VGA—Design choices for a new video subsystem

by S. Thompson

The VGA (Video Graphics Array) video subsystem is provided as standard on the system boards of the IBM Personal System/2® Models 50 and above. VGA was designed to meet the objectives set for these new systems and to support compatibility with older IBM offerings, while at the same time providing greater performance and increased function. The IBM Enhanced Graphics Adapter (EGA) was chosen as the compatibility base for VGA, since EGA had become the video standard for IBM-compatible computer systems. Six new modes of operation were designed to meet the needs of new business and consumer applications and to improve the ergonomics of the systems. Higher-performance video presents several design problems, including electromagnetic interference, physical design size, and cost. These design problems were contained by implementing the VGA function in a single-gate array and by using an analog display interface. The use of a video digital-to-analog converter (DAC) allows the VGA subsystem to show any color from a choice of 256K colors when a color display is used, or 64 gray shades when a monochrome display is used. The VGA subsystem was designed to provide a uniform interface for color and monochrome that allows a color mode to be selected when a monochrome display is used, or a monochrome mode to be used on a color display. A color-summing algorithm was designed and implemented in the BIOS (Basic Input/Output System) software that will allow colors to be shown as shades of gray on the monochrome display.

The VGA (Video Graphics Array) video subsystem is provided as standard on the system boards of the IBM Personal System/2® (PS/2®) Models 50 and above. The VGA is an IBM-developed VLSI module that provides new video function but at the same time is compatible with previous IBM video adapters such as the IBM Enhanced Graphics Adapter (EGA). Providing video on the system board is a change from the IBM PC, PC XT, and PC AT® strategy which required that a separate video adapter be purchased and installed in an option slot. Video and other functions, such as serial and parallel ports and a pointing device port, have been included on the system boards on the PS/2 machines to provide highly integrated and full-function systems. The VGA video function includes that which was available with the EGA, plus six new video modes. Three of the new modes are alphanumeric, and provide sharper characters which are easier to read. The remaining three video modes are APA (All-Points-Addressable) graphics modes which provide higher screen addressability or higher screen color content. The new video function is designed to meet the needs of today's business applications by providing support for enhancing graphs and charts, and by reducing eyestrain that can result from continued use of a computer CRT screen.

Another new feature introduced by the PS/2 systems is the video DAC (digital-to-analog converter) which translates digital color information into the analog red, green, and blue signals used by the PS/2 displays. The DAC has a color lookup table which allows each color shown on the color display to be selected from a choice of 256K (262 144) colors. (Throughout this paper, K = 1024 decimal.) The DAC allows up to 64 shades of gray to be shown on the PS/2 monochrome display.

Another new and powerful feature of PS/2 video is the ability to select color modes when a monochrome display is present, and, conversely, to select a monochrome mode when a color display is present. This feature allows software applications to run on *all* PS/2 systems, whether a particular machine has a monochrome or a color display. This means, for example, that an application running on any PS/2 system can always select a graphics mode to display charts and graphs. A new algorithm called "color summing" is implemented in the system Basic Input/Output System (BIOS) software which will translate color information into shades of gray for use on the monochrome display. Previous IBM video adapters limited the video-mode selection to the type of display connected to the video hardware. Monochrome modes were allowed only if a monochrome display was connected and color modes were allowed only if a color display was connected. Through the use of an analog display interface and the color-summing algorithm, PS/2 video has removed this restriction.

The six new modes include three alphanumeric and three graphics modes.

What's new about VGA video?

Six new operating modes. As previously stated, VGA introduced six new video modes in addition to those that were available on previous IBM video adapters. (A "video mode" is a certain format in which the display screen will operate.) The modes supported by the VGA have been chosen to provide compatibility with the existing software base, and to provide video function allowing the PS/2 family of machines to be used for applications that were not possible with the older machines. Higher screen addressability is a requirement for today's business applications; one limitation of previous machines was the inability to display photographs or complex diagrams without expensive add-on hardware. The new VGA function addresses these needs. The six new modes include three alphanumeric and three graphics modes. All modes, including the compatible modes, offer a choice of 256K selections from the color display or 64 shades of gray from the monochrome display. All modes with a vertical resolution of 200 lines are double-scanned by the VGA to occupy 400 lines on the display screen, producing a sharper image that is easier to read. Software still sees a 200-line screen addressability, since the operation is transparent and is produced by the hardware. Scan doubling also reduces the cost of the display, because it allows the use of three vertical sizes (350, 400, and 480 lines) rather than four (200, 350, 400, 480).

All modes except the 640 × 480-pel graphics modes were designed using a vertical screen refresh rate of 70 Hz to reduce the screen flicker perceived by the user. The refresh rate for the 640 × 480-pel modes is 60 Hz. The horizontal sweep rate for all modes is 31.5 kHz, which is twice the NTSC (National Television Systems Committee) standard of 15.7 kHz for television transmissions. The horizontal sweep rate is doubled to allow the IBM PS/2 displays to run noninterlaced rather than interlaced for all video modes, thereby reducing the screen flicker. Since all VGA modes occupy 350, 400, or 480 visible lines on the display screen, the 15.7-kHz rate would require the video image to be interlaced, allowing only half of the video image to be updated at a time. The odd scan lines would be updated in the first vertical refresh period, the even scan lines would be updated in the next vertical period, and the cycle would repeat. The effect of interlacing is to cut the actual screen refresh rate in half, increasing the amount of screen flicker. A noninterlaced scheme updates all scan lines within a single vertical refresh period and reduces the amount of flicker.

Alphanumeric modes. Table 1 shows the alphanumeric modes supported by VGA. There are three versions of modes 0, 1, 2 and 3, and two versions of mode 7; the different versions represent different character resolutions. VGA introduces new high-resolution text modes that use a 9 × 16-character box. The higher-resolution text is designed to improve the ergonomics of the PS/2 systems by reducing eyestrain in heavy-use applications such as business or secretarial. A 16-line character height was chosen, since heights larger than this do not significantly improve the clarity of a 9-pel-wide character but do require much more space in the BIOS ROM. The video BIOS function call that selects the character box size is separate from the function call that selects the mode type. The separate interfaces allow older, pre-VGA applications to select a text mode while the 9 × 16-character box is in effect, thereby taking advantage of the new high-resolution modes. Upon power-up, the PS/2 systems will automatically select the 9 × 16-character box size so that applications may take advantage of highest-quality alphanumeric modes when they request a mode through BIOS. If a lower-resolution alphanumeric mode is required, the appropriate character box size can be selected by the aforementioned BIOS call. The power-up default modes are mode 3+ if a color display is present, and mode 7+ if a monochrome display is present.

