IBM 9.2-Megapixel Flat-Panel Display: Technology and Infrastructure

Steven L. Wright,*
IBM T.J. Watson Research Center

ABSTRACT

Recently, IBM® introduced a very high-resolution high-information content flat-panel display, which incorporates a number of technological advances. This display contains 9.2 million pixels at a density of 204 pixels per inch, in a 3840x2400 pixel format. This is the first large-area display with a pixel density which satisfies normal visual acuity at normal reading distances, and increases the screen information content by a factor of 5 to 7 times over conventional displays. This display is appropriate for a wide variety of applications in which visualization of a large amount of data is important. At the same time, this technology advance heralds a number of infrastructure changes in system architecture, digital data protocol, and user interface design for both operating system and software applications.

Keywords: TFTLCD, LCD, liquid-crystal, flat-panel, displays, high-resolution, pixel density, high-information

1. INTRODUCTION

Flat panel display technology has evolved to the point that thin-film-transistor liquid-crystal displays (TFTLCDs) can outperform cathode-ray-tube displays (CRTs) in nearly all respects. In particular, the information content of TFTLCDs can now far exceed the capabilities of CRT technology. Over the past decade, the display group at IBM T.J. Watson Research Center has built several prototype displays with increased information content.[1] In 1996, this group made a 157 pixel per inch (ppi) 10.4" SXGA prototype display.[2] In 1998, this group fabricated a prototype 16.3" TFTLCD, code-named "Roentgen", which had 201 ppi and a pixel format of 2560x2048 (QSXGA).[3] Both of these prototypes had twisted-nematic (TN) liquid-crystal mode which limited the applicability of this technology. In 2000, a prototype 22.2" TFTLCD, code-named "Bertha", was made in a joint effort between IBM Research and IBM Japan, which utilized the dual-domain in-plane-switching (IPS) liquid-crystal mode. This display had a pixel format of 3840x2400 (QUXGA-W) with 204 ppi.[4] A small number of these prototype displays were delivered under contract to Lawrence Livermore National Laboratories, starting in Sept. 2000. In June 2001, the first product version of this display was announced. The interface electronics section of this product was improved, and a follow-on replacement product was announced in Nov. 2001. International Display Technology, Ltd., a subsidiary of IBM and Chi Mei Optoelectronics Corporation, currently manufactures this display. Some of the specifications for the 9.2 Mpixel display product are given in Table 1.

Feature	Specification
screen size	478 x 299 mm, 560 mm diagonal
pixel format	3840x2400 (QUXGA – Wide)
number and density of pixels, subpixel size	9.22 million, 204 ppi, 0.0415 x 0.1245 mm
number of colors, brightness, contrast ratio	16.7 million (24 bit), 235 cd/m^2, 400: 1
viewing angle	> 170^o (CR>15)
response time (sum of on and off times)	50 ms
LC mode, aperture ratio	Dual-domain IPS, 28%

Table 1. Specifications.

This display can be used with stand-alone workstations as well as high-end visualization systems with multiple rendering nodes. Applications include computer-aided design, geographic information systems, image analysis, digital pre-press soft proofing, digital camera image processing, medical imaging, finance and securities trading, intensive programming, and data visualization. Examples of high-end data visualization include thermonuclear modeling, seismic data analysis, weather modeling, and flight simulation.

For avionics, liquid-crystal displays are widely used, with very high brightness and dimming ratio. For large-area, very-high resolution panels such at that described here, the system integration issues and optical and other performance requirements will likely delay incorporation of this technology into the aircraft cockpit. However, there are other airborne applications for this technology which can be envisioned in the near future, such as command and control, intelligence data analysis, weapons systems, navigation, radar and weather analysis; perhaps working in conjunction with ground-based systems using similar display technology.

This paper discusses some key technology components, which converged to make this display possible, and some of the infrastructure issues. The challenges IBM faces introducing this new technology are primarily not with the manufacturing technology or design of the display itself. The increase in pixel density, information content and data bandwidth introduce a number of infrastructure issues outside the display. These issues include both hardware and software. Table 2 shows how display performance has historically lagged behind all other computer hardware performance. As digital standards and the need for higher data bandwidth evolves, the graphics hardware performance continues to improve. It should be recognized that a major driving force behind this development is to enable video computer games, not to meet specific industrial or military visualization requirements. Operating system and application user interface design currently lag far behind hardware capabilities. For high-resolution display technology to succeed, the user interface must be fully scaleable with pixel density.

Component	**1996**	**2002**	**2002/1996**
CPU	32 bit, 133 MHz	32 bit, 1.7 GHz	13 x
Memory	16 MB	512 MB	32 x
Hard disk	1 GB	60 GB	60 x
Communication	28.8 K bps	11 M bps	393 x
Display	SVGA (0.5 Mpixel)	SXGA (1.3 Mpixel) QUXGA-W (9.2 Mpixel)	3 x 19 x

Table 2. Relative change in computer technology component performance over the past six years.

The technological advances in design and process of the pixel array glass and module which enabled the manufacture of the IBM 9.2 Mpixel display have been described in the literature.[3-8] Other advances in interface electronics and digital interfaces will play a key role in future growth of this display technology. Some of these advances are listed in Table 3.

