Stop
Rewind
Fast Forward (by playing only I frames)
Fast Reverse (by playing on I frames)
Random Seek

The player optionally allows the user to control the decoding of various frame types to provide a scalable performance knob by skipping the decoding of B frames.

Following is a listing of the API supported by the decoder:

1) MpgDecInfo
2) MpgDecStart
3) MpgDecEnd
4) MpgDecFrame
5) MpgSeekFrame
6) MpgRewind

This API allows for random seeking, so that the player can implement fast forward, reverse play type of operation. For bitstreams containing only I Frames, this is a relatively simple operation. For bitstreams containing I, P and B frames the MpgSeekFrame operation becomes a little complicated since we cannot decode P or B frames without decoding the reference frames on which they are dependent. We overcame this problem by implementing the MpgSeekFrame function to always find the nearest I frame in the direction of the seek.

7.0 REFERENCES

[1] Patel, Ketan, et. al., "Performance of a Software MPEG Video Decoder", Proceedings of the ACM Multimedia conference, 1993.

[2] Ulichney, R., *Digital Halftoning*, MIT Press, Cammbridge, Mass. 1987.

[3] "Coded Representation of Picture, Audio and Multimedia/Hypermedia Information", Committee Draft of Standard ISO/IEC 11172, December 6 1991.

[4] "Microsoft Windows 3.1 SDK: Programmers Ref. Vol 1. - Overview", Microsoft Press, 1993.

Highly Integrated Controller Eases MPEG-1 Adoption

MPEG-1 Decoder Lowers The Cost Of Adding Video And Audio To PCs.

DAVE BURSKY

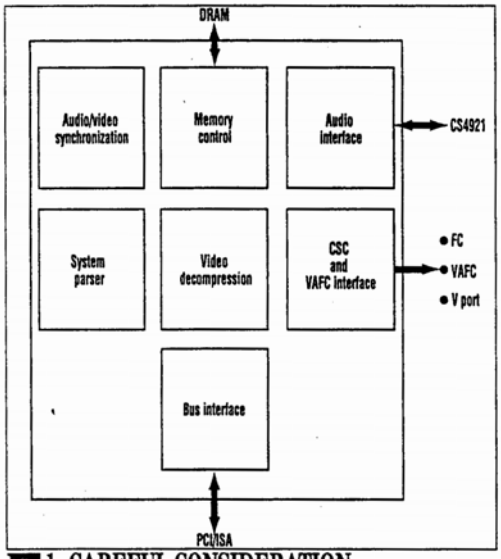
The drive to add multimedia capabilities to the personal computer, either by offering add-in cards or by building the capability directly on the motherboard, is forcing card and chip suppliers to find new ways to reduce costs. Individual add-in MPEG-1 decoder cards, although reasonable at several hundred dollars, must have their costs cut in half. The aim is to trim the user's cost of adding in MPEG-1 decoding to $100 for an off-the-shelf card, and even less if the system manufacturer is to include the capability as part of the base feature set of the PC.

With that in mind, designers at Cirrus Logic studied system partitioning issues and came up with a three-chip solution that trims the cost of a full MPEG-1 subsystem to less than $50 in components *(Fig. 1)*. The three chips include the newly-designed CL-GD5520 MPEG-1 video decoder chip, the already available CS4921 audio decoder, and a commodity, 256-kword by 16-bit DRAM. The DRAM buffer can be expanded by adding a second 256k by 16 DRAM. The larger buffer improves the quality of the displayed video and allows the subsystem to handle larger audio/video streams.

With the three chips, designers can build systems that decode full-motion MPEG video from a variety of video sources. That includes CDs, MPEG-1 CD-i movies, DOS OM-1 (Open MPEG consortium) compatible titles, and Microsoft Windows MPEG MCI standard video.

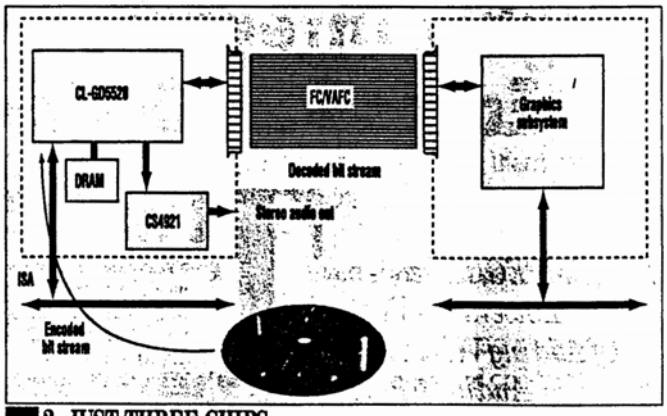
Designed from the ground up to offer the simplest interface in the PC environment, the CD-GL5520's MPEG-1 video decompressor is based on the MPEG-1 core technology licensed from CompCore Inc. The core is surrounded with all the functions it needs to communicate with the rest of the system at data rates of 80 Mpixels/s off the video port, and at up to 132 Mbytes/s on the host-bus interface. Unlike several other highly integrated MPEG-1 chips that incorporate the audio playback channel on the chip, designers at Cirrus Logic decided to keep the function off the chip. That's because the economics of integrating the sound onto the video decoder chip shows that the all-in-one approach doesn't really lower the cost of materials.

The decompression chip also includes both PCI and ISA host-bus interfaces (including PCI bus master-



1. CAREFUL CONSIDERATION to system partitioning has resulted in a highly-integrated MPEG-1 decoder developed by Cirrus Logic, the CL-GD5520. To simplify system design, the chip includes video decompression logic, a host-system interface to ISA PCI buses, DRAM control logic, system parsing control and audio/video synchronization logic. A multi-featured output port provides a VESA advanced feature connector (VAFC), a standard feature connector interface, or a proprietary V-port enhanced interface for transferring video images.

HIGHLY-INTEGRATED MPEG DECODER



2. JUST THREE CHIPS are needed to implement a full MPEG audio and video solution. These include Cirrus Logic's CL-GD5520 video decoder and CS4921 audio decoder, and a 512-kbyte DRAM (one 256-kword by 16-bit DRAM).

ing), a VESA advanced feature connector (VAFC) video output (which can also implement an FC or V-port interface), and an integrated color-space converter *(Fig. 2)*. Furthermore, due to its built-in arbitrary scaling and zoom capabilities, the chip can deliver windows of almost any size without degrading image quality. Pixel interpolation is used in the X direction and pixel replication in the Y direction when the images are resized.

To ensure that the MPEG-1 video and audio channels stay synchronized, a key issue in multimedia playback, designers at Cirrus Logic incorporated dedicated synchronization logic on the chip. The chip also includes a system parser to structure the MPEG-1 system-layer bitstream. For the audio channel, the chip delivers the parsed audio stream to the companion CS4921 audio codec developed by Crystal Semiconductor, Austin, Texas, a subsidiary of Cirrus Logic.

Support for both Windows 95 and the Plug-and-Play initiative is also included in the decoder chip. In addition, the PCI-bus mastering capability allows video data transfers to take place directly over the PCI bus, and directly to the graphics-controller's frame buffer. That eliminates the need for a ribbon cable that goes over the top of the cards. This solution also allows the graphic controller to display graphics at higher resolutions.

The integrated color-space conversion circuitry supports 5:5:5, 5:6:5, 8-bit error-diffused 3:3:2, and true-color 8:8:8 RGB formats, as well as 16-bit 4:2:2 YUV and AccuPak 8-bit YUV formats. That broad format support allows the video data to be delivered to just about any display controller subsystem. The chip also supports both NTSC and PAL video resolutions, thus expanding the potential market beyond the U.S. border.

Cirrus Logic's designers have also been busy developing the extensive driver support the chips will need for integrating them into a PC. Drivers are available for Windows 95, Windows 3.11, OM-1 DOS, VideoCD, and CD-i. The CL-GD-5520 is also fully compatible with the company's previously available graphics controllers and audio components. Reference design kits are available so that designers can quickly check compatibility MPEG-based software titles. Software development kits are also available for users who wish to incorporate MPEG video into their applications. □

PRICE AND AVAILABILITY

The CL-GD5520 is housed in a 208-lead PQFP. It sells for $32.00 apiece in lots of 1000 units. Samples are now available.

