIA GeForce 7800; ATI CrossFire, NVIDIA SLI motherboard chipsets and compatible cards
10th	2007–2008	DirectX 10, Windows Vista	ATI HD 2xxx-4xxx, NVIDIA GeForce 8-9 series
11th	2009+	DirectX 11, Windows 7	AMD Radeon HD 5xxx-6xxx, NVIDIA GeForce 2xx-5xx series

With virtually every recent graphics card on the market featuring DirectX 10 or greater capabilities, you don't need to spend a fortune to achieve a reasonable level of 3D graphics. Many cards in the \$50–\$150 range use lower-performance variants of current high-end GPUs, or they might use the previous year's leading GPU. These cards typically provide more-than-adequate performance for 2D business applications. Most current 3D accelerators also support dual-display and TV-out capabilities, enabling you to work and play at the same time.

However, keep in mind that the more you spend on a 3D accelerator card, the greater the onboard memory and faster the accelerator chip you can enjoy. If money is no object, you can buy a graphics card featuring its fastest GPU for more than \$500. Fortunately, there are plenty of choices using either NVIDIA or AMD GPUs in the less than \$500 price range that offer plenty of 3D gaming performance, including support for dual-GPU operations (NVIDIA SLI or AMD CrossFire X), which split rendering chores across the GPUs in both video cards for faster game display than with a single card. GPUs that support DirectX 10/11 are the preferred choice for anyone who wants to play current generation games.

Mid-range cards costing \$100–\$300 are often based on GPUs that use designs similar to the high-end products but might have slower memory and core clock speeds or a smaller number of rendering pipelines. These cards provide a good middle ground for users who play games fairly often but can't cost-justify high-end cards.

The basic function of 3D software is to convert image abstractions into the fully realized images that are then displayed on the monitor. The image abstractions typically consist of the following elements:

- **Vertices**—Locations of objects in three-dimensional space, described in terms of their x, y, and z coordinates on three axes representing height, width, and depth.

- **Primitives**—The simple geometric objects the application uses to create more complex constructions, described in terms of the relative locations of their vertices. This serves not only to specify the location of the object in the 2D image, but also to provide perspective because the three axes can define any location in three-dimensional space.
- **Textures**—Two-dimensional bitmap images or surfaces designed to be mapped onto primitives. The software enhances the 3D effect by modifying the appearance of the textures, depending on the location and attitude of the primitive. This process is called *perspective correction*. Some applications use another process, called *MIP mapping*, that uses different versions of the same texture containing varying amounts of detail, depending on how close the object is to the viewer in the three-dimensional space. Another technique, called *depth cueing*, reduces the color and intensity of an object's fill as the object moves farther away from the viewer.

Using these elements, the abstract image descriptions must then be rendered, meaning they are converted to visible form. Rendering depends on two standardized functions that convert the abstractions into the completed image that is displayed onscreen. The standard functions performed in rendering are as follows:

- **Geometry**—The sizing, orienting, and moving of primitives in space and the calculation of the effects produced by the virtual light sources that illuminate the image
- **Rasterization**—The converting of primitives into pixels on the video display by filling the shapes with properly illuminated shading, textures, or a combination of the two

A modern video adapter that includes a chipset capable of 3D video acceleration has special built-in hardware that can perform the rasterization process much more quickly than if it were done by software (using the system processor) alone. Most chipsets with 3D acceleration perform the following rasterization functions right on the adapter:

- **Scan conversion**—The determination of which onscreen pixels fall into the space delineated by each primitive
- **Shading**—The process of filling pixels with smoothly flowing color using the flat or Gouraud shading technique
- **Texture mapping**—The process of filling pixels with images derived from a 2D sample picture or surface image
- **Visible surface determination**—The identification of which pixels in a scene are obscured by other objects closer to the viewer in three-dimensional space
- **Animation**—The process of switching rapidly and cleanly to successive frames of motion sequences
- **Antialiasing**—The process of adjusting color boundaries to smooth edges on rendered objects

Note

To learn more about 3D techniques such as fogging, Z-buffering, T-buffering, programmable shading, and rendering methods, see the sections "Typical 3D Techniques," "Advanced 3D Filtering and Rendering," and "Single- Versus Multi-Pass Rendering," in Chapter 12, "Video Hardware," in *Upgrading and Repairing PCs*, 19th edition, included in its entirety on the disc packaged with this book.

Software Optimization

It's important to realize that the presence of an advanced 3D-rendering feature on any given video card is meaningless unless game and application software designers optimize their software to take advantage of the feature. Although various 3D standards exist (OpenGL and DirectX), video card makers provide drivers that make their games play with the leading standards. Because some cards do play better with certain games, you should read the reviews in publications and websites to see how your favorite graphics card performs with them. Typically, it can take several months or longer after a new version of DirectX or OpenGL is introduced before 3D games take full advantage of the 3D rendering features provided by the new application programming interface (API).

Some video cards allow you to perform additional optimization by adjusting settings for OpenGL, Direct 3D, RAMDAC, and bus clock speeds, as well as other options. Note that the bare-bones 3D graphics card drivers provided as part of Microsoft Windows usually don't provide these dialog boxes. Be sure to use the drivers provided with the graphics card or download updated versions from the graphics card or GPU vendor's website.

APIs

APIs provide hardware and software vendors a means to create drivers and programs that can work quickly and reliably across a variety of platforms. When APIs exist, drivers can be written to interface with the API rather than directly with the OS and its underlying hardware.

Currently, the leading game APIs include SGI's OpenGL and Microsoft's Direct3D (part of DirectX). OpenGL and Direct3D are available for virtually all leading graphics cards. At one time, a third popular game API was Glide, an enhanced version of OpenGL that is restricted to graphics cards that use 3dfx chipsets, which are no longer on the market.

OpenGL

The latest version of OpenGL is version 4.1, released on July 26, 2010. OpenGL 4.1 is based on OpenGL 3.x, and adds 64-bit double-precision floating point support, new shader stages and subroutines for faster and better quality of 3D rendering, and better performance. OpenGL 4.1 is backwards-compatible with OpenGL 3.x.

Although OpenGL is a popular gaming API, it is also widely used in 3D rendering for specialized business applications, including mapping, life sciences, and other fields. Windows XP and newer can support OpenGL either through software or through hardware acceleration. For a particular graphics card to support hardware acceleration of OpenGL, the driver developer must include an installable client driver (ICD). The ICD is distributed as part of the driver package that the video card or GPU vendor provides. Thus, driver updates can improve OpenGL performance as well as DirectX (Direct3D) performance.

To learn more about OpenGL, see the OpenGL website at www.opengl.org.

Microsoft DirectX

Direct3D is part of Microsoft's comprehensive multimedia API, DirectX. Although the most recent versions of DirectX (9.0c and up) provide support for higher-order surfaces (converting 3D surfaces into curves), vertex shaders, and pixel shaders, significant differences exist between DirectX versions in how these operations are performed.

DirectX 9.0c uses separate pixel and vertex shaders to create 3D objects. Although DirectX 9.0c provides greater precision in data handling as well as support for more instructions, more textures, and more registers than its predecessors, its use of separate shaders can still lead to slow 3D rendering when more pixels must be rendered than shaders, or vice versa. Shader Model 3 (used by DirectX 9.0c) is simply a development of the split-function design first developed for Shader Model 1 (used by DirectX 8.0) back in 2001, adding support for more instructions and greater numerical accuracy.

DirectX 10, developed for Windows Vista, includes a completely rebuilt Direct3D rendering engine with a brand-new shader design, Shader Model 4. Shader Model 4 adds a geometry shader to the vertex shader and pixel shader design used in earlier shader models to improve the handling of real-time scene changes such as explosions. However, the biggest single change in Shader Model 4 is the use of unified shaders that can be switched between vertex, pixel, and geometry shader operations on the fly, eliminating bottlenecks and improving performance, no matter what types of 3D data exist in a scene.

Note

With the replacement of dedicated vertex and pixel shaders in the DirectX 10 3D rendering pipeline, DirectX 10 GPUs are rated in terms of the number of stream processors on board. Each stream processor performs vertex, geometry, and pixel shading as needed.

When you are comparing two otherwise-equal DirectX 10 GPUs (same GPU, memory size and speed, motherboard and memory bus designs), the GPU with a larger number of stream processors will be faster.

Other architectural changes in DirectX 10 include process optimizations to reduce the load on the CPU. In a sample of different types of images rendered, DirectX 10 reduced the command cycles by as much as 90% over DirectX 9.

DirectX 11 was originally developed for Windows 7 (and is also available for Windows Vista) and adds several new features:

- **Tessellation**—Provides additional pipeline stages that increase the number of visible polygons at runtime.
- **Multithreaded rendering**—Enables the execution of Direct3D commands on multiple processor cores.
- **Compute shaders**—Provides an additional stage independent of the Direct3D pipeline that enables general-purpose computing on the graphics processor, part of Shader Model 5. This feature is supported by processor with integrated graphics.
- **Dynamic shader linkage**—A limited runtime shader linkage that allows for improved shader specialization during application execution.

To obtain a version of DirectX 11 for Windows Vista, see KB971512 at <http://support.microsoft.com> for details.

It's important to realize that DirectX 10/11 GPUs retain full compatibility with DirectX 9.0c and earlier DirectX versions, so you can play the latest games as well as old favorites with a DX10 or DX11-compliant video card. Updates for DirectX are provided via www.windowsupdate.com. More information about DirectX is available from Microsoft at www.microsoft.com/directx.

Dual-GPU Scene Rendering

In Table 12.16, I placed the development of dual PCIe graphics card solutions as the ninth generation of 3D acceleration. The ability to connect two cards to render a single display more quickly isn't exactly new: The long-defunct 3dfx Voodoo 2 offered an option called *scan-line interfacing* (SLI) that pairs two Voodoo 2 cards on the PCI bus, with each card writing half the screen in alternating lines. With 3dfx's version of SLI, card number one wrote the odd-numbered screen lines (one, three, five, and so on), whereas card number two wrote the even-numbered screen lines (two, four, six, and so on). Although effective, use of SLI with Voodoo 2 was an expensive proposition that only a handful of deep-pocketed gamers took advantage of.

A few companies also experimented with using multiple GPUs on a single card to gain a similar performance advantage, but these cards never became popular. However, the idea of doubling graphics performance via multiple video cards has proven too good to abandon entirely, even after 3dfx went out of business.

NVIDIA SLI

When NVIDIA bought what was left of 3dfx, it inherited the SLI trademark and, in mid-2004, reintroduced the concept of using two cards to render a screen under the same acronym. However, NVIDIA's version of SLI has a different meaning and much more intelligence behind it.