Technology	**Impact**
dual-domain IPS	wide viewing angle characteristics
liquid crystal “one-drop fill”	four-sided drive, uniform cell, reduced cost
photolithographic post spacers	uniform narrow cell gap, reduced light leakage, enables one-drop fill
ion-beam alignment	uniform cell, reduced cost
aluminum alloy gate metal	reduced gate line RC delay
advanced array design and materials, polymer film on array	high aperture ratio, reduced parasitics, reduced liquid crystal switching speed
FPGA input electronics, display frame buffer	reprogrammable interface, multiple input modes, pixel replication
dual-DVI interface	reduced cabling, increased data refresh
Digital Packetized Video Link	advanced data protocol w/reduced bandwidth requirements, daisy-chain displays
Visualization Frame Buffer	low-cost and flexible DVI compositing
Scaleable Graphics Engine	low-cost, modular high-end remote visualization, based on ethernet

Table 3. Key hardware technology components which enable high-resolution display manufacture and deployment.

2. ARRAY DESIGN AND PROCESS

To achieve acceptable viewing angle characteristics, an alternative to TN liquid crystal mode must be used. The dual-domain IPS mode has emerged as a leading wide-viewing angle technology. A major design and manufacturing challenge is to produce a high-resolution pixel array with dual-domain IPS mode with acceptable aperture ratio. This has been achieved with an advanced array process, utilizing a polymer insulator film. This insulator allows conductive layers to be offset and overlapped, thereby achieving a 28% aperture ratio and also reducing parasitic capacitance To achieve a narrow and uniform cell gap with reduced light leakage, traditional spacer balls have been replaced with photolithographic post-spacers. Prototype versions of the 9.2 Mpixel display have been produced using a liquid crystal "one-drop fill" process combined with a new process for liquid crystal alignment which does not require rubbing. In this process, an ion-beam is used to chemically alter the surface of a thin diamond-like carbon layer to align the liquid crystal. These technological improvements combine to improve performance and reduce manufacturing cost.

The connections to the data lines are a "dual-bank" interdigitated configuration, in which the columns are addressed either from above or below, shown conceptually in Figure 1. In the Roentgen prototype display, alternating groups of six lines were configured in this fashion. The rows have gate driver chips connected on both sides, using "double-sided" drive. This configuration enables relaxed ground rules for input/output fan-out, reduced signal distortion and cooling requirements, but makes the liquid crystal fill process somewhat more difficult. This configuration also enables the use of low-cost techniques for in-process repair of line defects,[10] but high manufacturing yields have obviated the need for repair. Over 5 km of wiring a few microns in width are used in the array, without a single break. There are a number of issues with materials and array parasitics, affecting the detailed electrical characteristics. The use of advanced array design and process to make short-channel length TFTs and to reduce parasitic capacitance is important. A critical requirement to drive a large number of rows within each frame time is the use of high-conductivity Al-Nd alloy gate metal with double-sided row drive. This reduces gate line delay by reducing both parasitic resistance and capacitance. In the future, even higher conductivity Cu gate metal has potential application.[2]

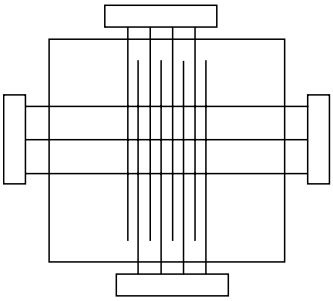

Figure 1 Drive Configuration

3. ERGONOMICS AND MONITOR DESIGN

3.1 Pixel density

A Snellen visual acuity of 20/20 corresponds to the eye's resolution of features spanning an angle of 1 arcmin. This visual acuity is an average over the population distribution. For children and young adults the average visual acuity is about 20/16, and it is not uncommon for individuals in occupations demanding high visual performance to have a visual acuity of 20/10. To obtain a driver's license in the United States requires 20/40 visual acuity. Typical reading distances are 40 to 50 cm, and at a viewing distance of 40 cm, a display with a pixel density of 100 ppi can render features as small as 2.2 arcmin. This means the "effective display acuity" at this distance is approximately 20/40. Any information attempted to be rendered which demands a higher visual acuity than 20/40 will be lost. This characteristic is borne out in measurements of visual acuity using displayed letter charts or word charts.[11] At large viewing distances, character recognition is limited by the subject visual acuity, but as the viewing distance is decreased, approaching the minimum accommodation distance, the recognition of small letters becomes limited by the display pixel density.

Typically, printed text letter sizes are chosen to be a factor of two to three larger than that size which is just recognizable with 20/20 visual acuity. A factor of two "acuity reserve" corresponds to letters approximate 8-pt size, the smallest letters that typical color CRTs can render. For printed letters this size and larger, legibility, reading speed, and user comfort reach a plateau. However, many printed documents have fine print as small as 4-pt, the "legally legible" size requiring 20/20 visual acuity to resolve. For newspapers, letter sizes as small as 6-pt are commonly used for sports scores and financial data.[11]

For Roman letters, a minimum of 5 pixels is needed in the vertical direction to render recognizable letters. Typically, 6 pixels are needed to render letters tested to be fully recognizable by subjects. At a viewing distance of 40 cm, one arcmin corresponds to a pixel density of 218 ppi, allowing recognizable 4-pt letters to be rendered. For Japanese characters, an even higher pixel density is needed for high quality. For 3.5 mm 24-dot Kanji characters, a minimum pixel density of 175 ppi is required.[12]