Cirrus Logic Inc., 3100 West Warren Ave., Fremont, CA 94538-6423; Saul Altabet, (510) 252-6286. CIRCLE 500

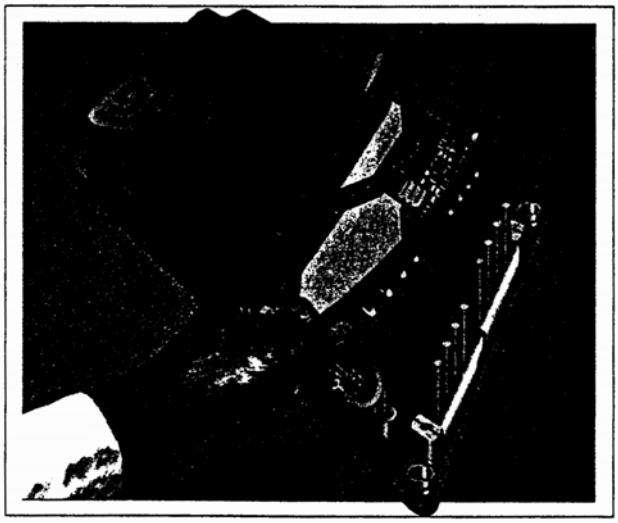
HOW VALUABLE?	CIRCLE
HIGHLY	525
MODERATELY	526
SLIGHTLY	527

ELECTRONIC DESIGN

AE

August 21, 1995 Volume 43, Number 17

COVER STORY



Modular Dc-Dc Converter Sends Power Density Soaring p. 59

A 500-W module with a 90 W/in.^{3} power density shrinks front-end supplies.

FEATURES

ELECTRONIC DESIGN (USPS 172-080; ISSN 0013-4872) is published twice monthly except for 3 issues in May and 3 issues in October by Penton Publishing Inc., 1100 Superior Ave., Cleveland, OH 44114-2543. Paid rates for a one year subscription are as follows: $105 U.S., $185 Canada, $210, $255 International. Second-class postage paid at Cleveland, OH, and additional mailing offices. Editorial and advertising addresses: ELECTRONIC DESIGN, 611 Route #46 West, Hasbrouck Heights, NJ 07604. Telephone (201) 393-6060. Facsimile (201) 393-0204. Printed in U.S.A. Title registered in U.S. Patent Office.

ELECTRONIC DESIGN

August 21, 1995 Volume 43, Number 17

DEPARTMENTS

Jesse H. Neal Editorial Achievement

Awards: 1967 First Place Award; 1968 First Place Award; 1972 Certificate of Merit; 1975 Two Certificates of Merit; 1976 Certificate of Merit; 1978 Certificate of Merit; 1980 Certificate of Merit; 1986 First Place Award; 1989 Certificate of Merit; 1992 Certificate of Merit

COVER ART: MARGARET ENDRES, BRUCE JABLONSKI, JOHN BEUKAMANN

AS

FA 17.1: An MPEG-1 Audio/Video Decoder with Run-Length Compressed Antialiased Video Overlays

Dave Galbi, Everett Bird, Subroto Bose, Eric Chai, Yen-Ning Chang, Pierre Dermy, Nishendra Fernando, Jean-Georges Fritsch, Eric Hamilton, Barry Hu, Ernest Hua, Frank Liao, Ming Lin, Ming Ma, Edward Paluch, Steve Purcell, Hisao Yanagi, Sun Yang, *Miranda Chow, *Takeya Fujii, *Akio Fujiwara, *Hiroyuki Goto, *Keiji Ihara, *Shinichi Isozaki, *Janny Jao, *Isami Kaneda, *Masahiro Koyama, *Tomoo Mineo, *Izumi Miyashita, *Goichiro Ono, *Shinji Otake, *Akihiro Sato, *Hideo Sato, *Akira Sugiyama, *Katsunori Tagami, *Kenji Tsuge, *Tomoyuki Udagawa, *Koji Yamasaki, *Sadahiro Yasura, *Tsuyoshi Yoshimura

C-Cube Microsystems, Milpitas, CA
*JVC, Yamato, Japan

This chip decodes MPEG-1 audio and video in real-time when connected to a single 80ns 256kx16 DRAM. The features of the chip are summarized in Table 1. A block diagram is shown in Figure 1. An MPEG-1 system stream, optionally embedded in a CD data stream, is sent to the chip on either an 8b host bus or a serial bus. The host interface contains a code FIFO that buffers input bit streams before they are written to the audio, video or overlay bit-stream buffers in DRAM. The MPEG system stream is processed by interrupting the on-chip CPU after a packet of compressed data has been written to DRAM. The CPU reads the system stream headers out of the code FIFO and initiates a block transfer of the next packet of compressed data to DRAM. The chip uses less than 5% of the clock cycles for system stream processing. The chip alternates between audio decoding and video decoding, with the audio portion using 15% of the clock cycles and video using 80%.

Audio and video bit streams are read from DRAM into a decoder FIFO. When decoding video, variable length codes (VLCs) are converted to fixed length codes (FLCs) by the VLC/FLC decoder. The VLC/FLC decoder writes video AC coefficients into ZMEM. When decoding audio, the VLC/FLC decoder extracts subband samples from the bit stream, performs degrouping and writes the results into ZMEM.

The signal processing unit (SPU) receives commands from the CPU and executes these commands in parallel with the rest of the chip. The SPU datapath is shown in Figure 2. The SPU performs three commands:

1. Dequantization and IDCT for video decoding
2. Dequantization and descaling for audio decoding
3. Matrixing and windowing for audio decoding

The Dequant/IDCT command reads an 8x8 block of AC coefficients from ZMEM and writes the results to a double buffered PMEM. During video decoding, the TMEM is used as the quarter-turn memory for the IDCT and the QMEM contains the quantizer matrix. The data flow for the SPU audio commands is shown in Figure 3. The Dequant/Descale command reads a vector of 32 audio subbands from ZMEM and writes the results to 32 locations in TMEM. The other 32 locations in TMEM are used to accumulate 32 partially-decoded audio samples. The Matrix/Window command reads 33 20b matrix results from PMEM and adds the product of matrix results and window coefficients to the partially decoded audio samples in TMEM. The Matrix/Window command then computes 4 20b matrix results that are written to PMEM. The DRAM controller writes matrix results in PMEM to DRAM and fetches previous matrix results for windowing. These DRAM transfers are in parallel with SPU operation. After 8 Matrix/Window commands, TMEM contains 32 decoded audio samples that are written to an audio output buffer in DRAM. The audio output unit receives decoded audio data from DRAM in an 8B FIFO and sends them out to the pins of the chip.

During video decoding, the motion- compensation unit receives reference blocks fetched from DRAM and half-pixel offsets them if needed. The offset reference blocks are added to the IDCT result in PMEM and the sum is stored back into PMEM. The motion compensation unit and SPU work in parallel on opposite halves of PMEM. After the offset reference blocks have been added to PMEM, the resulting decoded pixels are written to DRAM.

The video output unit receives decoded pixels from DRAM in a 112B luminance FIFO and a 128B chrominance FIFO. Luminance and chrominance are horizontally and vertically interpolated by 2x in each direction using a 7-tap horizontal filter and a 3-tap vertical filter. Compressed video overlays are read from DRAM into an overlay FIFO, decompressed and then blended with interpolated MPEG video. Finally, the pixels are optionally converted to RGB and output.

To decode both audio and video with only one 80ns 256kx16 DRAM, the chip must minimize the use of DRAM bandwidth and DRAM space. This is accomplished with the following techniques:

1. Decoded B frames are compressed before being written to DRAM to save about 200kb of DRAM space.
2. Video overlays are compressed to reduce the size of the overlay bit stream buffer in DRAM and to reduce the DRAM bandwidth needed to fetch the overlay bit stream.
3. The on-chip CPU has 96 CPU registers and a 16b instruction word. This gives a 1.9x improvement in instruction density compared to a conventional RISC CPU with 32 registers and a 32b instruction word.
4. The 20b audio matrix results are packed into 1.25 16b words before being written to DRAM.

Decoded B frames are compressed with a lossy DPCM compression technique to save DRAM space. Scan lines are DPCM-decoded in the video output unit. Video overlays are compressed with a run-length code with 4 symbol lengths: 4, 8, 12 and 20 bits. The symbols with N bits cover all runs of at least N/4 pixels so the maximum bit-rate of the compressed overlay bitstream is 4b/pixel. The typical overlay bit rate is 0.6b/coded pixel. The overlay symbols select a shadow color, a text color or transparent (Figure 4). To reduce jaggies and flicker, the MPEG/shadow color boundary and the shadow/text boundary are antialiased using a 2b blend factor indicated by the overlay symbols. The overlay can be gradually faded on or off with a 5b global fade factor. The on-chip CPU has an instruction set designed for instruction density and ease of implementation. The 16b instruction word contains two 6b register addresses and a 4b opcode. There are a total of 96 CPU registers of which 64 are accessible at one time. When a CPU interrupt occurs, 32 interrupt registers are used in place of 32 regular registers. The CPU datapath is 24b wide. CPU instructions are stored in DRAM and are read into a 1024x16 instruction memory as needed. A micrograph is shown in Figure 5.

Acknowledgments

The authors thank the following employees of Matsushita Electronics Corporation for contributions to the chip: K. Hamaguchi, A. Haza, J. Huard, Y. Ochi, Y. Okada, M. Suzuki and A. Yamamoto.