NVIDIA uses the term *SLI* to refer to scalable link interface. The *scaling* refers to load-balancing, which adjusts how much of the work each card performs to render a particular scene, depending on how complex the scene is. To enable SLI, you need the following components:

- A PCIe motherboard with an SLI-compatible chipset and two PCIe video slots designed for SLI operation
- Two NVIDIA-based video cards with SLI support

Note

Originally, you needed to use two identical cards for NVIDIA SLI. With the introduction of NVIDIA ForceWare v81.85 and higher driver versions, this is no longer necessary. Just as with the AMD CrossFire and CrossFire X multi-GPU solution, the cards need to be from the same GPU family, but they don't need to be from the same manufacturer. You can obtain updated drivers from your video card maker or from the NVIDIA website (www.nvidia.com).

In most cases, a special bridge device known as a *multipurpose I/O (MIO)* connects the cards. The MIO is supplied with SLI-compatible motherboards, but some SLI-compatible cards don't use it.

To learn more about SLI and for a list of SLI-compatible GPUs and nForce motherboard chipsets, visit NVIDIA's SLI Zone (<http://slizone.com>).

Figure 12.16 illustrates a typical SLI hardware configuration. Note the MIO device connecting the cards.

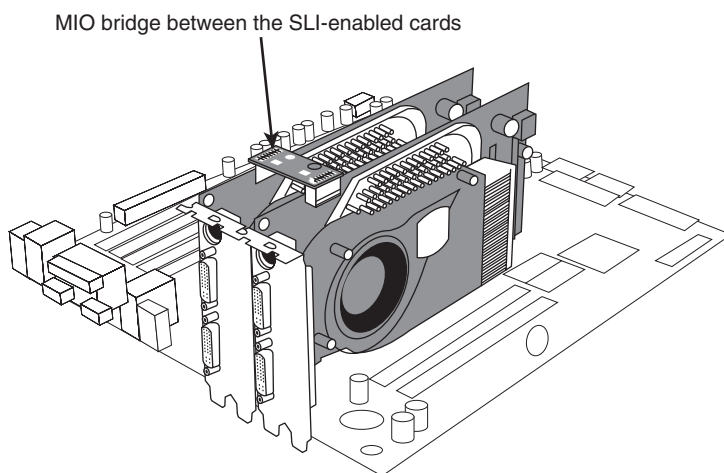


Figure 12.16 NVIDIA SLI installation using a multipurpose I/O (MIO) bridge device.

AMD CrossFireX

AMD's CrossFireX, an improved version of ATI's CrossFire multi-GPU technology, uses three methods to speed up display performance: alternate frame rendering, supertiling (which divides the scene into alternating sections and uses each card to render parts of the scene), and load-balancing scissor operation (similar to SLI's load-balancing). The ATI Catalyst driver uses alternate frame rendering for best performance, but it automatically switches to one of the other modes for games that don't work with alternate frame rendering.

To achieve better image quality than with a single card, CrossFire offers various SuperAA (antialiasing) modes, which blend the results of antialiasing by each card. CrossFire also improves anisotropic filtering by blending the filtering performed by each card.

To use CrossFireX, you need the following components:

- A PCIe motherboard with a CrossFireX-compatible chipset and two PCIe video slots designed for CrossFireX operation
- A supported combination of ATI CrossFireX-supported cards

Note

For specific models of motherboards, video cards, power supplies, memory, and cases designed to support CrossFireX, see <http://sites.amd.com/us/game/technology/Pages/crossfirex.aspx>.

First-generation CrossFire cards required users to buy special CrossFire Edition cards that contained the composting engine (an Xilinx XC3S400 chip) and used a proprietary DMS-59 port for interconnecting the cards. One of these cards was paired with a standard Radeon card from the same series via a clumsy external cable between the CrossFire Edition's DMS port and the DVI port on the standard card. Newer CrossFire-and CrossFireX compatible cards use the PCIe bus or a CrossFire bridge interconnect (similar in concept to the SLI MIO component) to connect matching cards.

CrossFire and CrossFireX can be disabled to permit multimonitor operation; CrossFire and CrossFireX cards can also be used to implement physics effects in games that use the HavokFX physics technology (www.havok.com).

Monitors

The monitors typically used with PCs come in a variety of sizes and resolutions and are typically based on one of two display technologies: liquid crystal display (LCD) or cathode-ray tube (CRT). Larger displays such as big-screen TVs and projectors can also use LCD or backlit LED technology but may use plasma or DLP technology as well. This section discusses the various features, specifications, and technologies used in PC monitors.

Display Specifications

A number of features and specifications differentiate one display from another. Some of these can be confusing, and some are more important than others. The following sections examine the features and specifications to look for when comparing or selecting displays.

Display Size

PC monitors come in various sizes, generally ranging from 15 inches to 30 inches in diagonal measurement. Displays smaller than 15 inches are available for specialized uses (and are often used on smaller laptop or palmtop/netbook systems). Displays larger than 30 inches are also available; however, these are generally categorized as large format multimedia displays or televisions rather than as

PC monitors. In general, the larger the monitor, the higher the price tag; however, there are often “sweet spots” where a certain size may have a price advantage over smaller sizes due to sales and manufacturing popularity.

Display sizes are measured diagonally, which is an artifact from the round tubes used in the first televisions, where the diagonal measurement was equal to the physical diameter of the tube. Whereas the diagonal display size measures the physical size of the display, the *viewable image size* refers to the diagonal measurement of the usable area on the screen (for example, the OS desktop). On an LCD or LED panel, the physical diagonal measurement and the viewable image size of the display are the same. With CRTs, however, the viewable image size is typically 1 inch smaller than the advertised diagonal display size. Therefore, when you’re comparing LCDs or LEDs and CRTs with the same diagonal size, the LCD or LED actually offers a larger viewable image.

Note

Many people became upset when they realized that the CRT monitors they were buying displayed images that were smaller than the advertised number. For example, a 17-inch CRT monitor would display only a 16-inch-sized image. This came to a head in the mid-1990s when several class-action lawsuits forced CRT monitor manufacturers to disclose the *viewable image size* in the immediate proximity of the physical diagonal measurement in any advertisements and specifications. LCDs were not affected by these lawsuits because LCDs have always had a viewable image size that is the same as the diagonal display measurement.

Tip

Larger, higher-resolution monitors retain their value longer than most other computer components. Although it’s common for a newer, faster processor to come out right after you have purchased your computer or to find the same model with a bigger hard disk for the same money, a good monitor can outlast your computer. If you purchase monitors with longer term usage considerations in mind, you can save money on your next system by reusing your existing monitor.

Resolution

Resolution indicates the amount of detail a monitor can render. This quantity is expressed in the number of horizontal and vertical picture elements, or *pixels*, contained in the screen. The total is usually expressed in the millions of pixels, or *megapixels*. As the resolution increases, the image consists of a greater number of pixels. With more pixels, you can see more of a given image, or you can view the image in greater detail.

As PC video technology has evolved, available resolutions have grown at a steady pace. Table 12.17 shows the industry standard resolutions (from lowest to highest) used in PC graphics adapters and displays as well as the terms or designations commonly used to describe them.

Table 12.17 Graphics Display Resolution Standards

Designation	Resolution (HxV)	Megapixels	Aspect Ratio
CGA	320×200	0.06	1.331
EGA	640×350	0.22	1.33 ¹
VGA	640×480	0.31	1.33
WVGA	854×480	0.41	1.78
SVGA	800×600	0.48	1.33
XGA	1024×768	0.79	1.33
HD 720	1280×720	0.92	1.78
WXGA ²	1280×768	0.98	1.67
XGA+	1152×864	1.00	1.33
WXGA ²	1280×800	1.02	1.60
WXGA ²	1366×768	1.05	1.78
WEXGA	1280×854	1.09	1.50
QVGA	1280×960	1.23	1.33
WXGA+	1440×900	1.30	1.60
SXGA	1280×1024	1.31	1.25
SXGA+	1400×1050	1.47	1.33
WSXGA-	1600×900	1.44	1.78
WSXGA	1600×1024	1.64	1.56
WSXGA+	1680×1050	1.76	1.60
UXGA	1600×1200	1.92	1.33
HD 1080	1920×1080	2.07	1.78
WUXGA	1920×1200	2.30	1.60
QXGA	2048×1536	3.15	1.33
WQXGA	2560×1600	4.10	1.60
QSXGA	2560×2048	5.24	1.25
WQSXGA	3200×2048	6.55	1.56
QUXGA	3200×2400	7.68	1.33
WQUXGA	3840×2400	9.22	1.60
HXGA	4096×3072	12.58	1.33
WHXGA	5120×3200	16.38	1.60
HSXGA	5120×4096	20.97	1.25
WHSXGA	6400×4096	26.21	1.56
HUXGA	6400×4800	30.72	1.33
WHUXGA	7680×4800	36.86	1.60

1. CGA and EGA used non-square pixels on a 4:3 display.

2. WXGA designates several slightly different resolutions.

Display adapters normally support a variety of resolutions; however, what a display can handle is usually much more limited. Therefore, the specific resolution you use is normally dictated by the display, not the adapter. A given video adapter and display combination will usually have an allowable maximum (usually dictated by the display), but it can also work at several resolutions less than the maximum. Because LCD and LED monitors are designed to run at a single native resolution, they must use electronics to scale the image to other choices. Older LCD panels handled scaling poorly, but even though current LCD and the new LED backlit panels perform scaling fairly well, you are almost always better off selecting the single specific resolution that is native to the display you are using. However, if the native resolution of the display is exceptionally high, you can choose lower resolutions to achieve larger and more readable icons and text.

Aspect Ratio

A given resolution has a horizontal and vertical component, with the horizontal component the larger of the two. The aspect ratio of a display is the ratio between the horizontal and vertical number of pixels. It is calculated as the width (in pixels) divided by the height. A number of different aspect ratios have been used in PC displays over the years. Most of the early display formats were only slightly wider than they were tall. More recently, wider formats have become popular. Aspect ratios of 1.5 or higher (numerically) are considered “widescreen” displays. Originally 1.33 (4:3) was the most common aspect ratio for displays because it matched the aspect ratio of the original standard definition televisions. The first generation of PC widescreen displays use a 1.60 (16:10 or 8:5) ratio, whereas 2nd generation displays use the 1.78 (16:9) ratio, which is also the most popular format for widescreen televisions. Although many PC users (myself included) preferred the additional vertical resolution of the 16:10 ratio over the 16:9, the LCD manufacturers found it more economical to produce the same 16:9 ratio display for PCs as they do for televisions, thus eliminating 16:10 versions as a choice. Table 12.18 shows the various aspect ratios and designations.

Table 12.18 Standard Aspect Ratios and Designations

Numerical Ratio	Fractional Ratio	Format Designation
1.25	5:4	Standard
1.33	4:3	Standard
1.50	3:2	Widescreen
1.56	25:16	Widescreen
1.60	16:10	Widescreen
1.67	5:3	Widescreen
1.78	16:9	Widescreen
1.83	11:6	Widescreen

1.50 and higher numerical aspect ratios are considered to be “widescreen.”