In a legibility study sponsored by IBM in 1998, no differences in reading speed or comfort were found for 11-pt letters rendered on a 12.1" 83 ppi notebook TFTLCD, 157 ppi 10.4" TFTLCDs and 100 ppi CRT, viewed at normal reading distances. However, there was a strong subject preference found for the high resolution TFTLCD over the other displays used in the study. In a second 1999 study with smaller letters, differences in reading speed and letter-counting speed were found for the different displays. Analysis of the data in this second study indicated that for 8-pt letters, both reading speed and letter-counting speed increased very roughly by 10-15% as the pixel density increased from 90 ppi to 160 ppi. However, more work needs to be done to confirm these results. Currently, the United States Display Consortium is sponsoring[12] a human factors study

While the majority of printed text uses letter sizes 8-pt and larger (figure captions in this paper are 9-pt letters), the ability to render smaller features is not useful solely for the purposes of displaying legible fine print. As the pixel density increases, the quality of letter serifs and natural image quality improves. While the effect of serifs on reading speed has been controversial, there is no doubt about the effect of pixel density on natural image quality. Accurate rendition of fine textures and patterns require pixels spanning at most one arcmin. In the case of fine, nearly parallel features, the human visual system can resolve offsets as small as 6 arcsecs. This vernier acuity (or "hyperacuity) is an order of magnitude more sensitive than the ability to resolve isolated features. For some fine patterns, extremely high pixel density would be required for the human visual system to resolve small offsets.

Clearly as the pixel density increases from 100 ppi to 200 ppi, the quality of both natural images and text sharply improves, with indications that both productivity and user comfort increase. It is this factor of two "effective acuity" which has the potential to be exploited. Unlike print, where there is little or no inherent grayscale capability, there is little incentive to further increase display pixel density beyond about 200 ppi. Although it is technically feasible to fabricate 300 ppi displays using amorphous Si TFTLCD technology, the small additional improvements in image quality probably do not outweigh the technical difficulties.

3.2 Screen characteristics

Historically, computer display pixel formats have followed a progression in which the aspect ratio is either 4:3 or 5:4. Each generation of pixel format typically represents an increase of pixel count by about 60% over the previous generation. Both of these aspect ratios are nearly square, and thus fit well with what CRT electron gun and electron optics design can most easily provide: a nearly square screen on the surface of a cube. Deviations from these general dimensions significantly increase the manufacturing cost of CRTs. However, for flat-panel displays, these restrictions do not apply. The screen can be made as large as the motherglass layout will allow, with the combination of total number of pixels and pixel density limited by the materials and lithographic process capabilities. It is clear that the standard 4:3 or 5:4 pixel formats have limited utility for rendering either pages of text or moving pictures for entertainment. The standard HDTV 1920x1080 format follows a 16:9 aspect ratio. This wide-screen format fills more of the horizontal visual field, thereby contributing to a sense of "presence" for the observer.

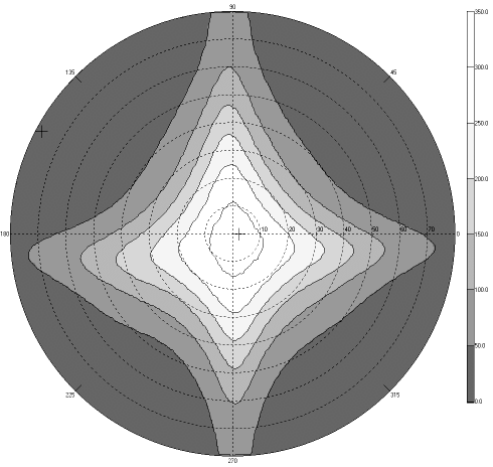

Figure 2 Polar plot of contrast ratio for dual-domain IPS, with 50:1 isocontours.

The IBM 9.2 Mpixel display has a screen diagonal of 22.2" with an aspect ratio of 16:10. This size is large enough to fully accommodate two full pages of standard 8.5"x11" or A4-sized paper. At the same time, this wide screen format also encompasses four HDTV formats, important for film editing and viewing. The 3840x2400, 16:10 format of this display allows for pixel replication of both standard 1600x1200 computer format and standard 1920x1080 HDTV format.

At a viewing distance of 40 cm, the corner-to-corner viewing range is +/- 35 degrees. An important technology requirement is that for a single viewer, variations in color or contrast viewed from different regions of the display are minimized. Figure 2 shows a polar plot of typical contrast ratio data for dual-domain IPS mode. The on-axis contrast ratio for this particular display was about 330:1. At a viewing inclination angle of 40 degrees, the average contrast ratio remained larger than 100:1. With the dual-domain IPS mode, the variation of just-noticeable-difference ΔE_{lab} values have been measured to be less than about 6 for the bright state over a viewing cone of 40 degrees extended over all azimuths. Since these values are smoothly varying over the viewing cone, the variation is hardly noticed, and also compares favorably to color accuracies achievable with print.

3.3 Moving picture quality

The moving picture quality of CRTs is excellent, and further improvements in LCD motion picture quality are needed to match that of CRTs. Recently, the switching speed of most liquid crystal display materials and modes has increased to the point that image smearing no longer remains as the dominant problem, including many TN mode displays. However, other aspects of TN mode can cause moving image artifacts. One characteristic of TN mode is that the turn-on time is much slower than the turn-off time, and another is that the switching time between two closely spaced levels is much larger than the time to switch between widely separated levels. The dual-domain IPS mode used in the IBM 9.2 megapixel panel has symmetrical switching behavior, and the switching speed between levels is largely independent of the separation of the two levels. However, for conventional IPS displays, with low pixel density, acceptable switching speeds can only be achieved using very high pixel voltages, beyond the range of suitable driver integrated circuits. However, for the high-density pixels in the IBM display, both small inter-electrode spacing and reduced cell gap are combined with improved liquid crystal material properties to achieve acceptable switching times.