0-7803-2495-1/95/$4.00/© 1995 IE

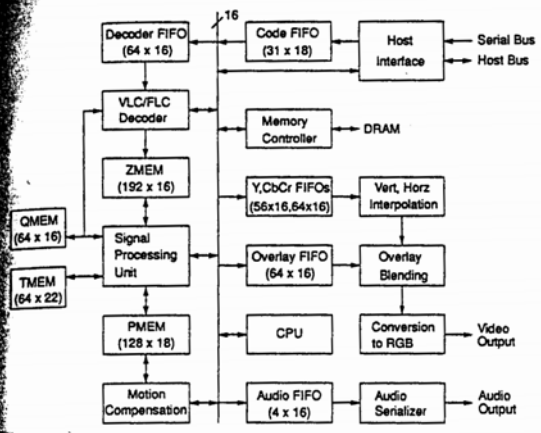


Figure 1: Block diagram.

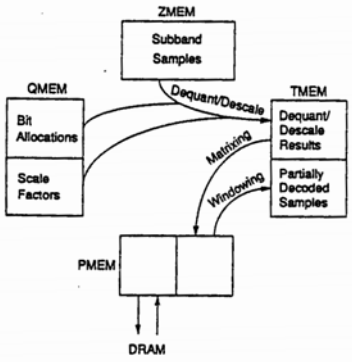


Figure 3: SPU audio commands.

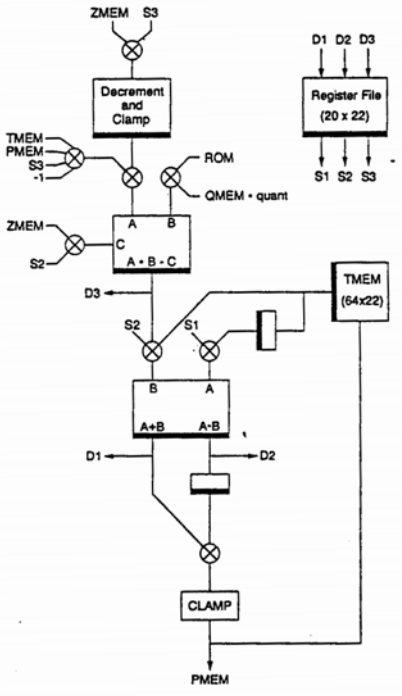


Figure 2: Datapath of signal processing unit.

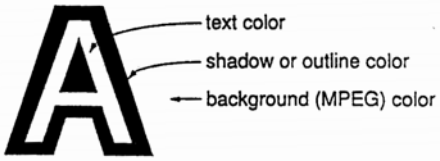


Figure 4: Video overlay colors.

Figure 5: See page 381.

Audio decoding performance	2 channels of 48kHz audio
Video decoding performance	352x240 @ 30Hz or 352x288 @ 25Hz
Video overlay resolution	up to 768x576
Process technology	0.5μm (drawn) 2-layer metal CMOS
Die size	11.5x11.5mm²
Logic transistors	305k
Memory transistors	485k
Clock frequency	40MHz
Operating voltage range	2.7V to 3.6V
Typical power consumption	600mW at 3.3V, Ta = 25°C
Max. power consumption	740mW at 3.6V, Ta = 70°C
Package	128-pin PQFP (18x18mm² body)

Table 1: Feature summary.

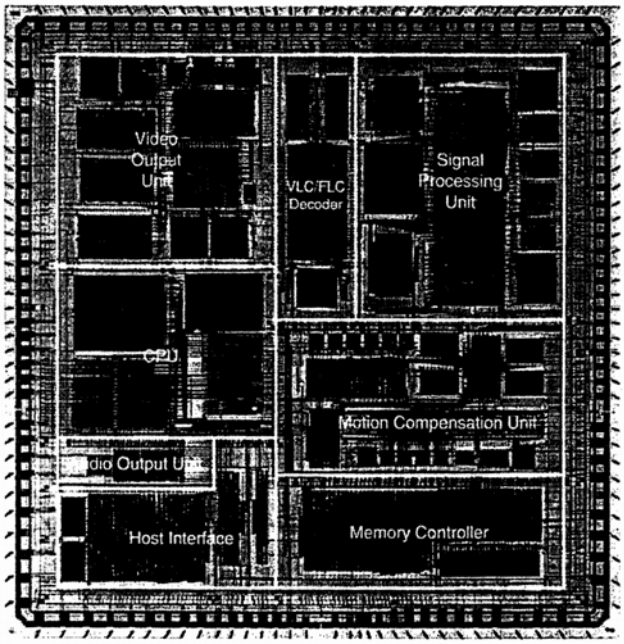


Figure 5: Micrograph of MPEG-1 A/V decoder.

FA 17.2: A Half-pel Precision MPEG2 Motion-Estimation Processor with Concurrent Three-Vector Search

(Continued from page 289)

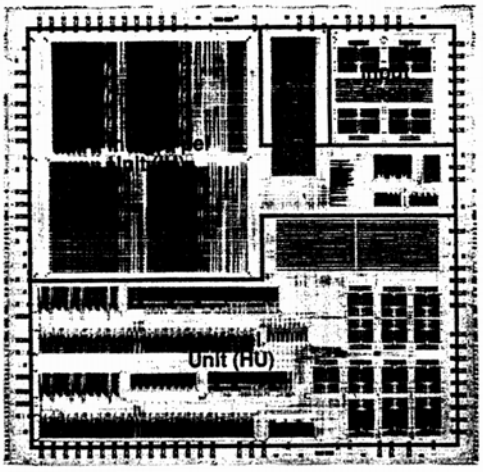


Figure 4: ME2 micrograph.

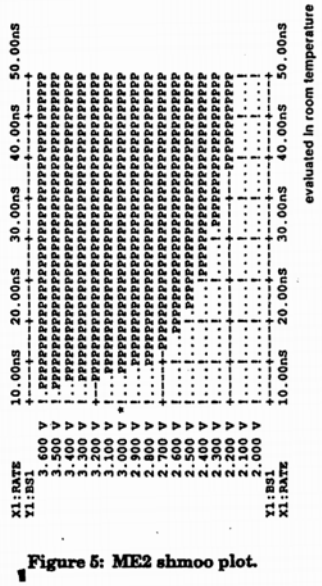


Figure 5: ME2 shmoo plot.

AY

SINGLE CHIP MPEG AUDIO DECODER

Greg Maturi
LSI Logic Corporation
Milpitas, California

ABSTRACT

An IC has been designed and fabricated which can take an MPEG System or MPEG Audio stream and decode LayerI and Layer II (MUSICAM) encoded audio into 16 bit PCM data. Audio/Video synchronization, cue and review is provided via its external channel buffer.

SUMMARY

The Single Chip MPEG Audio Decoder will take an MPEG Layer I or II (MUSICAM) System or Audio stream, and provide complete decoding into 16 bit serial PCM outputs. In addition, presentation can be delayed and audio frames skipped by means of the channel buffer, an external 256K x 4 DRAM controlled by this IC. This allows coarse synchronization of audio and video for skews up to 1 second for Layer I and 2.5 seconds for Layer II. Control over which frames are played or skipped provides cue /review features. Except for the channel buffer DRAM and a 25 - 30 Mhz clock, no other hardware is required. The IC is controlled by an 8 or 16 bit microprocessor, but can operate as a stand alone device with reduced flexibility.

The IC can receive data up to a 15 Mbits/second either serially or through microprocessor interface (selectable for 8 or 16 bits). An input fifo allows the IC to handle burst rates of up to 7.5 Megabytes/sec for up to 128 bytes. The IC will strip out the audio streams from MPEG system streams and provide presentation time and parametric information to the host. The audio frames will then be stored in the channel buffers.

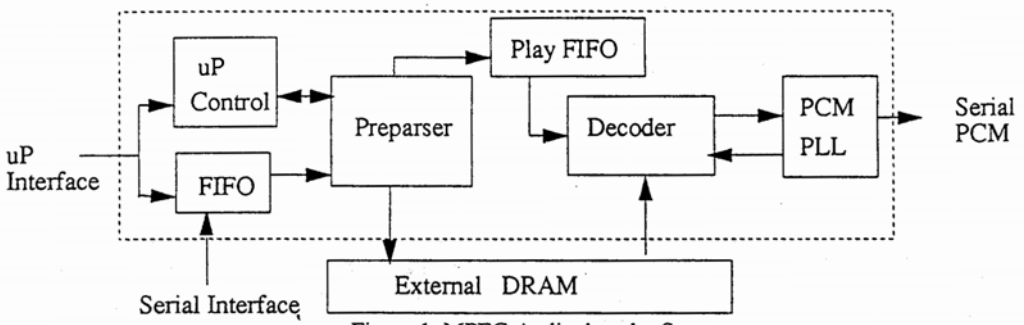


Figure 1. MPEG Audio decoder System

The IC is divided into 4 major parts: the preparser, the decoder, the DRAM controller, and PCM interface.