Figure 12.17 shows the physical difference between standard and widescreen aspect ratio display formats.

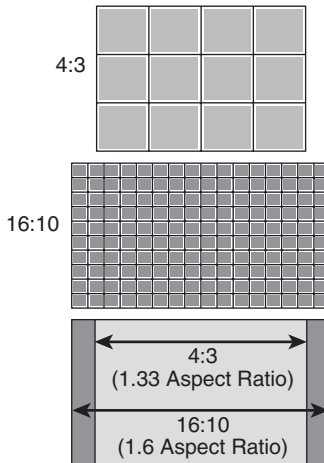


Figure 12.17 Standard (4:3 or 1.33) versus widescreen (16:10 or 1.60) display aspect ratios.

Pixels

In a color monitor, each picture element (pixel) consists of three red, green, and blue (RGB) subpixels. By varying the intensity of each of the subpixels, you can cause the overall color and brightness of the pixel to be anything from black (all off) to white (all on) and almost any color or level in between. The physical geometry of the RGB subpixels varies depending on the type of display, but the shape is normally either rectangular stripes or round dots. LCD monitors normally have the three subpixels arranged as rectangular vertical stripes in a linear repeating arrangement. CRTs may also use linear stripes, or they can have staggered stripes or dot triads.

When you're choosing a display, the most important considerations are the combination of size and resolution. The overall combination of size and resolution is normally expressed in pixels per inch (ppi), but it can also be expressed in pixel pitch, which is the distance between pixels in millimeters. A higher ppi number (or lower pixel pitch) means that fixed size images such as icons and text will be smaller and possibly harder to read. Pixel pitch is also sometimes called *dot pitch*, in reference to the dot-shaped subpixels used on some displays.

For a given size screen, higher resolution displays have a higher ppi number, which corresponds to a lower pixel pitch number. As a result, the picture elements are closer together, producing a sharper picture onscreen. Conversely, screens with a lower ppi number (which equals a larger pixel/dot pitch) tend to produce images that are more grainy and less clear.

Figure 12.18 illustrates a dot-shaped subpixel arrangement most commonly found on shadow-mask base CRTs, where the pixel or dot pitch is the shortest distance between same-color subpixels.

Figure 12.19 and 12.20 show striped subpixel arrangements in both linear and staggered forms. Of these, the linear form is the most common, used on virtually all LCDs and LEDs and most aperture grille-based CRTs. With striped subpixels, the pitch is measured as the horizontal or vertical distance between same color subpixels.

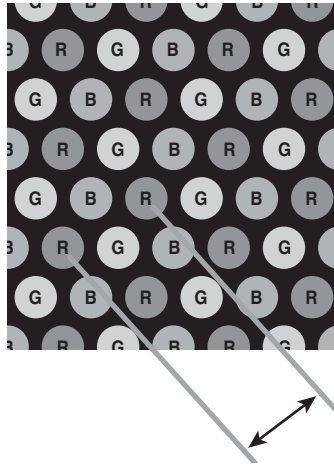


Figure 12.18 Dot-shaped subpixels, where pixel or dot pitch is the shortest distance between same color subpixels.

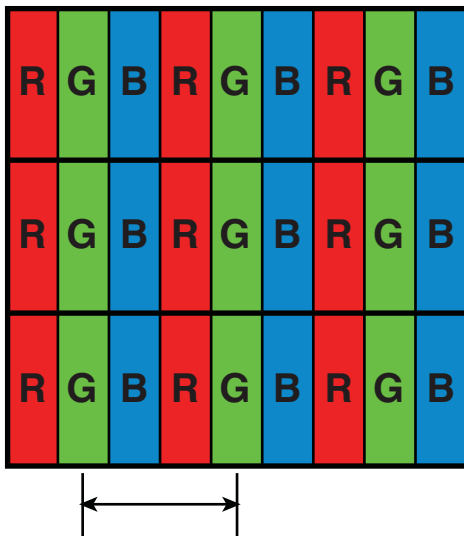


Figure 12.19 Stripe-shaped subpixels in a linear arrangement, where pixel pitch is the distance between same-color subpixel stripes.

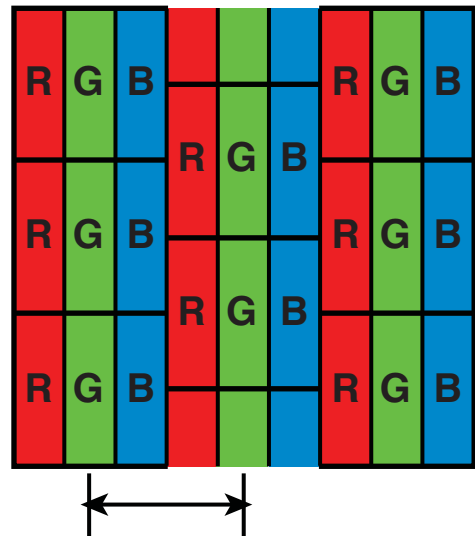


Figure 12.20 Stripe-shaped subpixels in a staggered arrangement, where pixel pitch is the distance between same-color subpixel stripes.

Generally, the higher the resolution, the larger the display you will want. Why? Because OS and application program icons and text normally use a constant number of pixels, higher display resolutions make these screen elements smaller onscreen. By using a larger display, you can use higher resolution settings and still have icons and text that are large enough to be readable. Although it is possible to change icon and text size, this often causes other problems with formatting in various windows and dialog boxes, such that in most cases it is best to stick with the default sizes.

At lower resolutions, text and onscreen icons are large. Because the screen elements used for the Windows desktop and software menus are at a fixed pixel width and height, you'll notice that they shrink in size onscreen as you change to the higher resolutions. You'll be able to see more of your document or web page onscreen at the higher resolutions because each object requires less of the screen. Tables 12.19 and 12.20 show the sizes and resolutions for commonly available standard and widescreen LCD monitors.

Table 12.19 Common Sizes and Resolutions for Non-Widescreen (<1.50 Ratio) Monitors

Display Size	Designation	Native Resolution	Megapixels	Pixels Aspect Ratio	Pixel per Inch	Pitch (mm)
15.0 in.	XGA	1024×768	0.79	1.33	85	0.298
17.0 in.	SXGA	1280×1024	1.31	1.25	96	0.263
19.0 in.	SXGA	1280×1024	1.31	1.25	86	0.294
20.0 in.	UXGA	1600×1200	1.92	1.33	100	0.254

Table 12.20 Common Sizes and Resolutions for Widescreen (>1.50 Ratio) Monitors

Display Size	Designation	Native Resolution	Megapixels	Pixels Aspect Ratio	Pixel per Inch	Pitch (mm)
16.0 in.	WXGA	1366×768	1.05	1.78	98	0.259
17.0 in.	WXGA+	1440×900	1.30	1.60	100	0.254
19.0 in.	WXGA+	1440×900	1.30	1.60	89	0.284
19.0 in.	WSXGA+	1680×1050	1.76	1.60	104	0.244
20.0 in.	WXGA+	1440×900	1.30	1.60	85	0.299
20.0 in.	WSXGA-	1600×900	1.44	1.78	92	0.277
20.0 in.	WSXGA+	1680×1050	1.76	1.60	99	0.256
22.0 in.	WXGA+	1440×900	1.30	1.60	77	0.329
22.0 in.	WSXGA+	1680×1050	1.76	1.60	90	0.282
24.0 in.	HD 1080	1920×1080	2.07	1.78	92	0.277
24.0 in.	WUXGA	1920×1200	2.30	1.60	94	0.269
26.0 in.	HD 1080	1920×1080	2.07	1.78	85	0.300
26.0 in.	WUXGA	1920×1200	2.30	1.60	87	0.292
30.0 in.	WQXGA	2560×1600	4.10	1.60	101	0.252

Because LCD monitors have a 1:1 relationship between resolution and pixels, any two displays having the same size and resolution always have the same pixel per inch or pixel pitch specification. Regardless of the actual resolution, any two displays having the same number of pixels per inch will display text, icons, and other elements on the screen at the same sizing. Although having a higher resolution display is generally considered better, you need to be careful when selecting small LCD displays with higher resolutions because icons and text will be much smaller than you might be accustomed to. You can change the icon and font sizes in Windows to compensate, but this often causes

problems such as abnormal word wrapping in dialog boxes. Also, many fonts are fixed, and they will remain at the smaller size no matter what settings you change.

Depending on their ability to see and read small text, many people will have difficulty seeing the text and icons when the display is rated at 100 ppi or higher. If you are going to choose a display rated over 100 ppi, you may need to either sit closer to the screen or use bifocals or reading glasses to read it. Changing resolution to a lower setting is usually unsatisfactory with a flat-panel display because either the display will simply be smaller (text and icons will remain the same size, but you can't fit as much on the screen) or the display will attempt to scale the data to fit the screen. However, scaling invariably results in a blurred and distorted image. The bottom line is that LCDs really work well only at their native resolution—something you should strongly consider when purchasing.

As a consolation, even with their tinier text and icons, LCD and LED backlit screens are much sharper and clearer than CRT-based monitors. So, even though the text and icons are physically smaller, they are often more readable with less eyestrain because they are sharp and perfectly focused.

CRTs don't have a 1:1 relationship between resolution and pixels. Therefore, when you're comparing CRTs, the smaller the pixel pitch, the sharper the images will be. As an example, the original IBM PC color monitor from the early 1980s had a pixel pitch of .43mm, whereas newer color CRTs have a pixel pitch between .25mm and .27mm, with high-end models offering .24mm or less. To avoid grainy images on a CRT, look for those with a pixel pitch of .26mm or smaller.

Horizontal and Vertical Frequency

Analog display connections such as VGA are designed to transmit signals that drive the display to draw images. These signals tell the display to draw an image by painting lines of pixels from left to right and from top to bottom. For example, if the display resolution is 1024×768, that means there are 768 lines that would be drawn, one after the other, from top to bottom. Once the 768th line is drawn, the entire image would be completed, and the process would repeat starting again from the top.

The speed at which this image drawing occurs has two components, called the *horizontal frequency* and the *vertical frequency*. These frequencies are also called *scan* or *refresh rates*. The horizontal frequency is the speed in which the horizontal lines are drawn, expressed as the total number of lines per second. The vertical frequency (or vertical refresh rate) is the speed in which complete images are drawn, expressed in the number of images per second.