To obtain blur-free moving images, the switching speed must be at least as fast as the data refresh rate, preferably a factor of two or three faster. Good motion perception is obtained with film projection at 24 frames/sec and better moving image quality is achieved at 30 frames/sec video rates, corresponding to an upper limit for single-transition switching time of 33 to 42 ms. The present dual-domain IPS switching time of about 23 ms for single-transition between 10% and 90% of luminance difference, is marginally adequate to suppress motion blur.

Improvements in liquid crystal materials can further increase the switching speed, but even if the response time of the liquid crystal material were instantaneous, there would still be motion artifacts due to "hold-type" phenomena. During each screen refresh, the luminance of a well-designed LCD pixel remains relatively constant. This means that the pixel luminance is latched at a constant value until the next addressing cycle in the subsequent frame. As the eye smoothly tracks a target, this introduces a periodic error in the location of a moving feature. This behavior is very different than CRTs, where the pixel luminance lasts only a few milliseconds each time the electron beam excites the pixel phosphor. This "strobe"-like nature of CRTs reduces the periodic integrated error in smooth eye tracking, but also introduces flicker. In film projection, each frame is flashed two or three times to suppress flicker, but this flashing also leads to small periodic error. At this time, there appears to be incomplete agreement about the relative importance of hold-type characteristics on motion perception.

One active area of research and development for TFTLCDs is to improve moving picture quality by use of a pulsed backlight or pulsed data.[13-16] The quality of moving pictures improves at the expense of brightness reduction and the re-introduction of flicker. One technique being explored is to process the pixel data so moving portions of the image are modulated.[17] While both slow response speed and latching characteristics limit the quality of moving images on LCDs, these problems are important primarily for rendering the fastest moving images in movies or for very rapid target or signal identification such as for weapons systems. For high-resolution display use in analysis of large image files, the performance of the graphics card and software application are of greatest concern.

For image analysis, it is generally desired that smooth roam rates of about 1/2 of a screen width per second be achieved, corresponding to about 600 pixels per second, or about 10 degrees/second. At a data refresh rate of 30 Hz, this means that regions of the moving image cannot be updated with a spatial precision greater than about 20 pixels, similar to the latching characteristics described earlier. Human visual acuity begins to drop for images moving faster than a few degrees/second, so during panning of an image, sharp detail cannot be distinguished. Although there are issues for moving image quality due to properties of the liquid-crystal materials contributing to motion blur, the largest impact on moving image quality during roam is image tearing or breakup. This breakup occurs due to limitations of the graphics card, computer system hardware, and software application. When portions of the moving image are blacked out or streaked, this has a very disconcerting effect. The highest-performance graphics cards available today are typically designed for optimal game performance, utilizing hardware acceleration for 3D texturing and certain motion effects. However, rapid panning of an image without breakup requires fast bitblt performance. Clearly, the application software and graphics acceleration hardware must be designed to work well together for optimal performance.

The issue of roam rate performance arises due to the way in which image data is observed, using displays with low pixel density and comparatively low pixel count. If the target position is approximately known, then a telescopic approach is efficient- small regions of the image data can be viewed under magnification, roaming over a small area. However, if a large region needs to be scanned to locate and identify a target, then a large field of view can be utilized for initial search, provided the image quality is matched to the visual acuity of the observer. In nature, raptors have a highly developed visual system with high acuity. Hawks, eagles and falcons soar at moderately high altitudes to search for prey over a large field, even though the time required to dive to the ground is larger than flying at a lower altitude. Raptors may have evolved to optimize their search by at least partially stabilizing their retinal image during flight.[18]

For the IBM 9.2 Mpixel display, the pixel count and potential information content is increased by a factor of 4.8 over standard 1600x1200 monitors. The pixel density is matched closely to what can be resolved by normal 20/20 visual acuity. These characteristics call for a different search mode, which fully utilizes both the capabilities of the monitor and the visual acuity of the analyst. Since the image viewable search area is larger, the roam rate needed can be reduced. In this way, the overall large-area image data is more static and less image manipulation is required. As suspected targets are identified, the method by which more detail is obtained would then primarily be through a series of zoom magnifications, similar to what is done in nature. In this way, smaller regions of the image data around suspected targets are manipulated. Through proper design of the application user interface, the zoom process could be made as comfortable as the panning process, retaining similar "contextual" information of the image for comfortable orientation.

5. DRIVE ELECTRONICS

One of the limiting factors for widespread use of high-resolution flat panel displays is the development of the monitor interface drive electronics. There is no intrinsic compelling reason to use analog input signals other than to maintain compatibility with CRTs. The cost of low-voltage digital electronics is lower than high-voltage analog, and the cost and development time of building analog or hybrid analog/digital input circuitry has been an inhibitor. As developers gain experience working with digital interfaces, and these interfaces gain broader acceptance in the market, these problems should diminish. Digital drive has a number of advantages at the system level, as discussed later. For high resolution flat panel displays there is no choice: the bandwidth requirements far exceed what can be straightforwardly be done with analog circuits. We next briefly describe the drive electronics system, working outward from the pixel array.

The pixels are charged up and down through amorphous-Si TFTs. The addressing of pixels is a digital process, but the data provided to each pixel is analog. Each TFT has a drive current in the vicinity of 1 μA, with turn-on times less than 1 μs, and complete turn-off times in the vicinity of 1 ms. The pixel node capacitance is less than 50 fF. For 2400 rows to be addressed at a typical frame rates, each pixel capacitance must be charged and discharged in a few microseconds.