A VGA functional enhancement unique to the alphanumeric modes is the ability to load up to eight different character fonts into the VGA character generator RAM. (The EGA adapter allowed only four character fonts to be loaded.) Any two of the eight character fonts can be selected to appear on the display screen at one time, effectively increasing the number of simultaneously displayable characters to 512. Normally, only 256 different characters may be displayed at one time. This feature allows special effects in text modes, such as italics, boldface, or special graphics characters. The BIOS system provides support for loading and displaying different character fonts.

Table 1 **Alphanumeric modes (new modes shown in italic type)**

Mode	Adapter	Character Size	Screen Size	Attribute Type
0, 1	CGA	8 × 8	40 × 25	color
2, 3	CGA	8 × 8	80 × 25	color
0*, 1*	EGA	8 × 14	40 × 25	color
2*, 3*	EGA	8 × 14	80 × 25	color
7	MDA	9 × 14	80 × 25	monochrome
0+, 1+	*VGA*	*9 × 16*	*40 × 25*	*color*
2+, 3+	*VGA*	*9 × 16*	*80 × 25*	*color*
7+	*VGA*	*9 × 16*	*80 × 25*	*monochrome*

Table 2 **Graphics modes (new modes shown in italic type)**

Mode	Adapter	Screen Size	Attributes
4, 5	CGA	320 × 200	4 colors
6	CGA	640 × 200	2 colors
D	EGA	320 × 200	16 colors
E	EGA	640 × 200	16 colors
F	EGA	640 × 350	monochrome
10	EGA	640 × 350	16 colors
11	*VGA*	*640 × 480*	*2 colors*
12	*VGA*	*640 × 480*	*16 colors*
13	*VGA*	*320 × 200*	*256 colors*

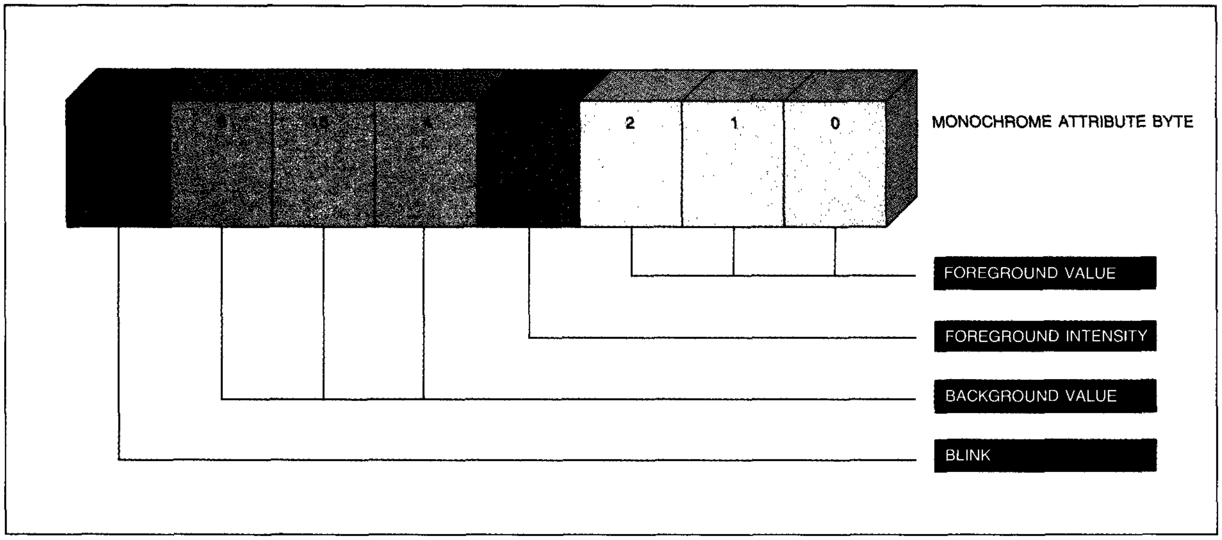
Graphics modes. Table 2 shows the graphics modes provided by the VGA subsystem, which includes all modes from CGA and EGA plus the two new 640 × 480-pel modes (modes 11 and 12) and the 256-color mode (mode 13). Mode 11 (640 × 480-pel screen addressability with two simultaneous colors) is available in both VGA and the PS/2 Model 30 video subsystem (Multi-Color Graphics Array, or MCGA). This mode is designed for applications that need 640 × 480-pel screen addressability and must run on all PS/2 systems including Model 30. Mode 12 is unique to VGA, and supports the 640 × 480-pel screen addressability with 16 simultaneous colors. This mode is designed to provide support for entry-level computer-aided design (CAD) systems, high-resolution business graphics, and imaging. The 640 × 480-pel screen addressability was chosen because it provides square pels (picture elements) on the screen. Square pels permit circles to appear round on the screen instead of elliptical, as is the case in a 640 × 200-pel or a 640 × 350-pel screen addressability, where the pels are larger vertically than horizontally.

Mode 13 is designed to provide a television-quality video image. This mode provides a screen addressability of 320 × 200 pels with 256 simultaneous colors, and is available on MCGA as well as VGA. The 256 colors can be chosen from a palette of 256K total colors, as has been stated previously. The high number of colors available simultaneously allows an image to appear to have a greater screen addressability than is actually provided. This phenomenon, which is termed "anti-aliasing," occurs when two areas of different colors are separated by a gradual color gradient instead of an abrupt color change. Photographs can be digitized by commercially available video-digitizing hardware and displayed quite accurately with mode 13, supporting applications such as personnel databases that store an individual's photo, and real estate databases that include pictures of the property for sale. The new version of IBM's Drawing Assistant program provides support for this and other VGA graphics modes.