Using a 1024×768 display as an example, if the vertical refresh rate is 60Hz, then all of the 768 lines that make up the image would need to be drawn 60 times per second, resulting in a horizontal frequency of 768 lines per image × 60 images per second, which equals 46,060 total lines per second, or a frequency of about 46KHz. If the vertical refresh rate were increased to 85Hz, the horizontal frequency would be $768 \times 85 = 65,280$, or about 65.3KHz. The actual figures are technically a bit higher at 47.8KHz and 68.7KHz, respectively, because there is about a 5% overhead called the *vertical blanking interval* that was originally designed to allow the electron beam to move from the bottom back to the top of the screen without being seen. Although there is no electron beam in LCD displays, the blanking time is still used for backward compatibility as well as to send additional data that is not part of the image. The exact amount of vertical and horizontal blanking time required varies by the resolution and mode. It is governed by the VESA CVT (Coordinated Video Timings) standard.

Displays with analog connections usually have a range of scan frequencies that they can handle, which effectively controls the minimum and maximum resolutions that can be displayed. Of the vertical and horizontal frequencies, for CRT displays the vertical refresh rate is more important because it controls flicker. A refresh rate that is too low causes CRT screens to flicker, contributing to eyestrain. The higher the refresh rate you use with a CRT display, the less eyestrain and discomfort you will experience while staring at the screen.

A *flicker-free refresh rate* is a refresh rate high enough to prevent you from seeing flicker on a CRT. The flicker-free refresh rate varies with the size and resolution of the monitor setting (larger and higher resolutions require higher refresh rates) as well as the individual because some people are more sensitive to it than others. In my experience, a 75Hz refresh rate is the minimum anybody should use with a CRT, especially at resolutions of 1024×768 and higher. Lower rates produce a noticeable flicker, which can cause eyestrain, fatigue, and headaches. However, although a 75Hz vertical refresh rate is sufficient for most, some people require a setting as high as 85Hz before the image is truly flicker-free. For that reason, 85Hz is considered by VESA to be the optimum refresh rate for CRT displays. Because a refresh rate that is too high can reduce video performance by making the adapter work harder to update the image more frequently, I recommend using the lowest refresh rate you are comfortable with.

Note

CRT manufacturers often used the term *optimal resolution* to refer to the highest resolution a given CRT monitor supports at the 85Hz VESA standard for flicker-free viewing.

Table 12.18 shows the correlation between resolution and refresh rates. As the resolution and vertical refresh rate increase, so must the horizontal frequency. The maximum horizontal frequency supported by the display is the limiting factor when selecting higher refresh rates at a given resolution.

Table 12.18 Refresh Rates Comparison

Resolution	Vertical Refresh Rate	Horizontal Refresh Rate*
1024×768	60Hz	47.8KHz
1024×768	75Hz	60.3KHz
1024×768	85Hz	68.7KHz
1280×1024	60Hz	63.7KHz
1280×1024	75Hz	80.3KHz
1280×1024	85Hz	91.5KHz
1600×1200	60Hz	74.5KHz
1600×1200	75Hz	94.1KHz
1600×1200	85Hz	107.2KHz

*Includes VESA CVT (Coordinated Video Timings) prescribed blanking overhead.

For example, let's say you have a CRT display that supports a maximum *horizontal frequency* of 75KHz. That display would be capable of an 85Hz refresh rate at 1024×768, but it would only be capable of a 60Hz refresh at either 1280×1024 or 1600×1200. Because the flickering on a CRT at 60Hz is unacceptable, using those resolutions would be out of the question. By comparison, a monitor capable of supporting a 110KHz horizontal frequency could handle even the highest 1600×1200 resolution at an 85Hz refresh rate, which would mean no flicker. Premium CRT displays offer flicker-free refresh rates at higher resolutions. Note that you can't use refresh rates higher than the display can support, and in some cases with older CRTs, selecting a refresh rate in excess of the monitor's maximum can actually cause damage to the display circuitry!

Windows supports Plug and Play (PnP) monitor configuration if both the monitor and video adapter support the DDC feature. Using DDC communication, Windows can read the VESA EDID from the display and use it to configure the graphics controller to match the display's capabilities, such as supported resolutions and refresh rates. This normally prevents the user from selecting refresh rates that the monitor cannot support.

LCD and backlit LED monitors aren't affected by vertical refresh rates like CRTs; LCDs avoid problems with flicker because of their design. LCDs use transistors to activate all the pixels in the image at once, as opposed to a scanning electron beam that must work its way from the top to the bottom of the screen to create an image. But most importantly, LCDs have a CCFL (cold cathode fluorescent lamp) or LED backlight that for all intents and purposes doesn't flicker. (It operates either continuously or at high frequencies of 200Hz or more.) In other words, although a vertical refresh rate setting of 60Hz is considered bad for CRTs, that is the standard rate used by most LCDs because they don't exhibit visible flicker. Although most LCDs can accept refresh rates of up to 75Hz, in most cases selecting rates higher than 60Hz will only force the video card to work harder and won't actually affect what you see on the display.

Tip

If you try to use a refresh rate higher than your display can support, you might see a message that you have selected an out-of-range frequency. If you use a dual-head video card, keep in mind that some models don't permit you to assign different refresh rates to each monitor. If you have both a CRT and an LCD connected to such a video card, use the highest refresh rate supported by both displays (usually 75Hz) to minimize flicker on the CRT.

Interlaced Versus Noninterlaced

Some monitors and video adapters can support both interlaced as well as noninterlaced modes. In *noninterlaced* (conventional) mode, the screen is drawn from top to bottom, one line after the other, completing the screen in one pass. In *interlaced* mode, the screen is drawn in two passes—with the odd lines first and the even lines second. Each pass takes half the time of a full pass in noninterlaced mode.

Early high-resolution CRT monitors used interlacing to reach their maximum resolutions; as with interlacing, the vertical and horizontal scan frequencies were cut in half. Unfortunately, this usually introduces noticeable flicker into the display, so in most cases you should avoid interlacing where possible. Fortunately, most modern monitors support noninterlaced modes at all supported resolutions, thus avoiding the slow screen response and potential flicker caused by interlacing.

Note

The 1080i HDTV standard is an interlaced mode that is sometimes used because it requires half the bandwidth of the 1080p (progressive) mode. However, in most cases, you will not see a DLP, LCD, or plasma TV flicker when it receives a 1080i signal. That is because the signal is normally converted internally into a progressive signal and scaled to the display's native resolution.

Image Brightness and Contrast

Although it's a consideration that applies to both LCDs and CRTs, the brightness of a display is especially important in an LCD panel because brightness can vary a great deal from one model to another. Brightness for LCD panels is measured in candelas per square meter (cd/m²), which is also called a *nit* (from the Latin *nitere*, "to shine") and often abbreviated as *nt*. Typical ratings for good display panels are between 200 and 450 nits—the brighter the better.

Contrast is normally expressed as the ratio between white and black, with higher ratios being better. There are unfortunately different ways to make the measurement, but the one that is most important is the *static contrast ratio*, which is the ratio from brightest to darkest that can be produced on a display simultaneously. Many display manufacturers like to quote dynamic contrast ratios instead because they are measured over time with different backlight brightness settings and produce significantly larger numbers. For example, a display with a 1000:1 static contrast ratio can also have an

8000:1 (or higher) dynamic contrast ratio. Even more confusion comes from the fact that many display manufacturers like to assign their own proprietary names to dynamic contrast ratios—for example, Acer calls it ACM (Adaptive Contrast Management), whereas ASUS calls it ASCR (ASUS Smart Contrast Ratio). I recommend comparing displays using only the static ratio.

Typical static contrast ratio values range from 400:1 to 1500:1. Anything higher than that is generally a dynamic ratio. Because of the capabilities of the human eye, static ratios over 1000:1 offer little perceptible visual difference. A good combination of both brightness and contrast is a brightness rating of 300 nits (or more) along with a static contrast ratio of 1000:1.

Note

When you evaluate an LCD TV monitor, be sure to note the brightness settings available in computer mode and TV mode. Many of these displays provide a brighter picture in TV mode than in computer mode.

Display Power Management Signaling

Monitors, like virtually all power-consuming computer devices, have been designed to save energy wherever and whenever possible. Virtually all monitors sold in recent years have earned the Environmental Protection Agency's Energy Star logo by reducing their current draw down to 15 watts (CRTs) or 5 watts (LCDs) or less when idle. Power-management features in the monitor, as well as controls provided in the system BIOS and in the latest versions of Windows, help monitors and other types of computing devices use less power.

►► For more information about power management, **see** Chapter 18, "Power Supplies," **p. 833**.

Display Power-Management Signaling (DPMS) is a VESA specification that defines the signals a computer sends to a monitor to indicate idle times. The OS normally decides when to send these signals, depending on how you have set up the power management settings.

Table 12.19 shows the various signal states and relative power consumption according to the DPMS state selected. Normally the monitor is placed in Suspend mode after a period of inactivity specified in the OS power management settings.

Table 12.19 Display Power Management Signaling States

State	Horizontal	Vertical	Video	Power Savings	Recovery Time
On	Pulses	Pulses	Active	None	N/A
Stand-By	No pulses	Pulses	Blanked	Minimal	Short
Suspend	Pulses	No pulses	Blanked	Substantial	Longer
Off	No pulses	No pulses	Blanked	Maximum	System dependent

Virtually all CRT monitors with power management features meet the requirements of the United States EPA's Energy Star labeling program, which requires that monitor power usage be reduced from up to 100 watts or more (when operating normally) to 15 watts or less in standby mode. LCD monitors comply with the far more stringent Energy 2000 (E2000) standard developed in Switzerland. E2000 requires that monitors use less than 5 watts when in standby mode. Note that LCD displays generally use one-third less power than CRTs in either operating or standby mode.

LCD Technology

Because of their light weight, smaller overall size, and much greater clarity, LCD panels have replaced CRT displays in virtually all new computer installations. Desktop LCD panels use the same technology that first appeared in laptop computers. Compared to CRTs, LCDs have completely flat, thin screens and low power requirements (30 watts versus up to 100 watts or more for CRT monitors). The color quality of a good active-matrix LCD panel can exceed that of many CRT displays, particularly when viewed from head on.

How LCD Displays Work

In an LCD, polarizing filters allow only light waves that are aligned with the filter to pass through. After passing through one polarizing filter, all the light waves are aligned in the same direction.

A second polarizing filter aligned at a right angle to the first blocks all those waves. By changing the angle of the second polarizing filter, you can change the amount of light allowed to pass accordingly. It is the role of the liquid crystal cell to act as a polarizing filter that can change the angle of polarization and control the amount of light that passes. The liquid crystals are tiny rod-shaped molecules that flow like a liquid. They enable light to pass straight through, but an electrical charge alters their orientation, which subsequently alters the orientation of light passing through them. In a color LCD, there are three cells for each pixel—one each for displaying red, green, and blue—with a corresponding transistor for each cell. The red, green, and blue cells that make up a pixel are sometimes referred to as *subpixels*. See Figure 12.21.