The data signals are applied to each column through data driver chips attached to the glass by tape-automated-bonding technology. Some manufacturers used chip-on-glass bonding technology. Each driver chip FIFO accepts digital LVDS input and an 8-bit scheme with non-linear voltage taps is used to establish the output signal levels. For IPS mode panels, typically 15V driver chips are used, with output voltage differences between the different levels as small as about 10

mV. The clock rates for typical data driver chips are in the range of 65 MHz. For medical imaging and other critical applications, 10-bit grayscale is needed. However, it is not clear if the performance/cost ratio is high enough to warrant large-scale production of 10-bit driver chips. The most straightforward way to achieve higher bit depth is through a combination of spatial and temporal dithering techniques.

A block diagram of the drive electronics in the latest product version is shown in Figure 3. The LVDS data signals to the panel module are divided into four vertical stripes of 960x2400. This hardware partitioning is transparent to the mode of operation "seen" by the graphics cards and operating system. To maintain stable and consistent front-of-screen performance, the screen refresh rate in the module is maintained at a constant value of 41 Hz, regardless of input data rate. In principle, screen refresh rates for TFTLCDs can be very low, depending upon the application. We have driven notebook computer displays as low as 20 Hz, to reduce power consumption. However, depending on the detailed array design, and data polarity inversion scheme, minimum refresh rates of 30-35 Hz are typically needed for TFTLCD monitors to avoid image artifacts.

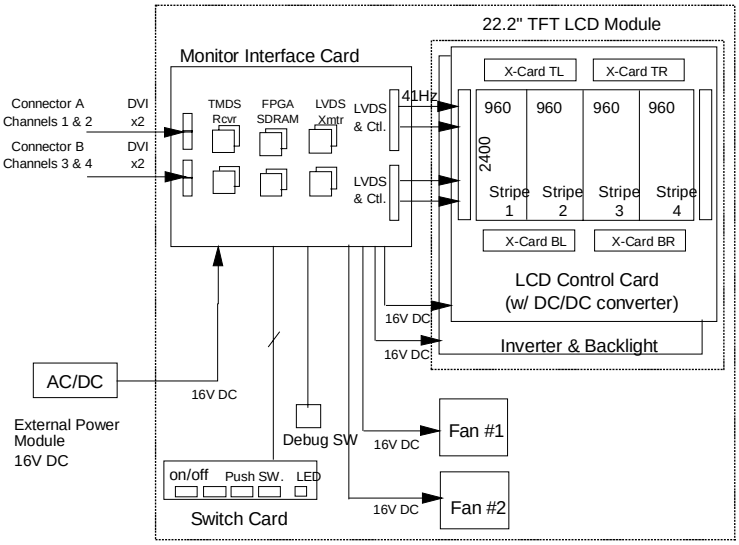

Figure 3. Block diagram of drive electronics.

Data refresh rate refers to the process of writing and reading data from a frame buffer, and can be very different from the screen refresh rate. Data refresh rate requirements are primarily determined by needs for moving-image quality. As discussed earlier, for all but the very most demanding applications there is no need to drive TFTLCDs with a data refresh higher than about 30 Hz.

The monitor interface card contains FPGA circuits, frame buffer memory, input TMDS receivers, LVDS transmitters, with connections to power supply and control circuits. This interface provides a number of important functions, and is reprogrammable. First, the receiver and transmitter sections convert the TMDS signals provided by the graphics card DVI output to LVDS signals used within the monitor. Second, the column pixel data is routed to the top and bottom of the array, as is appropriate. Third, the data is buffered in memory such that the output screen refresh is performed at 41 Hz, regardless of the input data rate. Fourth, the input data is composited from the various input DVI sources in the internal frame buffer, "remorphing" the data as needed for the four stripe partitions. Fifth, for cases in which the input DVI data does not match the full-screen resolution, the input data is pixel replicated to fill as much of the screen as possible, with the remaining portion made black. Sixth, limited timing differences between the different DVI inputs can be accommodated (synchronized) through the frame buffer. Currently, the maximum allowable phase shift depends upon operation mode. For the most stringent case, the maximum allowable delay between different DVI inputs is on the order of one line time or less, a small fraction of a frame time.

There are a maximum of four DVI input channels to the monitor, and four modes of data input, as shown in Figure 4. As far as the operating system is concerned there can be either a single desktop for the entire screen or logical “multiple monitor” screens, depending upon the graphics card device driver. If a single DVI input channel is used, the input digital pixel data format can be as large as the full 3840x2400 resolution of the screen. If input data is in a smaller format, such as XGA 1024x768, then the pixel data is replicated a factor of 3 in both vertical and horizontal directions, yielding a 3072x2304 image centered on the screen. For VGA 640x480 input, the pixel data is replicated by a factor of 5, yielding a 3200x2400 image. For the current product, if two DVI inputs are used, then pixel replication is not enabled, and panel mode is set to be two vertical stripes, which must be driven at 1920x2400 to fill the screen. If four DVI inputs are used, then depending upon the input data format, either a four-stripe mode or quad tile mode is selected.