A new (uniform) interface for monochrome and color

One of the objectives of VGA was to provide a subsystem that would support all video modes regardless

Figure 1 Monochrome attributes



of the type of display used, i.e., a uniform interface for color and monochrome. Color modes would show color on the color displays, and gray shades on the monochrome displays. Monochrome modes would show monochrome attributes on both color and monochrome displays. The uniform interface allows a low-end system configuration (PS/2 with monochrome display) to take advantage of most of the new video function. IBM's original PC video adapters only supported one type of display: The Monochrome Display was supported by the Monochrome Display Adapter, and the Color Display was supported by the Color Graphics Adapter. If an application was written to use a certain mode (for example, a color text mode), the appropriate adapter and display combination was required (the color display and adapter in the example). The IBM Enhanced Graphics Adapter supported both color and monochrome displays, but it would only allow color modes to be selected when a color display was used, and monochrome modes were only allowed on monochrome displays. By making the video modes essentially independent of the display type, VGA frees application software from the burden of supporting both types of attributes. An application is allowed to select any mode, whether it is in a monochrome or color video environment. Since the video mode is independent of the display type, the text mode attributes (monochrome or color) are also independent of the display type. This means that the "monochrome or color" referenced in an application configuration menu is *not* the same as the monochrome or color display that is connected to the system unit. An application may ask which type of display is being used so that it may select the video mode (and therefore the attribute type) that was required by the pre-VGA video adapters. Since VGA does not require that a certain display type use a certain video mode or attribute type, a user can select "color" in an application even if his or her system utilizes a monochrome display (which shows colors as shades of gray). And, of course, it is possible to select "monochrome" if a color display is connected. To understand how this is possible, one must differentiate between the display type and the attribute type.

The term *display type* refers to the type of display monitor or CRT that is physically connected to the system unit. The color display uses three color signals (red, green, and blue) to produce different colors on the screen. The monochrome display uses a single color signal and shows various intensities of a single color (white on the PS/2 monochrome display).

The term *attribute type* refers to the two sets of attributes that were introduced by the monochrome display adapter and the color graphics adapter. An "attribute" is a code that tells the video adapter hardware how to show a character on the display screen. Two types of attributes are described.

Monochrome attributes. IBM's first PC video subsystem consisted of the IBM Monochrome Display Adapter (MDA) and the IBM Monochrome Display.

Figure 2 Color attributes

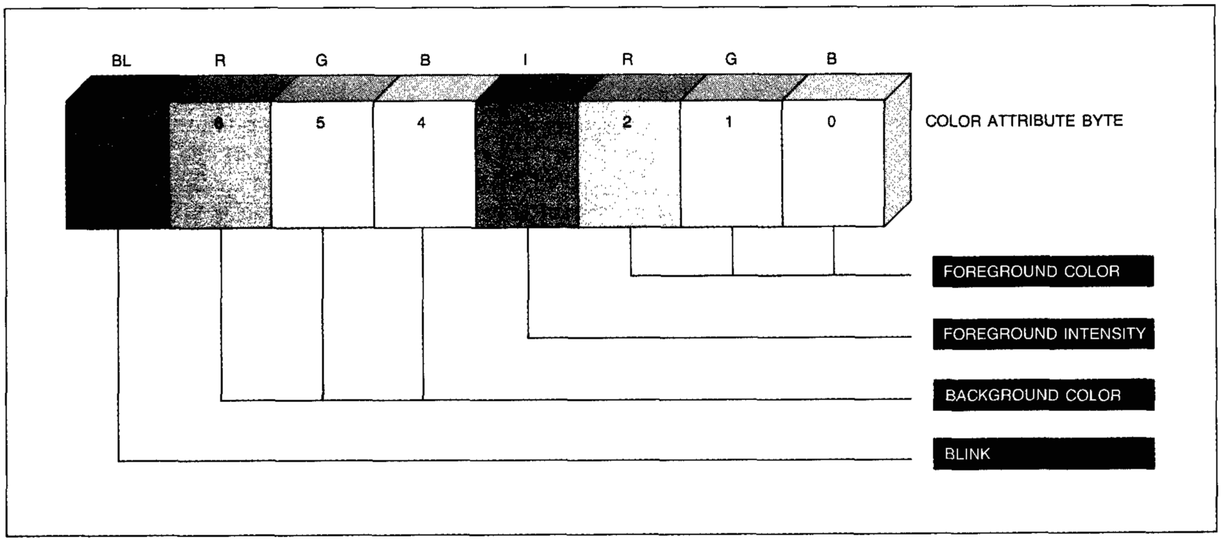


Table 3 Monochrome attribute definitions

Blink	Background	Intensity	Foreground	Results
0	000	0	000	Nondisplay
0	000	0	001	Underline
0	000	0	111	Normal Video
0	000	1	001	High Intensity Underlined
0	000	1	111	High Intensity
0	111	0	000	Reverse Video
1	000	0	001	Blink and Underline
1	000	0	111	Blink Normal Video
1	000	1	111	Blink High Intensity
1	111	0	000	Blink Reverse Video

This subsystem provided a high-resolution text mode with four available attributes defined as normal video, high-intensity video, underline, and blink. Figure 1 and Table 3 show the attribute combinations and codes defined for the monochrome video mode.

Color attributes. The IBM Color Graphics Adapter introduced color text and graphics modes for the IBM Personal Computer. This adapter provided ten attributes in text mode consisting of eight colors (black, blue, green, cyan, red, magenta, brown, and white), intensity, and blink. The attributes were used in combination to produce 16 foreground colors, eight background colors, and character blink. Figure 2 and Table 4 show the attributes defined by the CGA.

Table 4 Color attribute definitions

Blink	Intensity	Foreground/ Background	Results
0	0	000	Black
0	0	001	Blue
0	0	010	Green
0	0	011	Cyan
0	0	100	Red
0	0	101	Magenta
0	0	110	Brown
0	0	111	White
X	1	XXX	Intensify Foreground
1	X	XXX	Foreground Blink

When an application asks "What type of display is being used?" it may be asking "What video mode or attribute type must be used?" VGA now allows the user to make a choice of attributes, rather than having the video hardware dictate the choice based on display type. This capability serves to improve the user-friendliness of the VGA subsystem, and of the applications that run on it.