The preparser performs several functions: system/audio stream synchronization, stripping off of parametric and presentation time headers, syntax checking, CRC checking, and cataloging

Manuscript received June 5, 1992

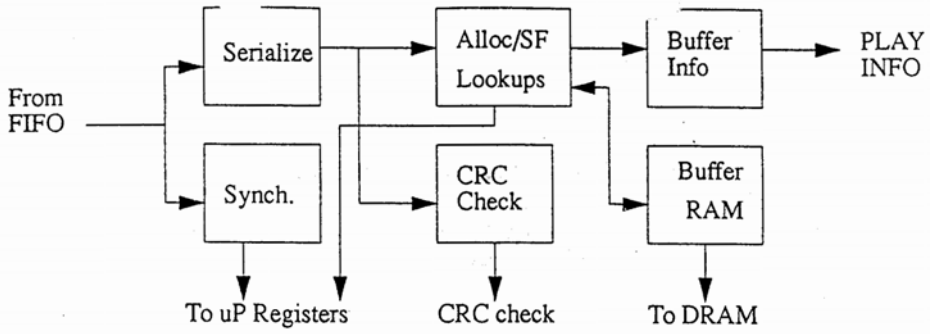


Figure 2.Preparser Architecture

frames. Error concealment (by repeating the last good frame) can be provided automatically. Since the frame must be partially expanded to obtain this information, the frame is stored in the channel buffer in a partially expanded form. A playlist is generated to tell the decoder which frame to decode next. The microprocessor can control which direction to fill the playlist, skip frames in the playlist and which direction for the decode to read the playlist.The decoder does most of the algorithmic work: It performs inverse quantization, scaling, and subband synthesis. It uses a 24 bit architecture. In addition, on Layer II it performs degrouping prior to dequantization. Filter coefficitents,dequantization values, scratchpad and vector memories are internal.

The DRAM controller provides RAS,CAS, address and data to the DRAM. It arbitrates between the preparser and the decoder. It also provides hidden refreshing. This controller requires a 256K x 4 DRAM (100 ns or faster)

The PCM interface buffers the PCM output from the decoder and provides 3 and 4 wire serial output compatible to most serial DACs. The serial clock is generated from the system clock using a fixed point divisor provided by the host (4 bits integer, 16 fractional). The decoder can also be bypassed, allowing serial PCM to be passed directly from input to output.

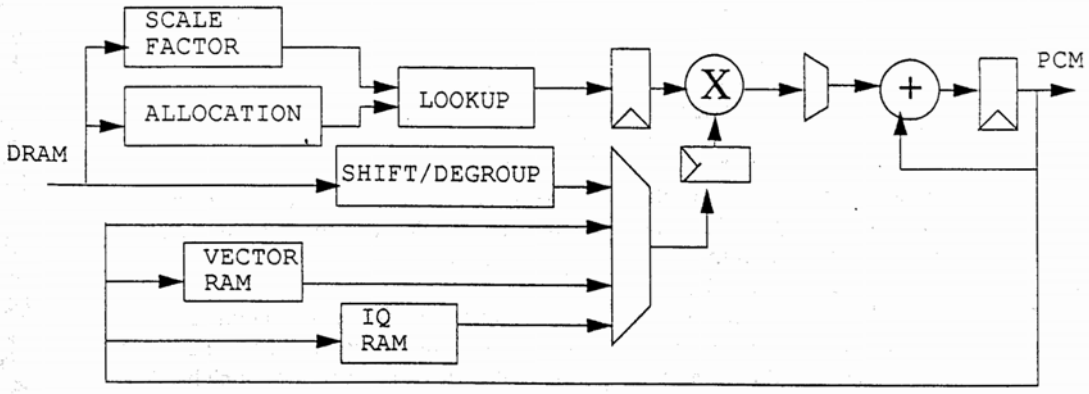


Figure 3.Decoder Architecture

Functional Block Operation

When initialized, the decoder synchronizes itself by monitoring the data stream and locating an audio frame in the data stream.When MPEG data is input, the chip strips away all unneeded information, retaining only the audio and control data. This data is then partially decompressed and stored in a channel buffer. When the appropriate control signals are seen, this stored data is played (fully decompressed and output in PCM format).

These activities are accomplished in general as follows:

Input Synchronization and Buffering

Data in either serial or parallel form enters the MPEG Audio decoder through the Controller Interface. The data is first synchronized to the system clock (SYSCLK), then is sent to the Input Data FIFO. The FIFO buffers data and supplies it to the Preparser. The FIFO can accommodate burst rates up to (input clock)/4 bytes/sec, for bursts of 128 bytes.

System Preparser

The Preparser performs stream parsing. For ISO System Stream parsing and synchronization, it detects the packet start code or system header start code and uses these to synchronize with packets. The parser reads the 16 bit "number of bytes" code in either one of these headers and counts down the bytes following. When count 0 is reached the next set of bytes should be a sync word, If not, the sync word seen was either emulated by audio or private data or a system error. The preparser will not consider itself synchronized until 3 consecutive good syncs have occurred. Likewise, it will not consider itself unsynchronized until 3 false are detected. This hysteresis is detailed in the flowchart in Figure 4. Upon synchronization, the preparser returns the presentation time stamp for use in audio-video synchronization.

Audio Synchronization

If the synchronization code is the selected audio stream or the input stream is only audio, the preparser will then synchronize to the audio stream. It first detects the 12 bit audio sync, if the bitrate is not free format, the bytes remaining in the frame are calculated from the bitrate and sampling frequency (extracted from the parametric values in the bitstream) according to the formula:

bytes = 48 * bitrate/sampling_frequency (I)

bytes = 144 * bitrate/sampling_frequency (II)

This value is loaded into a byte counter. As with the system synchronization, when the counter down counts to zero the preparser verifies the next 12 bits are a sync code. if the padding bit is set the counter will wait 4 bytes on Layer I and 1 byte on layer II before checking for the sync code. The hysteresis is similar to that of the MPEG system synchronization. The audio synchronization is identical for free format except one extra frame is required where the bytes in the frame are counted rather than calculated. Figure 5 shows the audio synchronization.

Storing in Channel Buffer

After synchronization, allocations and scalefactors are separated out and stored in the channel buffer. In layer I there are 32 bit allocations, each allocation 4 bits representing 0 to 15 bits per sample, 1 not allowed. In LayerII there are 8 to 30 allocations 1 to 4 bits in length, representing 0 to 16 bits per sample,1 not allowed. In Layer II, information on whether the samples are grouped (three samples combined into a single sample) is also stored with the allocation.