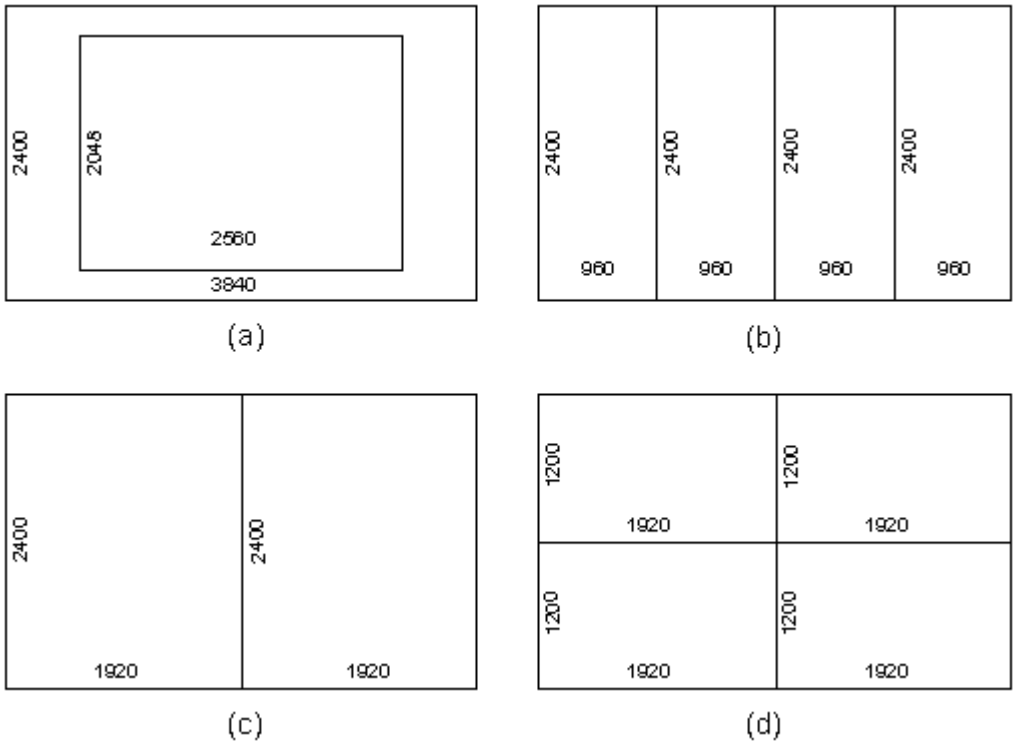

Figure 4. Input modes for Model T221. (a) singlel-screen with 1DVI input, example shown of 1280x1024 input with 2x replication. (b) 4-stripe with 4 DVI inputs. (c) 2-stripe with 2 DVI inputs. (d) quad tile with 4 DVI inputs

The reprogramable nature of the monitor interface allows for future input of dual-DVI signals and other new protocols, such as Digital Packetized Video Link (DVPL),[19] discussed later. In additional to the monitor interface, other portions of the monitor electronics are also reprogramable, such as EDID and controller section.

6. SYSTEM ARCHITECTURE

For conventional systems, the digital data generated from the software application and graphics card are converted to analog for input to the display. For analog flat panel displays, the input analog data must be converted back to digital format for input into the column drivers, where it is then converted back to analog signals to apply to the pixels. For digital high-resolution monitors, the retention of digital format for pixel data from the computer, through the graphics card, and into the display has a profound effect on the system architecture.

Just as the range of applications for high-resolution displays is broad, so too are the graphics performance requirements. There is a large difference in requirements for static 2D images viewed with a personal computer, 3D textured images on a standalone workstation, to massive polygon generation on high-end visualization systems with distributed rendering. The architecture of these systems can be described in terms of N computers connected to M displays, embedded in a distributed network for both computer-to-computer data exchange, and computer-to-display exchange.

For a standalone computer connected to one 9.2 Mpixel display, single graphics card solutions currently exist for the different bus and operating system platforms. The overall graphics performance is affected by the combination of many hardware and software components of the computer system. For computer-generated graphics, the intrinsic liquid crystal limitations discussed earlier are minor compared to limitations of the graphics card. The graphics card VRAM capacity, graphics acceleration performance, addressing capability, and number of DVI output all affect the performance. Currently, the best performing card for the 9.2 Mpixel display has 128 MB of memory and has two, single-channel DVI outputs providing a full-screen data refresh rate of 25 Hz. Multiple graphics cards can be installed in the same computer, and with appropriate software, the graphic acceleration required for rendering can be distributed

between the cards. However, generally it will be required that data output of the cards be synchronized, i.e. "genlocked" together. For cards allowing this feature, hardware genlocking can be achieved by connecting together the reset counter pins on the cards. In principle, software genlocking is also possible.

Some applications call for multiple monitors to be driven by a single computer. A straightforward approach is to utilize two or more graphics cards installed in the same computer, with each set of outputs connected to a different display. Using this approach, we have driven two 9.2 Mpixel displays from the same computer with a PCI bus card and AGP bus card, providing a total of 18.4 Mpixels. An alternative approach utilizes new digital data protocol, such as DVPL,[19] now proposed as a standard. A Bertha prototype monitor driven with DPVL protocol, was recently demonstrated.[20] With DPVL, packets of video data are used that contain only update information for the display. The unchanged pixel data remains stored in the display frame buffer. In this way, the bandwidth requirements are greatly reduced for situations in which only a small number of pixels are changing with time, or a large number of pixels are changing very slowly. Furthermore, each packet of video data can be tagged, so that only the monitor(s) with matching identification tags will respond to the data. With DPVL, many displays can be "daisy-chained" together on the same digital data bus, all driven from a single computer. One important application is a command and control center with a wall of maps.