Another new feature—color summing

The mechanism that shows colors as shades of gray on a monochrome display is called *color summing.* Summing could have been done in software or in hardware; a software implementation was chosen for versatility and hardware cost. To perform summing in the display hardware, the monochrome display would require three analog signals rather than one, increasing the cost of the cable and the display circuitry. The software approach, besides being simpler and less expensive, allows an application to disable the summing operation and control the gray level directly. The summing mechanism is invoked when an application uses the BIOS interface to set one or more of the DAC color registers. The 16 gray shades that correspond to the 16 default colors are preselected instead of summed in order to provide maximum contrast between colors. Summing combines 30 percent of the red, 59 percent of the green, and 11 percent of the blue components of the color attribute into a single value. Summing maps the 18 bits of color information into 6 bits. The individual weighting of the red, green, and blue signals is based on the sensitivity of the human eye to those colors, and results in a gray-level picture that minimizes the amount of screen information lost in the summing operation. Figure 3 graphically illustrates the color-summing mechanism.

Figure 3 depicts the summed value being sent to the display on all three color signals, allowing colors to be summed on a color display as well as a monochrome display. Summing on a color display can be done by an application developer to see how the application will look on a monochrome display, without requiring that a monochrome display actually be present. This capability was implemented to cut down on the amount of hardware a developer must invest in. Developers should note that the summed value is programmed into the video DAC, and the original red, green, and blue values are lost. This means that if the color components are to be

Figure 3 Color summing

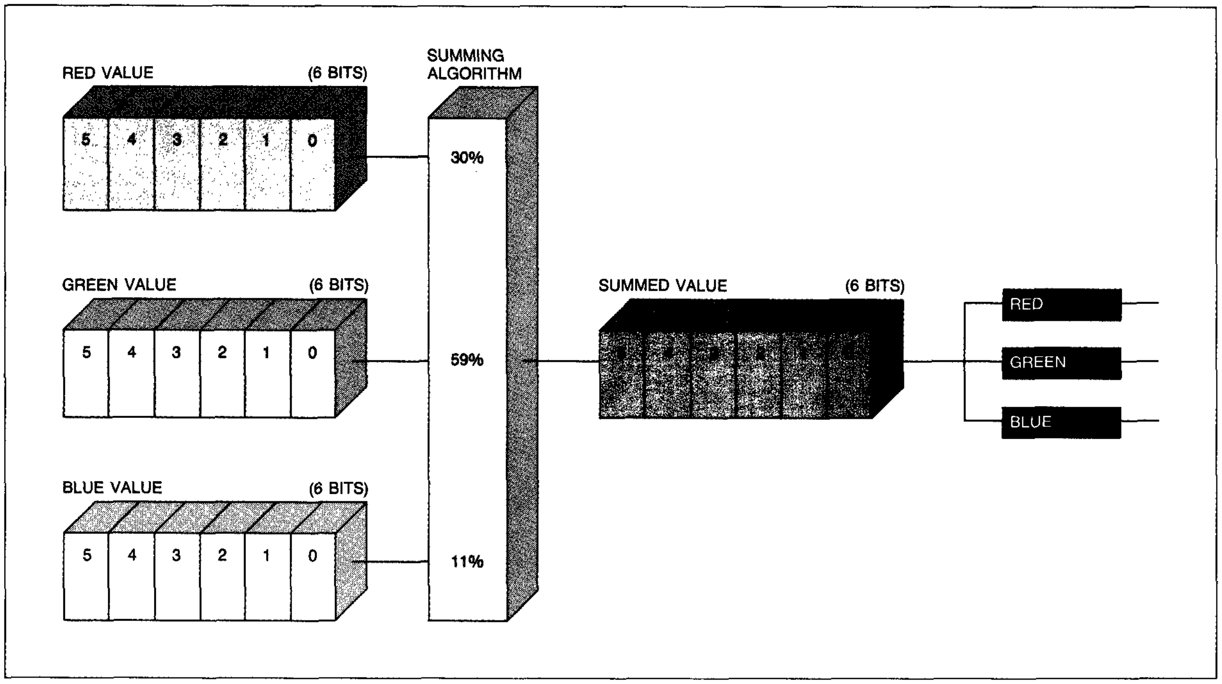
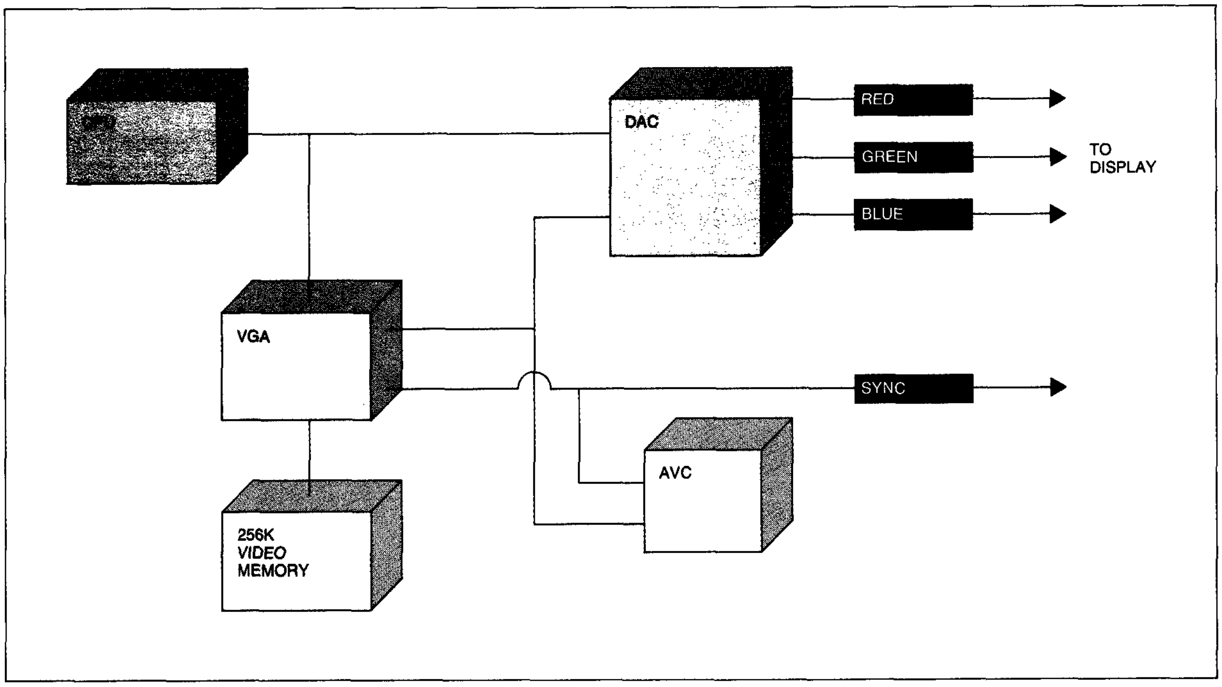


Figure 4 The VGA subsystem hardware



saved while summing is in effect, the application must store them in some way and not expect to read them out of the DAC.