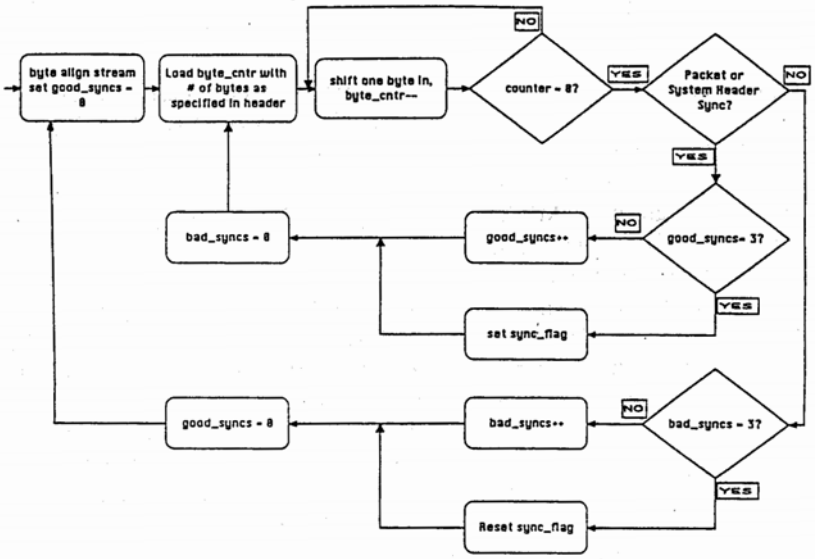


Figure 4. System Sychronization Hysterisis

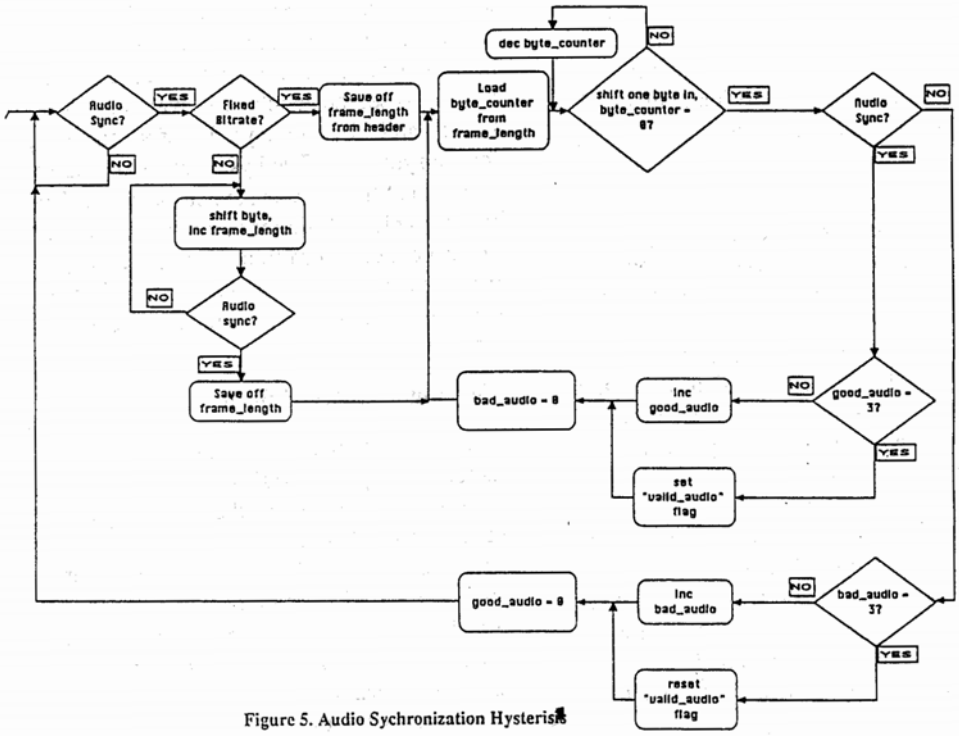


Figure 5. Audio Sychronization Hysterisis

Scalefactors are 6 bits indi to a lookup table, indicating the maximum amplitude of the samples in a subband. In Layer I, there is one scalefactor for each non-zero bit allocation. In Layer II there is 1 to 3 scalefactors per non-zero bit allocation. The actual number is determined by a 2 bit scalefactor select (again 1 per non-zero allocation). The preparser uses this information to separate out the scalefactors. The format that the allocations and scalefactors are stored in the memory is shown in Table I.

Table 1: channel buffer format

Row Address (HEX)	Information Stored
000:03F	allocations (channel 1)
040:07F	first scale index (channel 1)
080:0BF	second scale index (channel 1)
0C0:0FF	third scale index (channel 1)
100:13F	allocations (channel 2)
140:17F	first scale index (channel 2)
180:1BF	second scale index (channel 2)
1C0:1FF	third scale index (channel 2)

As allocations are being written to the channel buffer a small RAM records whether the allocation was non-zero. It then uses this information to separate out scalefactors without having to reread the channel buffer. On Layer II this is also done for scalefactor select bits. The same RAM holds these values for later use in scalefactor decoding.

Samples are left bitpacked when put into the channel buffer, just as received in the bitstream. They are stored immediately after the allocations and scale indices.

Audio data ca. .e parsed at input clock/2 bits per second. The limitation is the DRAM timing.

Parametric Data

The 20 bits following the audio sync word are called parametric data bits. These bits are used by the decoder and presented to the host interface. A maskable interrupt which is asserted as soon as these bits are read from the bitstream lets an optional microprocessor know these bits are available. Table II shows the bit definition.

Table 2: Parametric Data Format

Bit	Data
19 (MSB)	ID
18:17	Layer
16	Protection
15:12	Bitrate
11:10	Sampling Frequency
9	Padding
8	Private
7:6	mode
5:4	mode extension
3	copyright
2	original
1:0	emphasis

Ancillary data

The data immediately following the last data bit until the next frame sync is considered

ancillary data. The last bit of data is calculated from the decoded allocations and scale indices. This data is stored in a 16 X 8 bit FIFO. An interrupt indicates valid data in the FIFO, when the FIFO is half full, and when it has overflowed. If the ancillary data is less than 8 bits or a sync word is detected the ancillary bits are left aligned and written to the FIFO.

Play Buffer

The play buffer is a FIFO indicating the location in the channel buffer of the next frame to be played as well as minimum information the decoder needs to decode the samples. Usually, the play FIFO contains consecutive 4K block addresses for layer II and 2K block addresses for Layer I. However, if errors occur, the next address will be the last good frame stored.

The information that is passed in the play buffer is mode and mode extension, and a bit indicating if the frame should be blanked or played. This bit is set if an error occurs and error concealment is not selected. Since bitrate and sampling frequency are not allowed to be changed without resetting the decoder, the frame sizes remain the same. A time equivalent to the frame in error can be silenced with this method.

Decoder Operation

The decoder receives data for full decompression from the channel buffer. The location of this information and other required parameters are provided by the play buffer. The decoder performs all of the following functions: degrouping,dequnatization, denormalization and subband synthesis. Except degrouping, all functions are performed by use of a 2-cycle 24-bit multiplier -accumulator. A ROM provides lookup tables for scalefactors, quantization values, DCT and window coefficients. Two separate RAMS are provided, one for the dequantized coefficients, and one for the vectors generated in the subband synthesis. All memory is 24 bits, a block diagram is shown in Figure 3.

The decoding process begins with a start command being generated from the microprocessor or external start input. At that point the decoder reads parameters and channel buffer address information from the play buffer, and requests data from the channel buffer. The DRAM controller arbitrates between the preparser requesting to write data to the channel buffer and the decoder trying to read data for decompression.

In the first read, the decoder obtains allocation and scalefactor information. In the second read, the decoder obtains 1 to 5 nibbles containing the subband sample. If degrouping is required, the decoder implements the degrouping process:

```
For (i =0;i<3;i++)
{
  Sample[i] = c%nlevels;
  c = (int) c/nlevels;
}
```

by using a serial divider.

Dequantization is then performed by the following equation:

IQ[i] = (Sample[i] + D) * C

where C and D are both in lookup tables indexed by bit allocation.

Next is denormalization:

IN[i] = IQ[i] * scalefactor [scalefactor index]

This process is repeated for 32 samples. Each 24 bit denormalized sample is stored in the

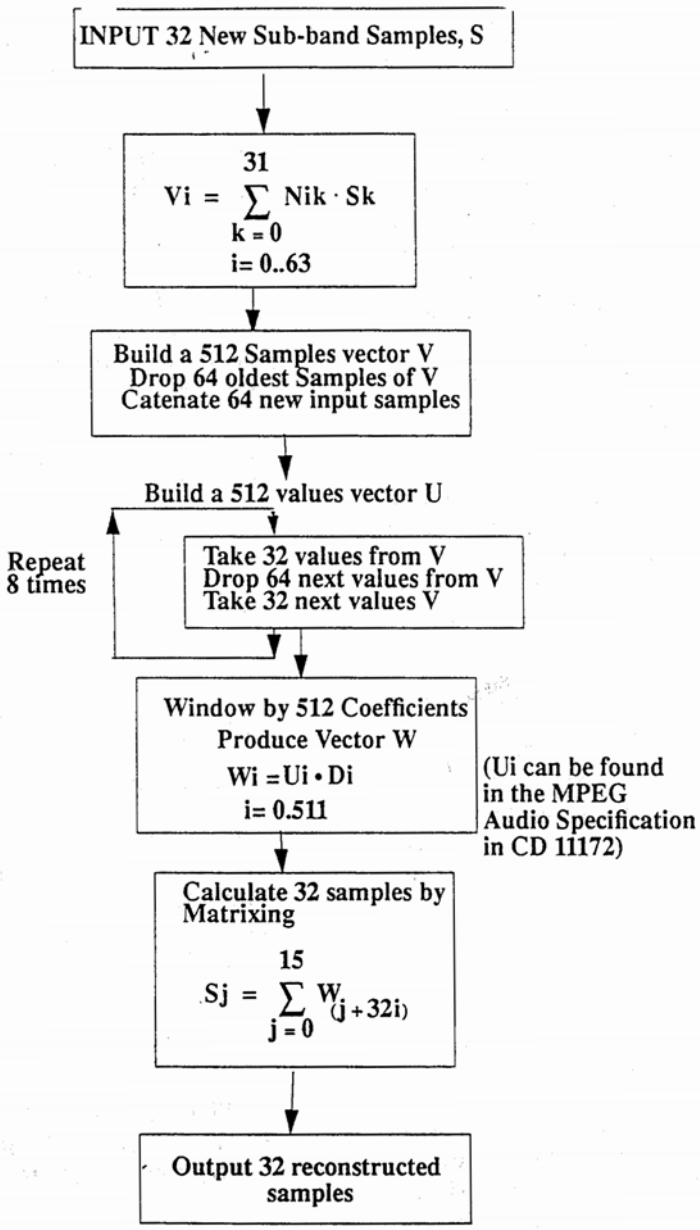


Figure 6. Subband Synthesis

denormalization RAM. These values are then used for subband synthesis.

Subband synthesis

The MPEG standard defines subband synthesis as shown in Figure 6. This process can be broken down into two functions. The first is an odd frequency inverse DCT, the second is a window function (with special addressing). The inverse DCT and window function are performed in parallel. That is the window and calculate samples is performed when the next PCM word has to be shifted out. In between the windowing function, the inverse DCT is performed. To insure that the data used for the window function is not from 2 different sets of sub-samples, the inverse DCT output is stored to a scratchpad portion of the vector ram. When the last of the 32 PCM samples has been transferred, this scratchpad is written to the correct section of the vector RAM.