For high-performance applications with demanding rendering requirements, multiple inter-connected computers can be used to distribute the computation and hardware acceleration. This approach requires the application to distribute the computation and rendering. If the graphics card outputs of each computer cannot be genlocked, then the output graphics data will need to be composited and synchronized with custom converter hardware. One low-cost approach currently being developed at IBM is the Visualization Frame Buffer (VFB).[21] The modular VFB hardware accepts multiple DVI inputs, buffers the data, and distributes the data among multiple outputs. The DVI outputs can be used to drive one or more displays, including projection displays. VFB circuitry could be designed or programmed to perform a variety of tasks, such as crosspoint switching, multiplexing, compression, and data protocol conversion.

For even more demanding visualization needs require many rendering nodes, such as a cluster of computers. Typically, custom, high-end visualization systems are used for these applications. However, cable length restrictions limit the flexibility of these systems. Maximum DVI cable lengths are approximately 5 meters, and this requires that the user visualization workstation or projection system be placed in close proximity to the rendering system. One low-cost solution utilizes Scaleable Graphics Engine (SGE) hardware,[22] being developed at IBM.[23] The SGE is designed to accept multiple data inputs via gigabit ethernet connections, and provides DVI output. In this way, the rendering nodes can be located remotely using a standard backbone for routing data. The memory buffer of the SGE is continually remapped to avoid loading effects, and the graphics cards in the rendering nodes perform the graphics acceleration. SGE technology provides a modular, flexible, low-cost and high performance solution as compared to conventional standalone high-end visualization systems. For high-end visualization needs, the combination of custom compositing hardware and the 9.2 Mpixel display provides a low-cost "mini-cave" environment at the desktop, within a distributed computing computing architecture.

7. INFRASTRUCTURE ISSUES

Nearly all data output from computers is presented to the user through the display. For some users with critical applications, it can be the single most important hardware component of a system. However, displays are widely viewed as a commodity, as compared to other computer hardware. The recent growth of flat-panel display monitor business has occurred due to space and power savings, and secondarily due to image quality and information content. As flat-panel prices decline, further growth of flat panel monitor business will occur. However, there are a number of infrastructure issues that currently have a strong effect on the growth of high-resolution digital TFTLCDs.

7.1 Digital Video Interface

A critical factor for the growth of flat panel displays is the development of graphics cards with digital output. Analog drive is useful only to maintain compatibility with CRTs, and introduces additional D-to-A and A-to-D steps between the graphics card and display that tend to degrade signal and image quality. Although analog scaling is important to driving CRTs with a variety of input pixel formats, to achieve optimal image quality, TFTLCDs should be driven at

native resolution. The emergence of DVI as a standard for graphics cards is an important enabling factor, and there are challenges for graphics card performance to drive very high-resolution displays. Pixel clock rate, video memory, and chip design for graphics acceleration are important. Digital drive, unlike analog drive for CRTs, does not require blocks of time to be reserved for electron beam flyback, blanking and signal settling. Ultimately, digital graphics card output will gain market share because it is less expensive to produce and more reliable than analog output.

A simple consideration for driving high-resolution displays is the number of DVI inputs. Figure 5 illustrates the maximum data refresh rates, which can be achieved versus the number of display pixels for various numbers of DVI channels. This plot neglects latency, i.e. time not allocated for providing pixel data. The bandwidth of a single DVI channel is roughly the same as for an analog channel. For the IBM 9.2 Mpixel display, however, a single DVI channel limits the data refresh rate to a maximum of approximately 16 frames/sec. For good moving image quality, a minimum of two DVI inputs is required. Depending upon the graphics card performance, and card pixel addressing capability, more DVI outputs may be needed.

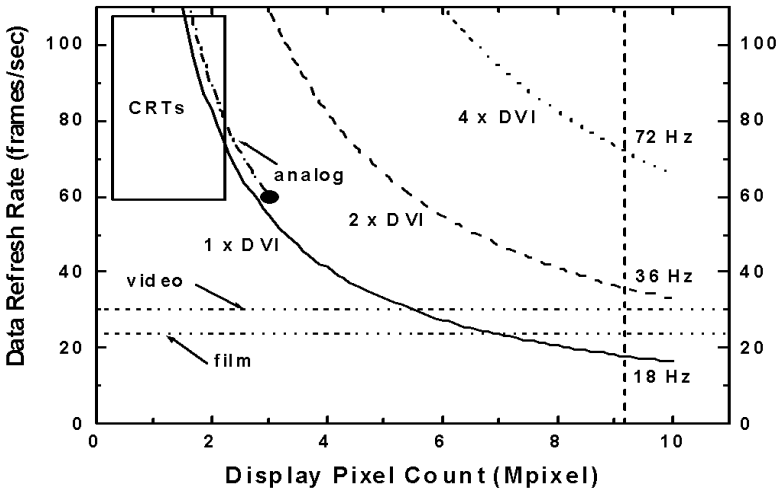

Figure 5. Maximum data refresh rates versus display pixel count for different numbers of DVI inputs.

7.2 Operating system and application user interface

In the early 1980's when display pixel densities were 72 ppi, attempts to provide what-you-see-is-what-you-get (WYSIWYG) capability were moderately successful. The unit of measure for scaling dialog boxes and other graphical user interface features was the pixel. When display pixel densities moved to 96 ppi, the first of a series of compensation fixes began. Font sizes, controlled through the operating system, were increased by a scaling factor to compensate for the increased screen pixel density. Today, pixel densities of 133 ppi are common for notebook displays, and the characteristics of the user interface have not been changed to accommodate this. There is no guarantee that letter sizes appearing on the screen are the same physical size they appear when printed. Although improvements have been made in the latest operating systems, not all features of the user interface are scaleable with pixel density. Examples include certain dialog boxes with text or radio buttons, and icon sizes. Controls do exist for setting many desktop features, but these must either be set manually, or by a separate software application.