The VGA subsystem hardware

The VGA and video memory. The VGA subsystem, shown in Figure 4, comprises the VGA, 256K bytes of video memory, the DAC, the analog display, and the auxiliary video connector. VGA is the largest gate array on the system board; it contains 10000 cells (a "cell" is a basic unit used to measure the size of gate arrays; it usually represents the area required for a simple logic gate), and contains the logic required to emulate the EGA and also support the six new modes. A single-chip implementation was chosen to minimize the board area required for the video subsystem and to increase the reliability of the entire computer system. Since the functional equivalent of four Personal Computer option cards as well as the CPU function was to be included on the system board, board real estate was at a premium. The video subsystem would not fit on the board if more than one gate array was required. A single-chip implementation also increases the reliability of the video subsystem through a reduced component count. Finally, electromagnetic radiation is drastically reduced with the single-chip design, since most of the switching logic is contained on an integrated circuit chip where switching currents and wire lengths are very small. Through a large effort in logic reduction and efficient design, the VGA function was squeezed into the 10000 cells available.

The large-chip design presented a problem in design verification. How could it be ensured that the design was correct? A TTL model of the VGA was deemed impractical, since such a model would be very large and difficult to maintain. Also, the TTL components available did not completely match the logic books that were used to implement the design. (Logic books are the logical building blocks such as AND gates, OR gates, etc. that make up the design.) Computer simulation of the VGA was the only practical means of verification. The task of simulating the 10000-cell VGA design was tackled by breaking it into five functional areas and simulating each section separately. Then the sections were combined, and simulation continued at the chip level. With this approach, a major portion of the design was verified,

The Micro Channel™ interface

The design of the Personal System/2® (PS/2®) Models 50, 60, and 80 introduced a new I/O structure known as the Micro Channel interface. It is an implementation of a new architecture designed to support the configuration, performance, reliability, and serviceability needs of multitasking, multiuser, and networked systems.

Configurability. In multitasking and multiuser systems, each task or user may require individual access to input/output devices (with some devices uniquely assigned to each user or task), thereby increasing the total number and diversity of devices required. Many I/O attachments exist to satisfy the large number of applications that utilize single-task personal workstations. These attachments can often conflict with one another because of simultaneous requests for the same system resource(s). In a single-task machine, concurrency is not a requirement, and these resources can be operated sequentially. Many system resources could not be readily shared in the original IBM PC architecture, and some desired configurations were not attainable.

The Micro Channel architecture introduces the Program Option Select (POS) function, which automatically resolves conflicts for system resources among attached devices. This function also eliminates all configuration switches in the system, along with the additional documentation typically required to support them. (As many as fourteen switches could have been required to prevent configuration conflicts for system resources prior to the Micro Channel architecture.) It is possible to assemble almost any combination of cards in a Personal System/2, including multiples of the same design. Thus, although standard asynchronous communication ports are normally limited to a maximum of two on an IBM PC AT® with OS/2™, a PS/2 with the Micro Channel interface supports eight ports with an expansion capability and no requirement to set switches. In this way, the Program Option Select function produces substantially improved ease of system configurability.

POS can also positively identify the type of card installed in each slot, enabling the diagnostic programs to automatically and accurately select appropriate diagnostic programs for each card. Redundant cards can even be held as active spares, waiting to take over if a card fails.

When diagnosing hardware problems, POS also permits the isolated operation of individual cards in the system, making all other cards inactive during the diagnostic operations. This effectively reduces the potential for two cards to interact, and supports rapid diagnosis with increased surety that the card identified for replacement is indeed a failing component.