PCM output

The PCM interface is responsible for obtaining data from the decoder, serializing it, and generating the control signals at the proper time for analog conversion by a serial DAC. In addition, refresh timing is based on the PCM clock.

The PCM contains registers that divide down the input clock to obtain the proper sampling frequency for the output PCM stream. These registers are either loaded on power up or written via the microprocessor port.

The first register is a 4 bit register, that indicates that is used to divide the input clock to obtain 2X the DAC serial clock. The 2nd register is 16 bits, and represents the fractional part of this divisor. Every time the 4 bit register counts down to zero, this fractional register is accumulated. Every time the accumulation exceeds 1 an extra clock cycle is added. Running with the slowest input clock and the fastest sampling frequency, this will produce a 10% variation in the serial clock, but the actual sampling frequency will be accurate to greater than 200 ppm.

As soon as the PCM word is loaded into the parallel to serial register, the PCM interface requests another from the decoder. The decoder completes its current DCT or inverse quantization, and then performs the window function described under subband synthesis. The decoder puts the PCM word into an output register, which the PCM interface will load into the parallel to serial converter when the previous word has been shifted out.

Conclusions:

The MPEG audio IC developed considers system level, interface and rate control issues, rather than just the number crunching involved in MPEG Audio compression. The IC can provide complete decoding from system to PCM with minimal additional hardware.

Acknowledgments:

The author would like to thank Peng Ang, Juergen Lutz and Simon Dolan. I would also like to thank Dave Auld and Neil Mammen for their incessant helpful suggestions during the development of this IC.

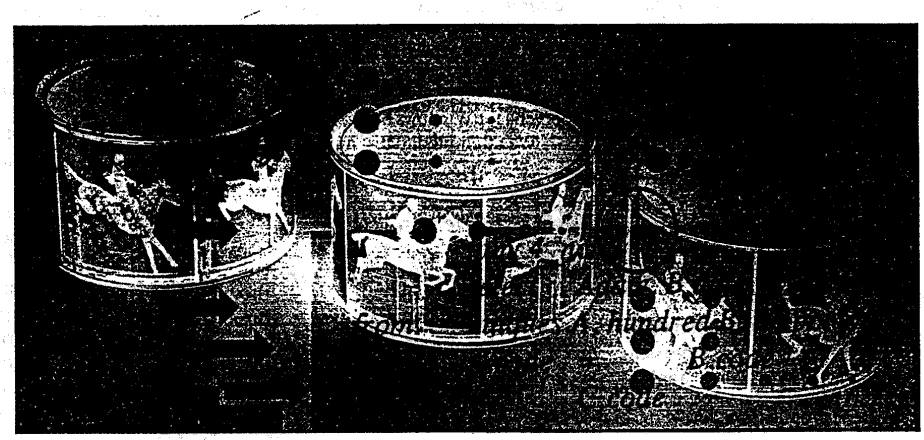
References:

1. ISO CD 11172-3: CODING OF MOVING PICTURES AND ASSOCIATED AUDIO AT UP TO ABOUT 1.5 MBITS/s.

Author Biography:

Greg Maturi received his B.S.E.E. degree in 1981 from the University of Virginia. From 1982 to 1985 he worked for Harris Aerospace division in Imploring, Florida implementing video compression algorithms. From 1985 to 1990, he worked for Texas Instruments, designing DSP circuitry for artificial intellignece systems. In 1991, he joined LSI Logic as an IC designer, where he is responsible for the development of the single chip MPEG Audio Decoder.

SQL DATABASES



The Great Leap *Forward*

The awkward years are over. PC-based SQL database servers have grown up to deliver on the promise of reliability and power.

By Brian Butler and Thomas Mace

Maturity has come quickly to PC-based structured query language (SQL) databases. In last year's roundup, we asked if 32-bit SQL databases for Intel processors were good enough to bet a business on. The answer, coming after months of ups and downs in the test labs, was only a qualified yes. While some of the products shone, many ran into serious difficulties and a few suffered outright failures (for details, see "PC-Based SQL: Time to Commit?", *PC Magazine*, October 12, 1993).

This year's testing offers a much rosier picture. Even though we more than quadrupled the size of our test database and boosted the complexity of our performance tests, all vendors came through with flying colors. While we still saw a wide range of performance

IN THIS REVIEW

Photography by Scott Van Sicklin

results, stability and reliability have improved dramatically across the board.

This is not to say that client/server SQL has become a trivial exercise. Integrating and debugging client and server operating systems, application code, networking components, and the SQL database itself demand real expertise. Even the best database must be backed up by solid client systems, networking cards and cabling, the network operating system and protocols, and the server hardware itself.

The advantages to client/server computing clearly predominate, however. For on-line transaction processing (OLTP) and decision-support applications, client/server offers reasonable hardware costs, faster application development, and for your end users, the familiar PC environment. Upsizing PC databases to client/ server carries with it the benefits of greater reliability, lower network loads, and centralized management. But whichever path you're on, SQL database servers for the Intel platform have made a quantum leap in quality.

OUR REVIEW LINEUP

This story covers most major 32-bit SQL database servers currently available for the Intel platform. All products covered in last year's story receive follow-up coverage and are reviewed in full if they have been released in major new revisions. Our main reviews, based on five months of lab testing, cover Informix OnLine for SCO Unix 5.01, Microsoft SQL Server for Windows NT 4.21, Oracle7 Server for NetWare 7.0.16, Sybase SQL Server for NetWare 10.01, Watcom SQL Network Server for NetWare 3.2, and XDB-Enterprise Server 4 for Windows NT.

We attempted to test and review the ASK Group's Ingres Server for OS/2 6.4.3. Ingres Server had serious difficulties with last year's tests, and many of the problems we found had not been resolved in the version we saw this year. During testing, The ASK Group was acquired by Computer Associates International, which withdrew Ingres Server from the market for debugging (for details, see the sidebar "Ingres Server: Still on Hold").

Client/server offers reasonable hardware costs, faster applications development, and for end users, the familiar PC environment.

Two databases covered in last year's story, Gupta SQLBase Server for NetWare and IBM DB2/2 have not been upgraded in the interim but will ship in major new revisions within the next few months. We were able to put a beta version of SQLBase Server through some of our tests (for details, see the sidebars "Preview: Gupta SQLBase Server" and "Coming Soon: A New DB2/2"). We were also able to examine a beta version of Borland International's new SQL entrant, The Borland InterBase Workgroup Server 4.0 (for details, see the sidebar "Preview: Borland's InterBase").

Cincom Systems and Raima Corp. declined to participate in this story because they could not free up support resources during our test cycle. Btrieve Technologies' NetWare SQL (formerly Novell's NetWare SQL) has not had a significant upgrade since it was last reviewed.

OS UNDERPINNINGS

Vendors were allowed to specify a 32-bit operating system for their server platform. Three OS's are represented this year: Microsoft Windows NT, NetWare, and SCO Unix.

Windows NT, which shipped during the past year, is a new platform for SQL. Its thread-based model, graphical administration tools, and strong networking support worked well for both Microsoft SQL Server and XDB-Enterprise Server. Other vendors, including IBM and Sybase, also plan to ship NT versions of their products.

NetWare, chosen by Oracle, Sybase,

HIGHLIGHTS

SQL Databases

ADVANCED FEATURE SETS are fast becoming standard as SQL databases grow ever more sophisticated. Most of the products we saw support ANSI cursors, triggers, stored procedures, and declarative referential integrity. All support BLOBs and cost-based query optimization.

SYMMETRIC MULTIPROCESSING is clearly the next performance frontier as SMP hardware becomes more common. Some engines use operating-system threads or processes to divide tasks over CPUs; others launch multiple instances of the database itself. Though some servers are SMP-ready, others require that you buy a special SMP version. Upcoming releases will add dedicated support for parallelizing queries, loading, and index creation. The only platform bucking this trend is Novell NetWare, which does not support multiple CPUs.

PRICES ARE DROPPING, driven by increased competition. Much of the pressure comes from sophisticated bundles targeted at the workgroup market. Client/server SQL remains an expensive proposition, however. The need for skilled administrators and the lack of turnkey software solutions means substantial outlays for software development and maintenance.

COMMAND-LINE ISQL, the traditional SQL interface to the database server, has just about seen its day. Most products now ship with menu-driven tools for setup, tuning, and administration. Some sport sophisticated Microsoft Windows–based interfaces, a trend that will be widely copied in the coming year.

and Watcom, remains somewhat controversial because the operating system, database, and any server-based utilities all run at Ring 0, the most privileged level of the Intel 386 protection scheme. In practice, we found NetWare to be problem-free once properly set up. Ring 0 operation is also extremely fast.