For applications, there is generally even less control over the appearance of the user interface. Enlargement of some features can be built-in, such as aids for the visually impaired. However, until recently there has been little incentive to design applications to take full advantage of increased pixel density. In our experience, many of these small features do not present a large problem to experienced users. Finding these features for mouse selection is not particularly difficult for a user who has performed that same operation many thousands of times in the past. However, for broad acceptance of high-resolution displays in the marketplace, applications must be designed with this control.

Clearly for successful design of user interfaces, the pixel density must be considered as an independent parameter. This is important not only for high-resolution displays and operating systems, but for common applications such as web browsers. These sorts of problems routinely emerge with web-site design today. In the future, formating for the display user interface must include both the pixel density and the screen size, similar to what is done for printers.

7.3 Intelligent display environment

The IBM 9.2 megapixel display represents a step toward an architectural environment in which various processing elements and communication functions are distributed between the computer and display. The digital input pixel data is processed and buffered in the display, similar to what is done in a printer. Just as a printer is connected to a data network, an intelligent display can also be connected to a data network. In this way, some of the pixel processing function of the graphics adapter is absorbed into the display, and new possibilities for efficient and secure visual data distribution and processing occur. In the future, it may occur that the graphics hardware acceleration is split between the computer motherboard chipset and display. The development of intelligent digital interface electronics for displays should enable a wide variety of input data to be processed and rendered on the screen. The range of input data could span standard computer display progressive scan, other video data formats, and printer data file formats. In this way, new ways are allowed to combine standalone workstation, local-area-network, and broadcast communication functions.

8. SUMMARY

Key technology, system architecture and infrastructure issues for IBM 9.2 Mpixel display have been discussed. The need for increased information content and quality will eventually lead to broad deployment of high-resolution digital display technology, but the rate at which this process will take place is dependent upon infrastructure issues only indirectly related to the display itself. High-end digital still-cameras and digital video camera development, the need for improved video conferencing, improved remote sensing and image data sharing capabilities are some of the developments, which should spur deployment. These needs occur in addition to conventional space and power savings. The gating factors are primarily related to infrastructure and cost, not underlying display manufacturing issues.

ACKNOWLEDGEMENTS

The support of K. Schleupen and his group, B. Wisnieff, and B. Artemenko is acknowledged, along with thanks to D. Hopper for invitation of this paper, and to the “other Steve” Millman for review of the manuscript.

REFERENCES

1. R.L. Wisnieff, J.J. Ritsko, IBM J. Res. Dev. **44**, 409 (2000).
2. E.G. Colgan et al., IBM J. Res. Dev. **42**, 427 (1998).
3. K. Schleupen et al., Proc. 18th Int. Display Research Conference, Seoul, Korea, p. 187, (1998).
4. T. Ueki, Proc. 1st Int. Display Manufacturing Conference, p. 177 (2000).
5. F. Hayashiguchi, Society for Information Display Int. Symp. Digest, paper 29.1, p. 480 (2001).
6. S. Odahara, , Society for Information Display Int. Symp. Digest, paper 56.4, p. 1358 (2001).
7. H. Kamiya, Society for Information Display Int. Symp. Digest, paper 56.3, p. 1354 (2001).
8. Y. Nakagawa, Society for Information Display Int. Symp. Digest, paper 56.1, p. 1346 (2001).
9. S.L. Wright, et al. IBM J. Res. Dev. **42**, 445 (1998).
10. S.L. Wright, I.L. Bailey, T-M Kuan, R.T. Wacker, J. Soc. Inf. Display **7**, 253 (1999), SID ’99 Digest, paper 24.3, p. 346 (1999).
11. T. Nose et al, Society for Information Display Int. Symp. Digest, paper 15.3, p. 180 (1999).
12. Request for Proposal RFP 01-71, United States Display Consortium, Dec. 7, 2001.
13. T. Kurita, Society for Information Display Int. Symp. Digest, paper 35.1, p. 986 (2001).
14. J. Hirata, et al., Society for Information Display Int. Symp. Digest, paper 35.2, p. 990 (2001)
15. T. Nose et al., Society for Information Display Int. Symp. Digest, paper 35.3, p. 994 (2001).
16. B-W Lee, D. Sakong, K. Chung, Information Display, **17**(12), p. 26 (2001).
17. K. Sekiya, H. Nakamura, Society for Information Display Int. Symp. Digest, paper 29.3L, (2002)
18. J. Larimer, NASA Ames Research Center, *personal communication*.
19. J. Mamiya et al., Society for Information Display Int. Symp. Digest, paper 6.1, p. 38 (2000).
20. J. Hanari et al., Society for Information Display Int. Symp. Digest, paper 15.2, p. 210 (2001).
21. K. Schleupen, IBM T.J. Watson Research Center, *personal communication*.
22. K. Perrine, D. Jones, IEEE Supercomputing, Nov. 2001.
23. T. Jackman, IBM T.J. Watson Research Center, *personal communication*; demonstration at SuperComputing 2001, Denver, CO.
* wrightsl@us.ibm.com; phone 914 945-3438; IBM T.J. Watson Research Center, PO Box 218, Yorktown Heights, NY 10598.