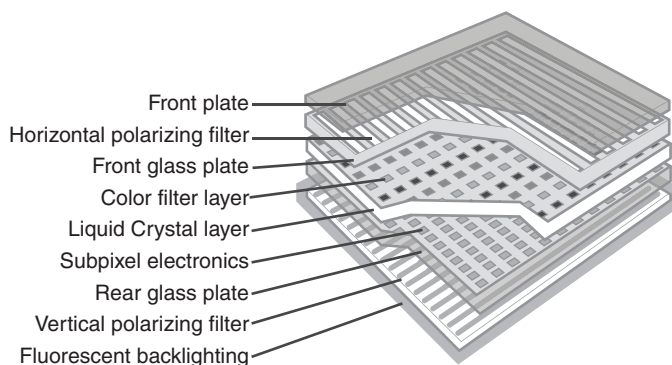


Figure 12.21 A diagram illustrating a typical color LCD display.

Active-Matrix Displays

LCD panels use a type of active-matrix technology known as a *thin-film transistor (TFT)* array. TFT is a method for packaging from one (monochrome) to three (RGB color) transistors per pixel within a flexible material that is the same size and shape as the display. Thus, the transistors for each pixel lie directly behind the liquid crystal cells they control.

Two TFT manufacturing processes account for most of the active-matrix displays on the market today: hydrogenated amorphous silicon (a-Si) and low-temperature polysilicon (p-Si). These processes differ primarily in their costs. At first, most TFT displays were manufactured using the a-Si process because it required lower temperatures (less than 400°C) than the p-Si process of the time. Now, lower-temperature p-Si manufacturing processes are making this method an economically viable alternative to a-Si.

To improve horizontal viewing angles in the latest LCDs, some vendors have modified the classic TFT design. For example, Hitachi's in-plane switching (IPS) design—also known as STFT—aligns the individual cells of the LCD parallel to the glass, running the electric current through the sides of the cells and spinning the pixels to provide more even distribution of the image to the entire panel area.

Hitachi's Super-IPS technology also rearranges the liquid crystal molecules into a zigzag pattern, rather than the typical row-column arrangement, to reduce color shift and improve color uniformity. The similar multidomain vertical alignment (MVA) technology developed by Fujitsu divides the screen into different regions and changes the angle of the regions.

Both Super-IPS and MVA provide a wider viewing angle than traditional TFT displays. Other companies have different names for the same technology—for example, Sharp calls it Ultra High Aperture (UHA). Manufacturers often like to think up their own buzzwords for the same technology because it makes their products seem different, but the results they generate are largely the same. Because larger LCDs can exhibit shifts in viewing angle even for an individual user, these more advanced technologies are often used on larger and more expensive panels.

Benefits of LCD Panels

LCD monitors offer a number of benefits when compared to conventional CRT glass tube monitors. Because LCDs use direct addressing of the display (each pixel in the picture corresponds with a transistor), they produce a high-precision image. LCDs can't have the common CRT display problems of pincushion or barrel distortion, nor do they experience convergence errors (halos around the edges of onscreen objects).

LCD panels are less expensive to operate than CRTs because they feature lower power consumption and much less heat buildup than CRTs. Because LCD units lack a CRT, no concerns exist about electromagnetic VLF or ELF emissions. Although LCDs offer a comparable mean time between failures (MTBF) to CRT displays, the major reason for LCD failures is the inverter or backlight, which might be relatively inexpensive to replace in some models. CRT failures usually involve the picture tube, which is the most expensive portion of the display and is often not cost-effective to replace.

LCD panels offer a significantly smaller footprint (front-to-back dimensions), and some offer optional wall or stand mounting. LCD panels also weigh substantially less than comparably sized CRTs. For example, a 17-inch LCD weighs less than 10 lbs., compared to the 50 lbs. weight of 19-inch CRTs with a similar viewing area.

LCD Monitor Selection Criteria

When selecting an LCD monitor, I recommend taking the following criteria into consideration:

- LCD monitors work best at their native resolution, and can vary greatly in how they scale to lower resolutions. Because higher resolutions result in smaller text and icons, make sure you don't purchase a display with a higher resolution than you can easily see or use. LCD monitors vary greatly in how well they scale from native to alternative resolutions, so you should evaluate the panel both at its native resolution and at any other resolutions you plan to use.
- For larger displays supporting higher resolutions, using a VGA analog connection will result in a poor quality image. In that case, you will want to use a digital connection, which means that both the video adapter and display will need compatible digital connections such as DisplayPort, DVI, or HDMI.
- LCDs have slower response times than most CRTs. For good performance with games, video, and animation, look for gray-to-gray (G2G) response times of 5ms or faster.
- LCD displays have lower viewing angles than CRTs. This can be an important consideration if you're planning to use your LCD monitor for group presentations. To improve the horizontal viewing area, several manufacturers have developed patented improvements to the basic TFT display, such as Hitachi's IPS, Fujitsu's multidomain vertical adjustment (MVA), and Mitsubishi's feed forward driving (FFD)—most of which have been licensed to other leading LCD makers. If you need wider angle viewing capability, look for LCD displays using these technologies to achieve horizontal viewing angles of 170° or more.

- A high-contrast ratio (luminance difference between white and black) makes for sharper text and vivid colors. Look for static contrast ratios of 1000:1 or more.
- Optional features such as integrated speakers, webcams, and USB hubs are available on many displays.

LED Backlit Technology

An increasingly popular variation on standard LCD display technology replaces fluorescent tubes with LEDs for backlighting. By using LEDs instead of fluorescent tubes, display warmup time is nil and mercury, a major environmental concern when disposed of, is no longer a concern when the display is scrapped. LED backlit displays also use about half the energy of standard LCD displays.

Figure 12.22 illustrates how an LED backlit display works.

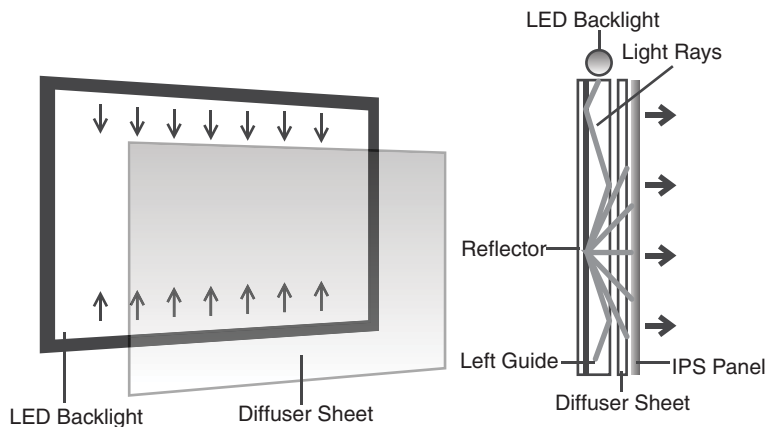


Figure 12.22 A diagram illustrating how an LED backlit display works.

CRT Display Technology

CRT technology is the same used in older television sets. In the past few years, CRTs have vanished store shelves, mainly due to the availability of lower-cost LCDs. However, many older systems still use CRT monitors. CRTs consist of a vacuum tube enclosed in glass. One end of the tube contains an electron gun assembly that projects three electron beams, one each for the red, green, and blue phosphors used to create the colors you see onscreen; the other end contains a screen with a phosphorous coating.

When heated, the electron gun emits a stream of high-speed electrons that are attracted to the other end of the tube. Along the way, a focus control and deflection coil steer the beam to a specific point on the phosphorous screen. When struck by the beam, the phosphor glows. This light is what you see when you watch TV or look at your computer screen. Three layers of phosphors are used: red, green, and blue. A metal plate called a *shadow mask* aligns the electron beams; it has slots or holes that divide the red, green, and blue phosphors into groups of three (one of each color).

Monitors based on Sony Trinitron or Mitsubishi DiamondTron picture tubes used an aperture grille type mask to separate red, green, and blue phosphors, resulting in strips of square pixels in a linear arrangement, similar to an LCD. NEC's ChromaClear monitors used a variation on the aperture grille called the *slotted mask*, which is brighter than standard shadow-mask monitors and more mechanically stable than aperture grille-based monitors. This resulted in a staggered pixel arrangement.

Various types of shadow masks affect picture quality, and the distance between each group of three (the *dot or pixel pitch*) affects picture sharpness.

◀◀ See “Pixels,” p. 643 (this chapter).

Figure 12.23 illustrates the interior of a typical CRT.

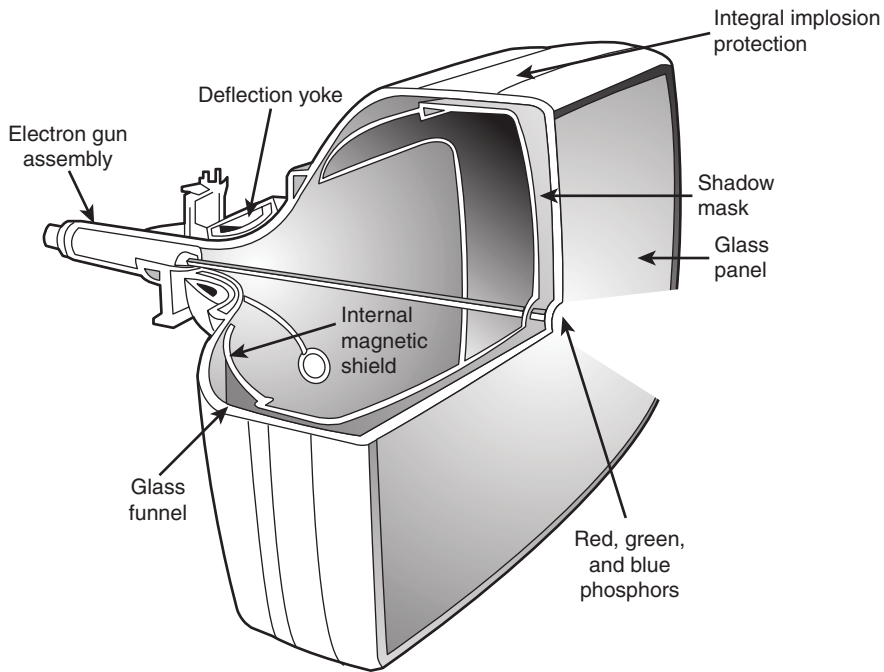


Figure 12.23 A typical CRT monitor is a large vacuum tube. It contains three electron guns (red, green, and blue) that project the picture toward the front glass of the monitor. High voltage produces the magnetism that controls the electron beams that create the picture displayed on the front of the CRT.

The phosphor chemical has a quality called *persistence*, which indicates how long this glow remains onscreen. Persistence is what causes a faint image to remain on your TV screen for a few seconds after you turn off the set. The vertical scanning frequency (also called the *refresh rate*) of the display specifies how many times per second the image is refreshed. You should have a good match between persistence and refresh rate so the image has less flicker (which occurs when the persistence is too low) and no ghost images (which occurs when the persistence is too high).