SCO Unix, chosen by Informix, is now a mature, highly stable product. Its only drawback is that it demands solid expertise on the administrator's part. OS/2, which was not chosen by any of this year's entrants, seems to be lagging in popularity as a SQL server OS.

ROBUST FEATURE SETS

This year marks the first time that relational database technology on the PC can be considered a broad success. All the tested products showed solid transaction-processing technology, the core of any SQL database. Log managers and lock managers all functioned smoothly, and we saw none of the instabilities, crashes, and data loss that plagued last year's lineup. These products offer what mainframe users take for granted: the ability to recover from a total system shutdown with data integrity intact.

We also saw a clear trend toward converging feature sets where many formerly cutting-edge features are fast becoming commodities. All the reviewed products except Watcom SQL and XDB-Enterprise Server support both triggers and stored procedures, and both Watcom and XDB will add them in upcoming releases. All the reviewed products support storage of binary large objects (BLOBs) in the database. All but XDB-Enterprise support two-phase commit, although only Oracle7 supports it transparently.

There may be not-so-subtle differences in how common features are implemented, however. For example, while all the reviewed products except Microsoft SQL Server support declarative referential integrity, XDB-Enterprise Server offers the most flexible approach. Child records can be automatically updated by changes to a parent record (a feature called *cascading update*) or deleted if the parent is dropped (*cascading delete*). Oracle7 supports cascading deletes but not cascading updates. Informix OnLine, Sybase SQL Server, and Watcom SQL take another tack, simply restricting any operation that tries to remove a parent with references to existing children. Comparable differences exist among implementations of triggers, stored procedures, and two-phase commit.

SPEED LIMITS

How fast a database delivers your data is always a big concern. While this year's performance results look lower than last year's because of our revamped tests and larger test database, performance has actually improved, in some cases significantly. Part of the reason is cost-based optimization, now used by all the reviewed products. Another is the maturing of lock and cache managers.

As giant applications strain the limits of existing servers, the next performance horizon is clearly the use of symmetric multiprocessing (SMP). Intel-based SMP hardware is becoming more common and under optimal conditions can deliver doubled performance when CPU and disk resources are doubled (for more information, see the sidebar "Intel-based SMP: How Strong?").

One performance question left unanswered in last year's story was how well Intel-based SQL servers stack up against heavyweight RISC platforms. In tests using Sybase System 10 that pit five high-end RISC servers against an Intel-based SMP server, we saw the Intel system

• Oracle7 Server for NetWare, Version 7.0.16

Designing a SQL database server is a tremendous challenge. It must provide the safety and integrity of a mainframe. It must be fast, robust, and above all, completely stable. Our Editors' Choice award for SQL database servers goes to Oracle7 Server for NetWare, the product that comes closest to meeting the ideal. Oracle7 is a virtual compendium of the industry's best features, and its solid core technology, including a multiversioning consistency model and row-level locking, give the product a clear performance edge. It ran friction-free through our punishing test suite, finishing in first place in most categories. Its impressive scores were obtained with almost no tuning. Oracle7 ships with a strong suite of administration tools and is well suited for distributed databases. Oracle7 demands deep pockets and professional administration skills, but it remains the overall best choice in high-stress transaction-processing environments.

An honorable mention goes to Microsoft SQL Server for Windows NT. Its strong performance, superb graphical administration tools, easy setup, and tight integration with the Windows NT operating system make for a compelling package. Significant enhancements to the base kernel include implementation of native Windows NT threads, making the package SMP-ready out of the box. Virtually everything needed for success is included in this attractive bundle.

Rounding out the top three contenders, Sybase SQL Server for NetWare delivered superb performance and the benefits of Sybase's sophisticated, feature-rich engine. We look forward to the release of Sybase System 10 add-ons for this product.

Our Contributors: BRIAN BUTLER, who directed testing for this story, is the president of Client/Server Solutions, a St. Louis–based firm specializing in SQL database performance testing and applications development. LORI MITCHELL is an associate project leader, and KASON LEUNG and ANATOLIY NOSOVITSKIY are technical specialists at Ziff-Davis Labs. THOMAS MACE was the associate editor in charge of this story, and MARK JONIKAS was the project leader at Ziff-Davis Labs.

clearly holding its own (for details, see the sidebar "Competing with RISC").

SHINY NEW TOOLS

The days are gone when simple command-line Interactive SQL (ISQL) tools were the state of the art. Administrators increasingly expect a bundled set of menu- or GUI-based tools for database creation, administration, and tuning chores. Long-established vendors such as Informix, Oracle, and Sybase are playing catch-up while Microsoft, with its experience in application interfaces, is a clear leader in sophisticated Windows-based tools. Oracle has shipped Windows-based administration tools with its Workgroup Server product (not reviewed here) and Sybase has Windows tools in beta testing. Watcom and XDB will move to GUI tools in future releases.

SIMPLER PRICING

SQL database prices are clearly on a downward trend, driven by new packages aimed at the departmental and workgroup markets. Pricing models have also gotten simpler. Where most vendors used to charge separately for users, client software, and networking components, all of the products in this story except Informix OnLine are priced on a per-user basis. While prices for the reviewed products vary widely, a price/performance analysis shows most products deliver similar bang for the buck (for details, see the sidebar "The Price of Performance").

Features, price, and performance can create daunting choices. Despite the hurdles, the news is good: PC-based SQL has never been stronger. The reviews that follow will help you find the database best suited to your needs.

Suitability to Task: SQL Databases

SQL databases were created for huge mainframe applications, but in today's PC-based client/server incarnations, they find employment in a wide range of tasks. Qualities that shine in one area can be drawbacks in another, and products that are tuned to excel in certain operations may falter elsewhere. Taking feature sets and test performances into account, we examine each reviewed product from four different perspectives.

For **production OLTP** applications, a database must be absolutely stable and offer excellent multiuser performance, a function of its locking model, cache management, and transaction-log management. We also look for support for triggers, stored procedures, declarative referential integrity, and on-line backup. Support for transparent two-phase commit and symmetric multiprocessing hardware are a plus. Strong scores on our Random Write Transaction Mix test contribute substantially to the rating.

To judge suitability for **decision support** applications, where large volumes of data are regularly moved to a decision-support server for analysis, we look for excellent loading, indexing, and ad hoc query performance. We also look for an efficient cost-based optimizer. Engine support for bidirectional scrollable cursors is a plus for easing development of GUI-based applications. Compatibility with industry-standard mainframe databases earns additional points.

Workgroup database servers frequently exist outside of an IS framework and pose a different set of demands. Here, we look for easy installation, ease of use, few tunables, and high-quality documentation that does not assume expert knowledge on the reader's part. A strong set of visual administration tools is a plus, as are an overall low initial acquisition cost and a good price/performance ratio. Databases that demand professional administration skill and intimate knowledge of the underlying operating system did not fare as well in this category.

Connectivity and deployment affect many other tasks. Here, we look for support for a wide range of network protocols and client environments. Server support for multiple concurrent protocols earns extra points, as do strong tools for monitoring and tuning the network, the operating system, and the database itself. We also look for a good selection of precompilers, a well-documented call-level C API, and the availability of gateways and connectivity products, either from the vendor or from a third party. A robust set of intuitive, GUI-based administration tools is a plus.

SUITABILITY TO TASK

Company Name	
Production OLTP	EXCELLENT
Decision support	GOOD
Workgroup database	FAIR
Connectivity & deployment	POOR

Informix Software Inc.

• Informix OnLine for SCO Unix

ALL REVIEWS BY BRIAN BUTLER AND THOMAS MACE

Last year's roundup of SQL databases wasn't easy on Informix. While Informix OnLine for NetWare offered many strong features, it was plagued during our multiuser tests by numerous crashes.

This year, we looked at a major new release on a different platform: Informix OnLine for SCO Unix, Version 5.01. The bugs are gone, testing ran smoothly, and the product's feature set has been significantly enhanced. Informix OnLine's performance scores, which fell in the midrange of the review lineup, are comparable with last year's results. But its high price—more typical of the Unix world than of the competitive PC marketplace—gave it the poorest price/performance ratio of any product in the lineup.

We caught Informix OnLine right before a major new 6.0 release that will address a number of performance issues. The version we reviewed in last year's story, Informix OnLine for NetWare, Version 4.1, has not been upgraded, and the company has no plans to bring it up to date with its Unix cousins.

Informix OnLine has long offered a robust set of engine features, including a cost-based optimizer, engine-driven backward-scrollable cursors, cursor context preservation, mirroring of databases and transaction logs, and on-line backup.

NEW TO THIS RELEASE

The new release adds a number of features that are quickly emerging as industry standards. These include stored procedures (which can return multiple rows), triggers (added in Release 5.01), and declarative referential integrity (Restrict only). Restrict ensures that a user cannot delete parent records that have dependent child records. Automatic deletion of child records is not supported and must be coded using triggers. The database also supports entity integrity by enforcing acceptable data values (including default values) for particular columns. This release does not support group-level security or audit trails, however.