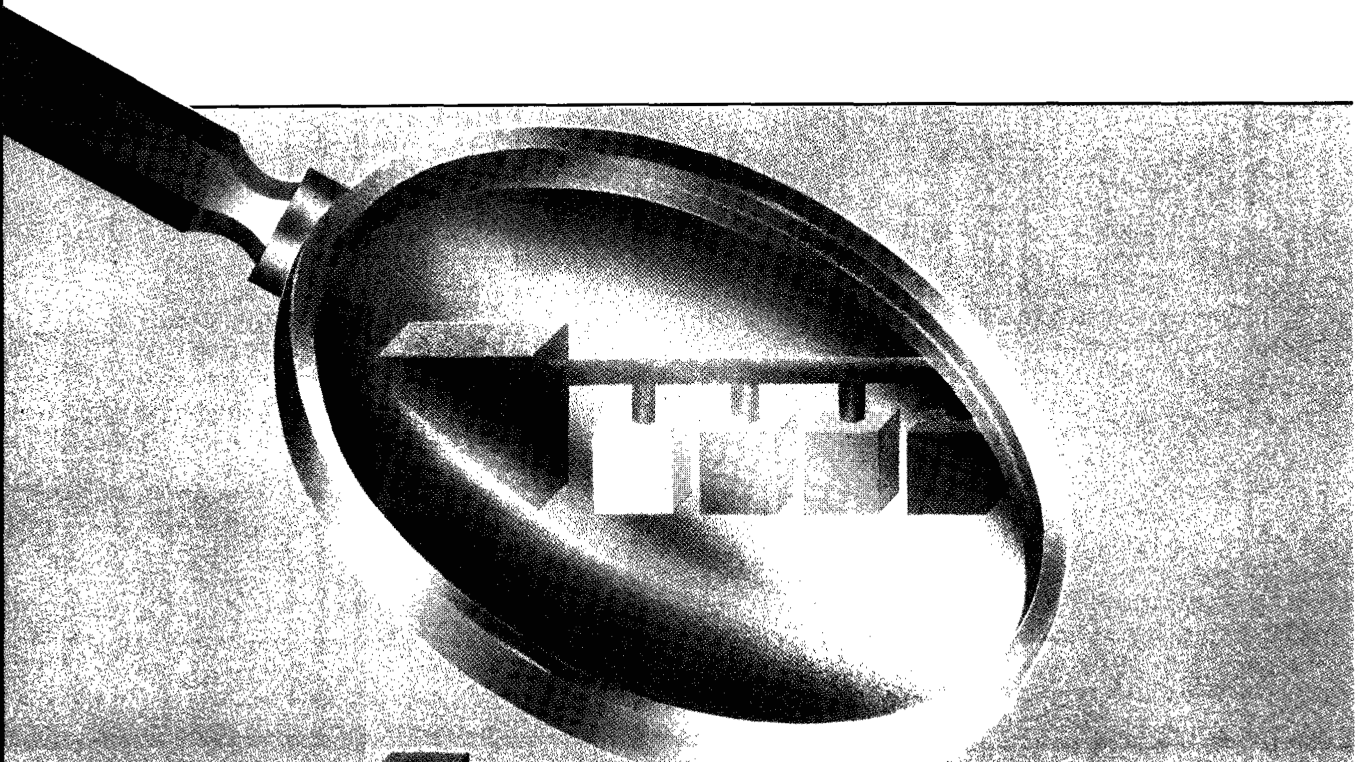
Extendability. The design of the Micro Channel bus supports new approaches to distributing function between the main circuit board on the PS/2 (the planar board) and plug-in adapter cards. For example, the high-performance graphics adapter (VGA) mounted on the PS/2 planar board can communicate with a display extension function, such as the IBM 8514A high-resolution graphics subsystem, via a section of a Micro Channel connector. This eliminates the need to repeat the VGA logic on the IBM 8514A or other display extension card, thereby reducing the cost of the display extension. The inclusion of the basic display function is an improvement over the IBM PC display extensions that do not have this capability and must implement the circuitry required to reproduce the display modes on each extension adapter card. Thus, the Micro Channel interface eliminates this need for redundant functional circuitry.

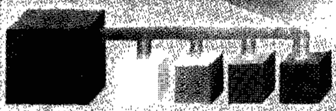
Performance. The Micro Channel bus is designed to support Multi Master capability, permitting certain attachments (such as a coprocessor adapter card) to act as the bus master. Up to fifteen bus masters can be supported in addition to the processor. These Micro Channel attachments can transmit more data in less time because they can communicate directly with I/O devices and memory without depending on the processor on the planar board. Masters can communicate with other I/O devices more efficiently than older IBM PC direct memory access (DMA) devices, and can independently access all systems memory and I/O devices attached to the Micro Channel bus.

Masters allow the efficient distribution of processing responsibility throughout the system, freeing main processor resources for other tasks. This is analogous to human organizations, where managers must be capable of delegating responsibility to others in order to allow the *system* to perform more effectively. This Micro Channel feature is an improvement over the Personal Computer architecture, which could not as easily delegate processing responsibility.

Overlapping of many I/O operations, called concurrency, usually results in increased data traffic to and from the I/O devices. The Micro Channel architecture has the ability to support bus master I/O attachments that can communicate directly with system memory or other I/O devices. These device attachments can move blocks of data, in "bursts," to and from I/O devices at nearly 20 million characters per second without burdening the processor with data transfer chores. This ability permits concurrent I/O traffic to flow with less congestion. Data can travel over 32-bit paths to I/O devices, concurrently, via the Micro Channel interface at speeds typically

Personal System/2, PS/2, and AT are registered trademarks, and Micro Channel and OS/2 are trademarks, of International Business Machines Corporation.





three to five times faster than in IBM PC- or PC AT-based systems. Since the Micro Channel interface supports bus masters, 32-bit I/O interfaces, and burst data operations, the higher data-transfer capability of the I/O device can be achieved in these concurrent operations.

When the data on a fixed-disk file are organized sequentially one sector after another, the burst capability of the Micro Channel interface allows the data to be transferred from the file attachment card to system memory as fast as the file device can operate. Thus, a cache, or temporary storage for the file data in memory, can be replenished three times faster than was possible in PC AT systems.

Reliability and serviceability. The operational concurrency of multiple I/O devices significantly increases the frequency of requests for processor attention (processor interrupts). The Micro Channel bus uses *level-sensed interrupts*, the detection of relatively stable voltage levels to request attention from the system. It is more reliable than *edge-triggered interrupts*, a detection of transient voltage swings which is used in PC buses. Reliable detection is important in multitask and multiuser systems, where a lost interrupt causes a loss in synchronization between the system software and I/O and may require a restart of the failed operation at the last successful step. If the source of the spurious error cannot be isolated, the problem cannot be prevented from recurring. Whereas the Personal Computer is limited in its ability to isolate these interrupt errors, the design of the Micro Channel bus supports an interlock between software and I/O devices which allows detection of lost or spurious interrupts.

Micro Channel architecture defines a mechanism whereby intermittent errors in memory or other feature cards are identified during normal operations. The errors may be *logged* or recorded for subsequent analysis by appropriate system software. By reviewing the error log on all types of infrequent—but potentially devastating—failures, service personnel can more quickly locate and replace the failing component.

By improving reliability and reducing intermittent system failures, the Micro Channel interface improves overall data and system integrity to a level appropriate for the new generation of complex systems. The result should be substantially fewer service calls, and increased productivity and satisfaction for system users. Many PS/2 systems will be used as high-performance shared systems in which data integrity is of the highest priority. The Micro Channel architecture provides the capability to reduce catastrophic system errors and to diagnose and repair the system in an expedient manner. Whereas performance and compatibility comparisons are readily apparent, data integrity, reliability, and maintainability must be inferred from an analysis of the system design. As systems become more complex and performance increases, data integrity becomes one of the most important system characteristics.