The electron beam moves quickly, sweeping the screen from left to right in lines from top to bottom, in a pattern called a *raster*. The horizontal scan rate refers to the speed at which the electron beam moves laterally across the screen, measured in the number of lines drawn per second.

During its sweep, the beam strikes the phosphor wherever an image should appear onscreen. The beam also varies in intensity to produce different levels of brightness. Because the glow begins to fade almost immediately, the electron beam must continue to sweep the screen to maintain an image—a practice called *redrawing* or *refreshing* the screen.

Due to the lower persistence phosphors used in PC monitor CRTs, most have an ideal refresh rate (also called the *vertical scan frequency*) of 85Hz, which means the screen is refreshed 85 times per second. Refresh rates that are too low cause the screen to flicker, contributing to eyestrain.

Plasma Display Technology

Plasma displays have a long history in PCs. In the late 1980s, IBM developed a monochrome plasma screen that displayed orange text and graphics on a black background. IBM used a 10-inch version of this gas plasma display on its P70 and P75 briefcase portable systems that were released way back in 1988.

Unlike the early IBM monochrome plasma screens, today's plasma displays are capable of displaying 24-bit or 32-bit color. Plasma screens produce an image by using electrically charged gas (plasma) to illuminate triads of red, green, and blue phosphors, as shown in Figure 12.24.

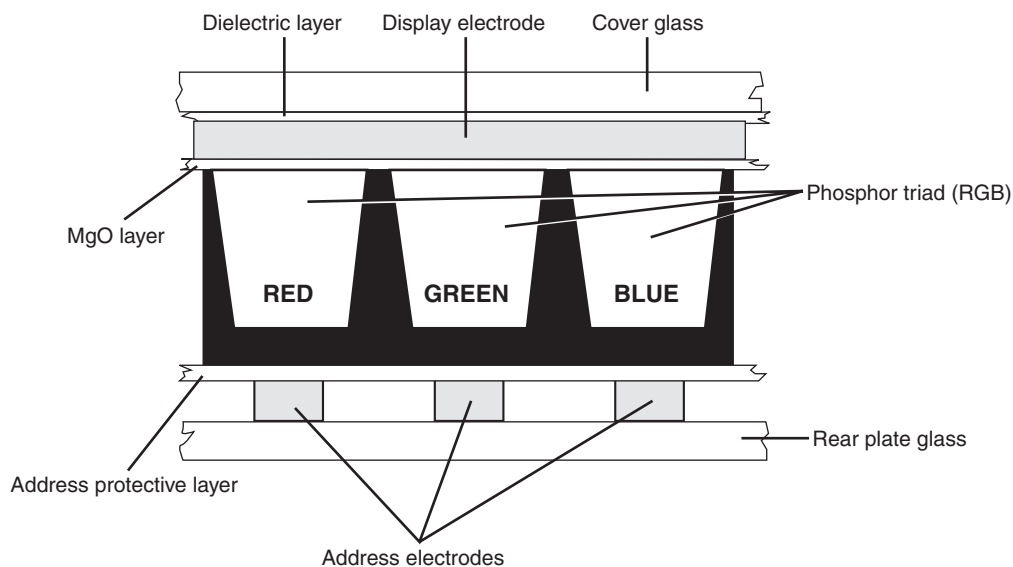


Figure 12.24 A cross-section of a typical plasma display.

The display and address electrodes create a grid that enables each subpixel to be individually addressed. By adjusting the differences in charge between the display and address electrodes for each triad's subpixels, the signal source controls the picture.

Typical plasma screens range in size from 42 inches to 50 inches or larger. Because they are primarily designed for use with DVD, TV, or HDTV video sources, they are sized and optimized for video rather than computer use. Their quick response time makes them good choices for enjoying sports, streaming HD video, or Blu-ray movies.

LCD and DLP Projectors

Originally, data projectors were intended for use in boardrooms and training facilities. However, with the rise of home theater systems, the increasing popularity of working from home, and major price reductions and improvements in projector technology, portable projectors are an increasingly popular alternative to large-screen TVs and plasma displays.

Two technologies are used in the construction of data projectors:

- Liquid crystal display (LCD)
- DLP

Instead of using triads of subpixels as in a flat-panel or portable LCD, an LCD projector works by separating white light into red, green, and blue wavelengths and directing each wavelength through a corresponding LCD panel. Each LCD panel's pixels are opened or closed according to the signals received from the signal source (computer, DVD, or video player) and are combined into a single RGB image that is projected onto the screen. A relatively hot projection lamp must be used to project LCD images, so LCD projectors require some cool-down time before they can be stored.

The other major technology for presentation and home theater projectors uses Texas Instruments' own DLP technology. DLP projectors use a combination of a rapidly spinning color wheel and a microprocessor-controlled array of tiny mirrors known as a *digital micromirror device* (DMD). Each mirror in a DMD corresponds to a pixel, and the mirrors reflect light toward or away from the projector optics. Depending on how frequently the mirrors are switched on, the image varies from white (always on) to black (never on) through as many as 1,024 gray shades. The color wheel provides color data to complete the projected image. Compared to LCD projectors, DLP projectors are more compact, are lighter, and cool down more quickly after use. Although DLP projects were originally more expensive than LCD projectors, they are now about the same price. Most current projectors also support HDTV resolutions of 720p and 1080i, enabling a single projector to work as either a PC or a TV display.

Figure 12.25 illustrates how a DLP-based projector produces the image.

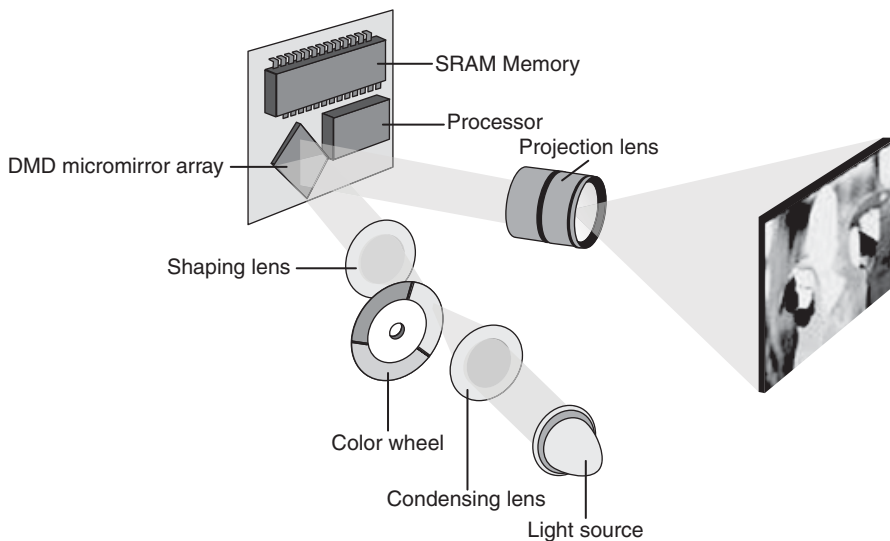


Figure 12.25 How a typical DLP projector works.

The earliest DLP projectors used a simple three-color (RGB) wheel, as shown in Figure 12.25. However, more recent models have used a four-segment (RGB and clear) or a six-segment (RGBRGB) wheel to improve picture quality.

Note

For more information about digital light processing, see the official Texas Instruments website about DLP technology at www.dlp.com.

Using Multiple Monitors

One of the most useful things you can do for increasing the usability and productivity of a PC is to attach multiple monitors. Adding additional monitors gives you more screen real estate to use for running multiple applications simultaneously. When you configure a system to use multiple monitors, the OS creates a virtual desktop—that is, a display that exists in video memory that combines the total screen real-estate of all of the attached displays. You use the multiple monitors to display various portions of the virtual desktop, enabling you to place the open windows for different applications on separate monitors and move them around at will.

Although having three or more displays is entirely possible, just adding a second monitor can be a boon to computing productivity. For example, you can leave an email client or word processor maximized on one monitor and use the other monitor for web browsing, research documents, and more.

Using multiple monitors requires an independent video signal for each display. Starting with Windows 2000, this is most commonly done by using Dualview—a single graphics adapter with two video outputs, also known as a *dual-head graphics adapter*. It is also possible to use two or more video cards, but this approach is used today only if three or more displays are used.

When multiple video adapters are installed, the system identifies one of the video adapters as primary. The primary adapter is sometimes called the *VGA adapter* and is the one that displays the POST (Power-On Self Test) and BIOS Setup screens. This is a function of the motherboard and motherboard BIOS. Most modern BIOSs allow you to choose the primary display adapter via a setting in the BIOS Setup. Normally the options are onboard (built in), PCI, or PCIe (AGP on older systems with an AGP slot). For the PCI and PCIe selections, if you have multiple adapters installed, the primary will be the one in the highest priority slot.

If the BIOS does not let you select which device should be the primary display, it decides solely based on the priority of the buses or slots in the machine. Depending on the BIOS used by your system, you might need to check in various places for the option to select the primary display adapter; however, in most cases it will be in the Video Configuration menu.

Dualview

Although older video cards often have only a single output, most current video cards have at least two. With the exception of laptops, the same is true for most motherboard-based video adapters. Laptops have almost always had dualview graphics adapters, whereas most desktop motherboards have not until recently.

A dualview card is preferable to using two separate cards because only one slot is used, as are fewer system resources and even power. The types of video outputs on dual-head cards can vary, so make sure you choose a card that has the outputs you need or that you can change the existing outputs via adapters. Digital outputs such as DVI, HDMI, and DisplayPort are preferred because these can normally be converted to others with inexpensive adapters.

Homogeneous Adapters

The best solution for running multiple graphics adapters in a single system is to ensure they are homogeneous. This means that they use the same driver, which implies that they must also have chipsets from the same manufacturer (such as AMD, ATI or NVIDIA) and from compatible families within that chipset manufacturer. Using multiple homogeneous adapters is supported in Windows 98/Me (up to nine total displays) and in Windows 2000 and later (up to 10 displays) using any combination of single- or dualview adapters. The installation is simple because only a single driver is required for all the adapters and displays.

Because ATI and NVIDIA use a unified driver architecture to support their current product lines, you can use two (or more) ATI-based or NVIDIA-based cards to create the desired homogeneous multiple-monitor configuration. The specific cards and even chipsets can be different, as long as they can use the same driver.

Heterogeneous Adapters

Heterogeneous adapters are those using chipsets from different manufacturers or from incompatible chipset families within the same manufacturer, which therefore require different drivers. This method has largely been superseded by Dualview or by using matching display adapters.

Note

For more information about using heterogeneous adapters, see "Heterogeneous Adapters," in Chapter 12, "Video Hardware," in *Upgrading and Repairing PCs*, 19th edition, included on the disc packaged with this book.

Video Troubleshooting and Maintenance

Solving most graphics adapter and monitor problems is fairly simple, but costly, because replacing the adapter or display is the normal procedure. Except for specialized CAD or graphics workstation-oriented adapters, virtually all of today's adapters cost more to service than to replace, and the documentation required to service the hardware properly is not always available. You can't get schematic diagrams, parts lists, wiring diagrams, and other documents for most adapters or monitors. However, before you replace an apparently failed component, be sure you have exhausted all your other options. These include following the troubleshooting tips in the next section and checking the warranty status of your device.

Tip

If you're not in the habit of registering newly purchased products, it's a good habit to pick up. Registering your hardware helps to verify warranty status, sometimes provides longer warranty coverage for your equipment, and helps the vendor reach you with updates or service advisories.

Remember also that many of the problems you might encounter with modern video adapters and displays are related to the drivers that control these devices rather than to the hardware. Be sure you have the latest and proper drivers before you attempt to have the hardware repaired; a solution might already be available.

Troubleshooting Video Cards and Drivers

Use the following suggestions to help you solve problems with your display hardware.

The display works when booting and in the BIOS Setup, but not in Windows.

Solution

If you have an acceptable picture while booting or when in the BIOS Setup but no picture in Windows, most likely you have an incorrect or corrupted video driver installed. Boot Windows in either Safe Mode (which uses the Microsoft supplied vga.sys driver) or Enable VGA Mode (which uses the current driver with VGA resolution settings). If Safe Mode or VGA Mode works, get a correct or updated driver for the video card and reinstall.

If you have overclocked your card with a manufacturer-supplied or third-party utility, you might have set the speed too high. Restart the system in Safe Mode and reset the card to run at its default speed. If you have adjusted the speed of AGP/PCI/PCIe slots in the BIOS Setup program, restart the system, start the BIOS Setup program, and reset these slots to run at the normal speed.

Can't select desired color depth and resolution combination.

Solution

Verify that the card is properly identified in Windows and that the card's memory is working properly. Use diagnostic software provided by the video card or chipset maker to test the card's memory. If the hardware is working properly, check for new drivers. Use the vendor's drivers rather than the ones provided with Windows.

Can't select desired refresh rate.

Solution

Verify that the video adapter and the monitor have the latest drivers installed and are properly identified in Windows. If necessary, obtain and install updated drivers for the adapter and display.

Can't adjust OpenGL or Direct3D (DirectX) settings.

Solution

Install the graphic card or chipset vendor's latest drivers instead of using the drivers included with Microsoft Windows. Standard Microsoft drivers often don't include 3D or other advanced dialog boxes.

Can't enable multiple monitors.

Solution

If you are using a multihead graphics card, make sure the additional monitors have been enabled in the manufacturer's driver. This might require you to open the Advanced settings for your driver. (Make sure you are using the latest one.) If you are using two or more video cards in an SLI (NVIDIA), CrossFire (ATI) or CrossFireX (AMD) configuration, you must normally disable this configuration before you can enable additional monitors. If you are using multiple video cards in separate slots, check the BIOS Setup configuration for the primary adapter.

Can't enable SLI operation.

Solution

Make sure the MIO (SLI bridge) device is properly installed between your video cards (refer to Figure 12.16). If you are not using identical video cards, you must use NVIDIA ForceWare v81.85 or new drivers to enable SLI operation. Make sure both cards use the same GPU family. Make sure SLI operation is enabled in the ForceWare driver.

Can't enable CrossFire or CrossFireX operation.

Solution

If you are using first-generation CrossFire cards, make sure you have paired up a CrossFire Edition and standard Radeon card from the same GPU family. Also, make sure you have properly connected the CrossFire Edition and standard card.

If you use later-model cards that use a CrossFire or CrossFireX bridge interconnect, make sure the interconnect is properly attached to both cards. Make sure you are using pairs of cards with the same GPU.

Update the ATI CATALYST display drivers to the latest production version. Be sure you have enabled CrossFire or CrossFireX in your display driver.

Can't enable Aero 3D desktop in Windows Vista or Windows 7.

Solution

Make sure your video card or integrated video is running a WDDM (Windows driver display model) driver and supports DirectX 9.0 or later. Update the driver if the hardware can support Aero. Update the card if the card does not have WDDM drivers available.

DisplayMate

DisplayMate is a unique diagnostic and testing program designed to thoroughly test your video adapter and display. It is somewhat unique in that most conventional PC hardware diagnostics programs do not emphasize video testing the way this program does.

I find it useful not only in testing whether a video adapter is functioning properly, but also in examining video displays. You easily can test the image quality of a display, which allows you to make focus, centering, brightness and contrast, color level, and other adjustments much more accurately than before. If you are purchasing a new monitor, you can use the program to evaluate the sharpness and linearity of the display and to provide a consistent way of checking each monitor you are considering. If you use projection systems for presentations—as I do in my PC hardware seminars—you will find it invaluable for setting up and adjusting the projector.

DisplayMate also can test a video adapter thoroughly. It sets the video circuits into each possible video mode so you can test all its capabilities. It even helps you determine the performance level of your card, with respect to both resolution and colors as well as speed. You can then use the program to benchmark the performance of the display, which enables you to compare one type of video adapter system to another. The Video Edition also supports various types of multimedia projectors and TVs used in home theater and presentation environments.

For more information on DisplayMate, visit www.displaymate.com.

Video Drivers

Drivers are an essential, and often problematic, element of a video display subsystem. The driver enables your software to communicate with the video adapter. You can have a video adapter with the fastest processor and the most efficient memory on the market but still have poor video performance because of a badly written driver.

Video drivers generally are designed to support the processor on the video adapter. All video adapters come equipped with drivers the card manufacturer supplies, but often you can use a driver the chipset maker created. Sometimes you might find that one of the two provides better performance than the other or resolves a particular problem you are experiencing.

Most manufacturers of video adapters and chipsets maintain websites from which you can obtain the latest drivers; drivers for chipset-integrated video are supplied by the system, motherboard, or chipset vendor. A driver from the chipset manufacturer can be a useful alternative, but you should always try the adapter manufacturer's driver first. Before purchasing a video adapter, you should check out the manufacturer's site and see whether you can determine how up to date the available drivers are. At one time, frequent driver revisions were thought to indicate problems with the hardware, but the greater complexity of today's systems means that driver revisions are a necessity. Even if you are installing a brand-new model of a video adapter, be sure to check for updated drivers on the manufacturer's website for best results.

Note

Although most devices work best with the newest drivers, video cards can be a notable exception. Both NVIDIA and AMD use unified driver designs, creating a single driver installation that can be used across a range of graphics chips. However, in some cases, older versions of drivers sometimes work better with older chipsets than the newest drivers do. If you find that system performance or stability, especially in 3D gaming, drops when you upgrade to the newest driver for your 3D graphics card, revert to the older driver.

The video driver also provides the interface you can use to configure the display your adapter produces. The driver controls the options that are available for these settings, so you can't choose parameters the hardware doesn't support. For example, you cannot select resolutions your display does not support, even though your video card might support them.

In most cases, another tab called Color Management is also available. You can select a color profile for your monitor to enable more accurate color matching for use with graphics programs and printers.

Video cards with advanced 3D acceleration features often have additional properties; these are discussed later in this chapter.

Maintaining Monitors

Because you can use a good monitor for several years on more than one computer, proper care is essential to extend its life to the fullest extent.

Use the following guidelines for proper care of your monitor:

- Although phosphor burn (where an image left onscreen eventually leaves a permanent shadow) is possible on CRT or plasma displays, LCDs are immune from this problem. To prevent burn-in on non-LCD displays, use the power management settings in your OS to blank the display after a period of inactivity. Screen savers used to be employed for display protection, but you should avoid them in lieu of putting the display into Standby mode.

- To prevent premature failure of the monitor's power supply, use the power-management features of the OS to put the monitor into a low-power standby mode after a reasonable period of inactivity. Using the power-management feature is far better than using the on/off switch when you are away from the computer for brief periods. Turn off the monitor only at the end of your computing "day."

How can you tell whether the monitor is really off or in standby mode? Look at the power LCD on the front of the monitor. A monitor that's in standby mode usually has a blinking green or solid amber LCD in place of the solid green LCD displayed when it's running in normal mode.

If the monitor will not go into standby mode when the PC isn't sending signals to it, make sure the monitor is properly defined in the Display properties sheet in Windows. In addition, make sure the Energy Star check box is selected for any monitor that supports power management, unless the monitor should be left on at all times (such as when used in a retail kiosk or self-standing display).

- Make sure the monitor has adequate ventilation along the sides, rear, and top. Because monitors use passive cooling, a lack of adequate airflow caused by piling keyboards, folders, books, or other office debris on top of the monitor will cause it to overheat and considerably shorten its life. If you need to use a monitor in an area with poor airflow, use an LCD panel instead of a CRT because LCDs run much cooler than CRTs.
- The monitor screen and case should be kept clean. Turn off the power, spray a cleaner such as Endust for Electronics onto a soft cloth (never directly onto the monitor!), and gently wipe the screen and the case.
- If your CRT monitor has a degaussing button or feature, use it periodically to remove stray magnetic signals. Keep in mind that CRTs have powerful magnets around the picture tube, so keep magnetic media away from them.

Testing Monitors

Unlike most of the other peripherals you can connect to your computer, you can't really tell whether a monitor suits you by examining its technical specifications. Price might not be a reliable indicator either. Testing monitors is a highly subjective process, and it is best to "kick the tires" of a few at a dealer showroom or in the privacy of your home or office (if the dealer has a liberal return policy).

Testing should also not be simply a matter of looking at whatever happens to be displayed on the monitor at the time. Many computer stores display movies, scenic photos, or other flashy graphics that are all but useless for a serious evaluation and comparison. If possible, you should look at the same images on each monitor you try and compare the manner in which they perform a specific series of tasks.

Before running the tests listed here, set your display to the highest resolution and refresh rate allowed by your combination of display and graphics card:

- Draw a perfect circle with a graphics program. If the displayed result is an oval, not a circle, this monitor will not serve you well with graphics or design software.
- Using a word processor, type some words in 8- or 10-point type (1 point equals 1/72 inches). If the words are fuzzy or the black characters are fringed with color, select another monitor.
- Display a screen with as much white space as possible and look for areas of color variance. This can indicate a problem with only that unit or its location, but if you see it on more than one monitor of the same make, it might indicate a manufacturing problem; it could also indicate problems with the signal coming from the graphics card. Move the monitor to another system equipped with a different graphics card model, and retry this test to see for certain whether it's the monitor or the video card.

- Display the Microsoft Windows desktop to check for uniform focus (with CRT displays) and brightness (with CRT and LCD displays). Are the corner icons as sharp as the rest of the screen? Are the lines in the title bar curved or wavy? Monitors usually are sharply focused at the center, but seriously blurred corners indicate a poor design. Bowed lines on a CRT can be the result of a poor video adapter or incorrect configuration of the monitor's digital picture controls. Before you decide to replace the monitor, you should adjust the digital picture controls to improve the display. Next, try attaching the monitor to another display adapter. If the display quality does not improve, replace the monitor.