Chester A. Heath

and design errors were quickly spotted, corrected, and new logic resimulated. With a design as programmable as the VGA, the number of possible hardware states is virtually infinite. Since it would be impossible to simulate all hardware states, only representative states were chosen. The different sections of the design were simulated over complete hardware cycles in the representative states. The job of VGA simulation was done by several engineers and required the power of a mainframe computer.

One of the new functions that contributes significantly to the size of VGA is readable registers, which facilitate hardware save/restore-state operations needed in a multitasking environment. Since most applications use the display screen directly, supporting save/restore in the video hardware was a must. Having readable registers also makes possible a read-modify-write operation which allows an application to modify selected bits in a register without disturbing the rest. The readable-register function was implemented using a treed multiplexer scheme rather than an internal three-state bus approach. There are no three-state buses inside VGA, since this type of bus cannot be tested well enough to meet IBM's high testability requirements.

So that the full VGA function is available on all machines without requiring a memory upgrade, 256K bytes of video memory is provided as standard. Applications for the PS/2 machines no longer need be concerned about how much memory is available in the video subsystem, i.e., whether or not they can use the high-function VGA graphics modes. Providing the full memory configuration as standard also serves to increase the reliability of the video subsystem. Since the memory modules can be soldered directly to the system board without using sockets, module-insertion errors and bad connections are eliminated.

The combination of VGA and the 256K bytes of video memory is comparable to a fully configured EGA card. The VGA subsystem adds two new items: the Digital-to-Analog Converter (DAC) and the auxiliary video connector.

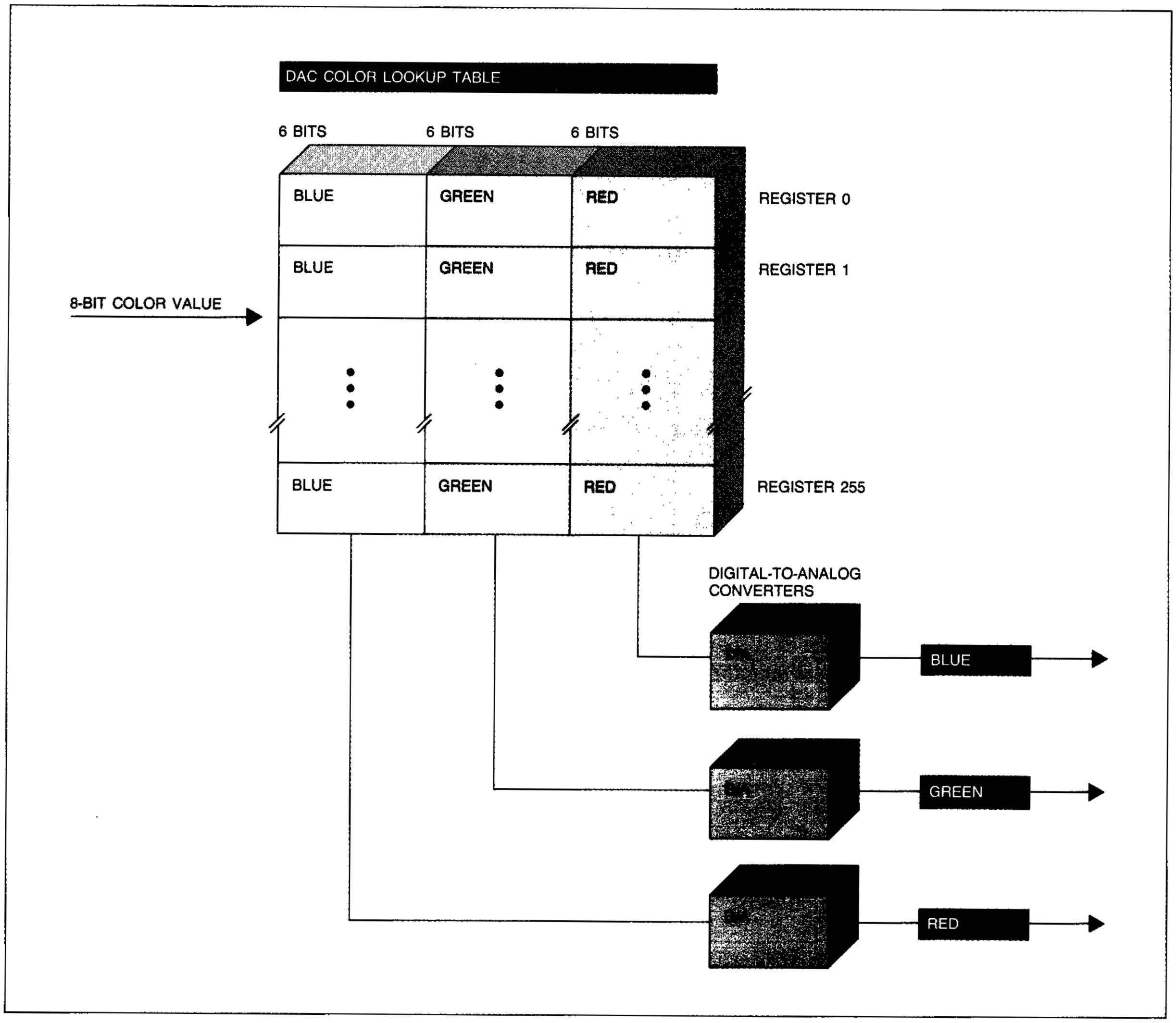
The video digital-to-analog converter (DAC). The DAC converts the 8 bits of digital color information from the VGA chip into the analog signals used by the new PS/2 displays. This particular DAC was chosen because it provides a color lookup table that will translate an 8-bit color value into three analog color signals with 6-bit resolution on each signal, allowing color selection from a palette of 256K colors. Figure 5 illustrates the operation of the DAC.

The 8-bit color value from VGA is used as an address to select one of the 256 color registers in the DAC color lookup table. Each of the color registers in the DAC is 18 bits wide and contains 6 bits for the red component, 6 bits for the green component, and 6 bits for the blue component. The 6-bit color components are then converted to analog signals by three digital-to-analog converters. The color lookup table is programmed by BIOS or by an application to produce the desired colors on the screen. The lookup table is also readable to satisfy the save/restore state requirement of multitasking. For compatibility, BIOS programs the DAC to produce EGA-compatible colors for all modes except the 256-color mode (mode 13 hex), so that applications that write to the EGA-compatible palette within VGA will produce the same results on VGA as on EGA. Of course, it is possible to change these colors by modifying the DAC color lookup table via a new BIOS call provided for this purpose. It is recommended that applications use the BIOS interface to write to the DAC, since the BIOS can convert the color information to shades of gray if a monochrome display is attached to the system. This conversion mechanism is the color summing addressed earlier in this paper.

The auxiliary video connector

The auxiliary video connector (AVC) is an extension of the Microchannel connector, which has been designed to allow for easy expansion in the video area and to eliminate the "throwing away" of hardware that in the past was replaced by a new card. The AVC provides access to the digital interface between VGA and the DAC by an option card, which may receive the digital signals from the VGA or may drive the DAC (and therefore the display) with its own digital signals. The AVC allows for video option cards that are incremental in nature; that is, the option card need not provide the VGA function for compatibility, since the VGA on the system board can still be used. The IBM 8514/A display adapter is an example of a video option card that is incremental in nature. This adapter provides two increased-function video modes for a display attached to the adapter: a 640 × 480-pel 256-color mode (available with any PS/2 color display) and a 1024 × 768-pel 256-color mode (available with the 8514 display only). The 8514/A will take the video from the VGA on the system board via the AVC and show it on the display connected to the 8514/A. When an 8514/A advanced-function mode is selected, the adapter will generate its own video signals and ignore the VGA signals.

Figure 5 Video DAC block diagram



Why use an analog display interface? The VGA video subsystem uses analog displays, a change from the older IBM video adapters which used digital displays (MDA, CGA, and EGA). One may wonder why IBM has chosen to support analog displays when the television industry seems to be tending toward digital television. The answer is that VGA is digital where TV is digital, and analog where TV is analog. TV uses digital processing to operate upon the picture information sent by the broadcast station. The digital data are then converted to analog signals for use by the cathode ray tube. IBM has chosen analog displays for several reasons, including the number of colors displayable, cable size, electromagnetic compatibility, and expandability.

In a digital display interface, the number of required digital signals increases as the number of displayable colors increases. The relationship is as follows:

number of signals = log base 2 (number of colors),

or

2 (number of signals) = number of colors.

For example, in CGA 16 colors were available, yielding number of signals = log base 2 (16) = 4,

or

2 (4 signals) = 16 colors.

In the EGA, 16 colors out of a choice of 64 colors were displayable. Although we can only display 16 colors at a time, the display interface must allow us to present 64 different color codes to the display, so the enhanced color display used six signal lines:

The VGA subsystem allows up to 256 colors to be displayed simultaneously.

number of signals = log base 2 (64) = 6.

The VGA subsystem allows up to 256 colors to be displayed simultaneously from a choice of 256K colors. A digital implementation of 256K colors would require 18 video signal wires. Such an interface would require a fairly large connector and cable, and would present problems in the areas of electromagnetic compatibility, reliability, and usability. Each wire in a cable acts as an antenna, producing electromagnetic radiation on which the FCC has set strict limits. The goal set by the FCC is to limit the amount of ambient radiation; many wires radiating at high video frequencies runs contrary to that goal. Since there are only three analog video signals, coaxial cable can be used cost-effectively. Coaxial cable allows the impedance of the video cable to be matched to that of the system board video output (the DAC and loading resistors) and the display input. Impedance matching reduces signal losses, reflections, and noise, producing a cleaner, sharper video image on the display screen. The coaxial cable itself also drastically reduces electromagnetic radiation from the video signals, since each signal is surrounded by a ground shield. An analog interface improves reliability, because the larger number of wires required by a digital interface presents a greater probability of cable failure due to a broken wire. Finally, usability is adversely affected by a digital implementation, since a large cable is stiff and difficult to control.

Another reason to select analog displays is future expandability. A digital display shows only the number of colors supported by the number of signal lines in its cable. An analog display provides a continuum of colors, the number of colors displayable being almost infinite. The number of color choices for an analog display is determined by the video adapter itself. Currently VGA allows a choice of 256K colors; future video hardware could increase that number and still use the current analog displays. The VGA achieves 256K colors by providing three analog color signals (red, green, and blue) that can take on any one of 64 different values. Thus,

(64 red values) · (64 green values) · (64 blue values) = 262 144 colors.

The monochrome display uses only one color signal, so it can display 64 gray levels:

(64 color values) = 64 shades of gray.

The analog displays accept color signals whose voltages range from 0.00 volts (black level) to 0.70 volts (white level). The VGA divides this voltage range into 63 steps (64 sequential color values define 63 color steps). Each step is

0.70 volts/63 = 11.1 millivolts.

VGA controls the analog color signals with voltage steps of 11.1 millivolts. If a future video adapter used smaller voltage steps, a larger number of steps would fit in the 0.00- to 0.70-volt range and would therefore produce a larger choice of colors on the existing displays.

Concluding remarks

Through the use of VLSI (Very-Large-Scale Integration) and an analog display interface, the VGA subsystem provides a high level of video function, performance, and reliability that opens new ground for personal computer applications. The single-chip design of VGA allows the entire video subsystem to be included on the system board and improves the EMI profile of the machines. The analog interface allows a large color selection and a uniform interface for color and monochrome, which allows any video mode to be used on any display, whether color or monochrome. VGA has also been designed to meet the requirements of compatibility and the support of multitasking. The PS/2 systems as a family have been designed to meet the needs of the business and consumer applications, both now and in the future. The VGA is an example of a subsystem whose design choices were made with these needs in mind.

Stephen Thompson *IBM Entry Systems Division, P.O. Box 1328, Boca Raton, Florida 33432.* Mr. Thompson is a senior associate engineer and codesigner of the IBM Video Graphics Array. He has presented the VGA subsystem at seminars in the United States and in Europe, and is credited with several innovations in the areas of video and memory subsystem design. Mr. Thompson joined IBM in 1983; he received a B.S.E.E. degree from Wayne State University in Detroit, Michigan.

Reprint Order No. G321-5318.