Adjust the brightness up and down to see whether the image blooms or swells, which indicates the monitor is likely to lose focus at high brightness levels. You can also use diagnostics that come with the graphics card or third-party system diagnostics programs to perform these tests.

- With LCD panels in particular, change to a lower resolution from the panel's native resolution using the Microsoft Windows Display properties settings. Because LCD panels have only one native resolution, the display must use scaling to handle other resolutions full-screen. If you are a web designer, are a gamer, or must capture screens at a particular resolution, this test will show you whether the LCD panel produces acceptable display quality at resolutions other than normal. You can also use this test on a CRT, but CRTs, unlike LCD panels, are designed to handle a variety of resolutions.
- A good CRT monitor is calibrated so that rays of red, green, and blue light hit their targets (individual phosphor dots) precisely. If they don't, you have bad convergence. This is apparent when edges of lines appear to illuminate with a specific color. If you have good convergence, the colors are crisp, clear, and true, provided there isn't a predominant tint in the phosphor.
- If the monitor has built-in diagnostics (a recommended feature), try them as well to test the display independently of the graphics card and system to which it's attached. A display with built-in diagnostics shows text or a test pattern onscreen if it is receiving power when the host system is turned off or if the monitor is not connected to a host system.

Adjusting Monitors

One embarrassingly obvious fix to monitor display problems that is overlooked by many users is adjusting the monitor controls, such as contrast and brightness. Although most recent monitors use front-mounted controls with onscreen display (OSD), other adjustments might also be possible.

Older CRT monitors, for example, may have a focus adjustment screw on the rear or side of the unit. Because the adjusting screw is deep inside the case, the only evidence of its existence is a hole in the plastic grillwork. To adjust the monitor's focus, you must stick a long-shanked insulated screwdriver about 2 inches long into the hole and feel around for the screw head. This type of adjustment can save you an expensive repair bill. Always examine the monitor case, documentation, and manufacturer's website or other online services for the locations of adjustment controls.

Virtually all recent monitors use digital controls instead of analog controls. This has nothing to do with the signals the monitor receives from the computer, but only the controls (or lack of them) on the front panel that enable you to adjust the display. Monitors with digital controls have a built-in menu system that enables you to set parameters such as brightness (which adjusts the black level of the display), contrast (which adjusts the luminance of the display), screen size, vertical and horizontal shifts, color, phase, and focus. A button brings up the menu onscreen, and you use controls to make menu selections and vary the settings. When you complete your adjustments, the monitor saves the settings in nonvolatile RAM (NVRAM) located inside the monitor. This type of memory provides permanent storage for the settings with no battery or other power source. You can unplug the monitor without losing your settings, and you can alter them at any time in the future. Digital controls provide a much higher level of control over the monitor and are highly recommended.

Digital controls make adjusting CRT monitors suffering from any of the geometry errors shown in Figure 12.26 easy. Before making these adjustments, be sure the vertical and horizontal size and position are correct.

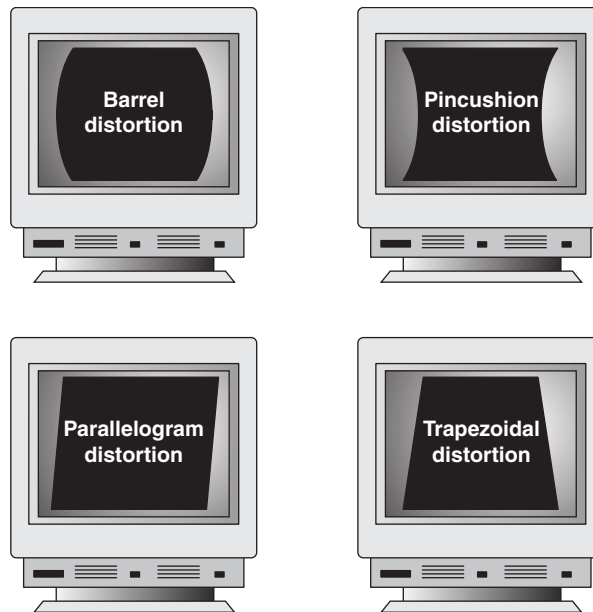


Figure 12.26 Typical geometry errors in CRT monitors; these can be corrected on most models that have digital picture controls.

Although LCD panels aren't affected by geometry errors as CRT monitors can be, they can have their own set of image-quality problems, especially if they use the 15-pin analog VGA video connector. Pixel jitter and pixel swim (in which adjacent pixels turn on and off) are relatively common problems that occur when you are using an LCD monitor connected to your PC with an analog VGA connector.

You can use the Auto-Tune option available in many LCD panels' OSDs to fix these and other LCD display problems.

Bad Pixels

A so-called *bad pixel* is one in which the red, green, or blue subpixel cell remains permanently on or off. Those that are permanently on are often called *stuck pixels*, whereas those that are permanently off are called *dead pixels*. Failures in the on state seem to be more common. In particular, pixels stuck on are noticeable on a dark background as tiny red, green, or blue dots. Although even one of these can be distracting, manufacturers vary in their warranty policies regarding how many bad pixels are required before you can get a replacement display. Some vendors look at both the total number of bad pixels and their locations. Fortunately, improvements in manufacturing quality make it less and less likely that you will see LCD screens with bad pixels.

Although there is no standard way to repair bad pixels, a couple of simple fixes might help. One involves tapping or rubbing on the screen. For example, I have actually repaired stuck pixels on several occasions by tapping with my index finger on the screen directly over the pixel location (with the screen powered on). Because I find a constantly lit pixel to be more irritating than one that is

constantly dark, this fix has saved me a lot of aggravation (when it has worked). A similar technique is to use the tip of a PDA stylus or ballpoint pen to apply pressure or to tap directly on the stuck pixel. I recommend you wrap a damp cloth over the tip to prevent scratching the screen. Some have had success by merely rubbing the area where the stuck or dead pixel is located.

Another potential fix involves using software to rapidly cycle the stuck pixel (as well as some adjacent ones), which sometimes causes the stuck pixel to become unstuck and function properly. The two main programs for doing this are Udpixel (<http://udpix.free.fr>) and Jscreenfix (www.jscreenfix.com).

Unfortunately, none of these fixes work all the time; in fact, in most cases the pixel remains stuck on or dead no matter what you try. If you have stuck or dead pixels that do not respond to any of the fixes I've detailed, you might want to contact the screen or laptop manufacturer to inquire about its bad pixel replacement policy. Because the policies can vary widely among different manufacturers and displays, to find the allowable defect limits for your specific display, I recommend you consult the documentation or contact the manufacturer directly.

Troubleshooting Monitors

Follow the suggestions in the following sections to help fix problems with your display.

No picture

Solution

If the LED on the front of the monitor is yellow or flashing green, the monitor is in power-saving mode. Move the mouse or press Alt+Tab on the keyboard and wait up to 1 minute to wake the system if it is turned on.

If the LED on the front of the monitor is green, the monitor is in normal mode (receiving a signal), but the brightness and contrast are set incorrectly; adjust them.

If no lights are lit on the monitor, check the power and power switch. Check the surge protector or power director to ensure that power is going to the monitor. Replace the power cord with a known-working spare if necessary. Retest. Replace the monitor with a known-working spare to ensure that the monitor is the problem.

Check data cables at the monitor and video card end. Check power cables to the video card if the video card requires additional power to operate.

Jittery picture quality

Solution

For LCD monitors, use display-adjustment software or onscreen menus to reduce or eliminate pixel jitter and pixel swim caused by running the digital LCD display with an analog (VGA) video source. Use a digital connection (DVI, HDMI, or DisplayPort) instead of the VGA connection to avoid digital-analog-digital conversion problems like these.

For all monitors, check cables for tightness at the video card and the monitor (if removable):

- Remove the extender cable and retest with the monitor plugged directly into the video card. If the extender cable is bad, replace it.
- Check the cables for damage; replace as needed.
- If problems are intermittent, check for interference. (Microwave ovens near monitors can cause severe picture distortion when turned on.)

For CRT monitors, check refresh-rate settings; reduce them until acceptable picture quality is achieved:

- Use onscreen picture adjustments until an acceptable picture quality is achieved.
- If problems are intermittent and can be “fixed” by waiting or gently tapping the side of the monitor, the monitor power supply is probably bad or has loose connections internally. Service or replace the monitor.

Repairing Monitors

Although a display often is replaced as a whole unit, some larger displays might be cheaper to repair than to replace. If you decide to repair the monitor, your best bet is to either contact the company from which you purchased the display or contact one of the companies that specialize in monitor depot repair.

Depot repair means you send in your display to a repair specialist who either fixes your particular unit or returns an identical unit it has already repaired. This usually is accomplished for a flat-rate fee; in other words, the price is the same no matter what the company has done to repair your actual unit.

Because you usually get a different (but identical) unit in return, the company can ship your repaired display immediately upon receiving the one you sent in, or even in advance in some cases. This way, you have the least amount of downtime and can receive the repaired display as quickly as possible. In some cases, if your particular monitor is unique or one the company doesn't have in stock, you must wait while your specific unit is repaired.

Troubleshooting a failed monitor is relatively simple. If your display goes out, for example, a swap with another monitor can confirm that the display is the problem. If the problem disappears when you change the display, the problem is almost certainly in the original display or the cable; if the problem remains, it is likely in the video adapter or PC.

Many of the better quality, late-model monitors have built-in self-diagnostic circuitry. Check your monitor's manual for details. Using this feature, if available, can help you determine whether the problem is really in the monitor, in a cable, or somewhere else in the system. If self-diagnostics produce an image onscreen, look to other parts of the video subsystem for your problem.

The monitor cable can sometimes be the source of display problems. A bent pin in the connector that plugs into the video adapter can prevent the monitor from displaying images, or it can cause color shifts. Most of the time, you can repair the connector by carefully straightening the bent pin with sharp-nosed pliers. A loose cable can also cause color shifts; make sure the cable is securely attached.

If the pin breaks off or the connector is otherwise damaged, you can sometimes replace the monitor cable. Some monitor manufacturers use cables that disconnect from the monitor and video adapter, whereas others are permanently connected. Depending on the type of connector the device uses at the monitor end, you might have to contact the manufacturer for a replacement.

If you narrow down the problem to the display, consult the documentation that came with the monitor or call the manufacturer for the location of the nearest factory repair depot. Third-party depot repair service companies are also available that can repair most displays (if they are no longer covered by a warranty); their prices often are much lower than factory service.

Caution

You should *never* attempt to repair a CRT monitor yourself. Touching the wrong component can be fatal. The display circuits can hold extremely high voltages for hours, days, or even weeks after the power is shut off. A qualified service person should discharge the CRT and power capacitors before proceeding.