Informix has enhanced the cost-based

Preview: *Borland's InterBase*

By Brian Butler and Thomas Mace

Borland International hasn't exactly been a leader in client/server databases, but the company is staking much of its future on a push into the client/server arena. While Borland's desktop databases and development tools will figure in this strategy, the cornerstone will be The Borland InterBase Workgroup Server, Version 4.0, a SQL database server due for release on a number of platforms this fall. Releases on Microsoft Windows NT and NetWare should be out by the time this article appears.

InterBase, created by InterBase Software Corp., is a technically advanced engine that found an early niche in the on-line complex processing (OLCP) market because of its pioneering support for features such as multiversioning, BLOBs, and multidimensional array data types. But the product languished after its initial sale to Ashton-Tate and up until now has seen little growth under Borland (at press time, InterBase 3.2 was the currently shipping version).

Our look at an early beta of the new InterBase, Version 4.0 for NetWare, revealed an enhanced product repositioned as an upsizing tool. The biggest change is that InterBase can now interface directly with Borland's desktop databases dBASE 5.0 for Windows and Paradox 5.0 for Windows.

STRONG CORE ENGINE

InterBase offers a strong core set of features that includes declarative referential integrity, triggers, stored procedures, event alerters, user-defined functions, a cost-based optimizer, BLOB support, and transparent two-phase commit. InterBase is based on a multiversioning database engine, an approach it shares with Oracle7.

Multiversioning provides transactions with a read-consistent view of the database: A given transaction sees the database as it was at the moment the transaction began, and multiple transactions can see the database in several different consistent states. The main advantage to multiversioning is that read transactions, especially long-running reads typical of decision-support applications, do not acquire locks that block write transactions, improving overall concurrency.

THE DESKTOP CONNECTION

InterBase offers a unique synergy with Borland's desktop database products dBASE for Windows 5.0 and Paradox 5.0 for Windows. Both can connect directly to InterBase, which in turn provides direct engine support for the desktop products' native record-navigation commands (in addition to InterBase's support for standard SQL). This lets developers migrate dBASE and Paradox apps to a client/server environment or use these PC databases as front-end development tools.

Borland is staking much of its future on a push into the client/server arena.

dBASE for Windows and Paradox for Windows share a common local database engine called the Borland Database Engine. This includes an IDAPI (Independent Database API) component for connecting to InterBase and other SQL databases. The IDAPI InterBase driver, called Client/Server Express, provides the direct low-level interface to InterBase. Since InterBase directly supports dBASE's and Paradox's record navigation, you can use commands such as dBASE's Skip -1000 and Go Bottom with InterBase data. The IDAPI component also includes Borland's SQL Link drivers for connecting to Microsoft SQL Server 4.21, Oracle7, Sybase SQL Server 10.01, and ODBC-compliant databases. These drivers also let you use dBASE or Paradox commands with third-party SQL databases, but only through a potentially slower SQL translation. The ODBC component of SQL Link will also let you develop applications for InterBase using non-Borland tools.

PACKAGING

In addition to the pending Windows NT and NetWare versions, ports for DEC Alpha OSF1, HP-UX, Sun OS, and Sun Solaris are scheduled to ship this fall. A Chicago version will ship soon after Microsoft's release of Chicago, and an OS/2 version is due by early next year. Borland also plans to bundle a version of Delphi (the code name for its upcoming Visual Basic competitor) for use in off-line applications development. Pricing is expected to be highly competitive, and client software, including the SQL Link and Client/Server components, ODBC drivers, a set of Windows-based administration tools, and the InterBase C API libraries, will be bundled free with every server.

A few holes remain in the current InterBase strategy. The product cannot operate with the new dBASE for DOS 5.0, so migrating existing dBASE apps to InterBase means porting them to Windows. Also, many of the engine's most advanced features are only accessible through proprietary interfaces. But InterBase's tie-in with Borland's Windows databases is compelling. dBASE and Paradox developers will certainly want to evaluate the product when it ships. □

optimizer to be more intelligent about its choices, and it now lets you set the optimization level of the query. The default is High Optimization, which performs an exhaustive search through all possible access plans and picks the one with the lowest cost. With complex queries involving many tables, this process can be more expensive than the actual execution of the query. In such scenarios, you can select Low Optimization, which will make a quick best guess.

SUITABILITY TO TASK Informix OnLine for SCO Unix	
Production OLTP	GOOD
Decision support	FAIR
Workgroup database	POOR
Connectivity & deployment	GOOD

The optimizer did not make any mistakes during our tests, but we did run into a problem updating the optimizer statistics. A bad value placed in the statistics page caused two subsequent queries to crash the server. This was fixed by modifying the statistics page manually.

The new release has also improved Informix OnLine's index-creation speed. Last year, it was the slowest product at indexing our test database by a huge margin. This year its indexing score, while not exactly zippy, was more in line with other competitors'. Under Informix's new indexing scheme, index entries are sorted prior to their insertion into the B+tree structure.

The Informix OnLine engine has always offered strong binary large object (BLOB) support. As with the previous version of the product, BLOBs are stored in a distinct Blobspace, allowing you to tune the associated page size separately for best performance. The maximum allowable BLOB size is 2GB. BLOBs are written directly to disk, not to shared-memory data buffers. This saves space in the transaction logs and keeps the pool of shared-memory data buffers from being swamped. With the optional Informix-OnLine/Optical add-on product, BLOBs can be stored on WORM (write-once-read-many) optical subsystems. Unfortunately, we did not get a chance to test Informix OnLine's BLOB throughput capabilities because of time constraints in our test cycle. This was not due to any problem with the product.

Informix OnLine provides locking by row, page, table, or database and the unique ability to configure locking on a table-by-table basis. You can have one table set for a page-level locking scheme, another with record-level locks, and yet another large lookup table set for table-level locking. Isolation levels are also highly tunable and include support for dirty reads (no isolation), committed read isolation, cursor stability, and repeatable reads.

Version 5.0 added support for distributed Informix OnLine databases through the separate Informix-Star product. Informix-Star adds a two-phase commit protocol and lets users transparently manipulate multiple Informix OnLine databases at several locations. The current lets you update multiple databases on a single Informix OnLine server instance in a single transaction.

KNOW YOUR UNIX

While database administration does not usually call for much knowledge of the underlying operating system, this version of Informix OnLine demands a good working knowledge of Unix. During installation, we had to modify some SCO kernel parameters to get the package up and running. This process is documented in the machine notes file on the system.

Like most Unix database vendors, Informix recommends that you set up raw file partitions, a task that can be a tricky process. Because the Unix file system has its own cache, the database has no way of ensuring that writes have been physically committed to disk. This can lead to serious integrity problems if the system crashes. Using raw file partitions bypasses the Unix file system, the only way to ensure integrity loss doesn't happen.

Once the database is installed, you can use the supplied DB-Monitor utility to configure various system parameters including buffers, locks, users, and tables. This menu-driven utility also lets you change the server's mode of operation to on-line, off-line, or quiescent mode (a single-user administration mode). DB-Monitor also provides backup, recovery, and a window into virtually everything the engine is doing. It can display a multitude of statistics to aid in the tuning process, such as cache hits, disk reads and writes, and checkpoints.

The database also ships with a menu-driven setup tool called DB-Access for creating databases and tables and executing SQL statements. A set of command-line utilities, which can be driven by scripts, provides additional administrative functions.

COMING DOWN THE PIKE

We narrowly missed the next major release of Informix OnLine, Version 6.0, which should be shipping on the SCO Unix platform by the time this story appears. Where the 5.0 release is generally targeted at broadening engine functionality, Version 6.0 is primarily aimed at boosting performance.

Informix has rebuilt large portions of the database server, replacing the current process-based engine with an internal multithreaded system. The most important change will be the ability to exploit symmetric multiprocessing hardware through the addition of parallel index creation, parallel thread-level sorts, and parallel backup and restore capabilities. An upgraded 6.0 optimizer will be able to maintain data-distribution histograms. Declarative referential integrity support will be extended to cover cascading deletes.

▶ FACT FILE

Informix OnLine for SCO Unix, Version 5.01

List price: Server software, one development system, 60 client connections, and client software: $29,395. **Requires:** Server: 386-based PC or better, 2MB RAM, 5MB hard disk space, SCO Unix System V 3.24 or later. DOS client: 286-based PC or better, 700K RAM, 3.2MB hard disk space. **In short:** Version 5.01 of Informix OnLine adds significant enhancements to this veteran Unix database. New features include triggers, stored procedures, and declarative referential integrity. Unix knowledge is required, but the bundled administration tools make for easy setup and tuning. Informix OnLine ran smoothly through our benchmark tests, although its performance scores remain in the midrange. Its high price gives it the worst price/performance ratio of the roundup. By the time this story appears, a new Version 6.0 should be available that offers major performance enhancements.

Informix Software Inc., 4100 Bohannon Dr., Menlo Park, CA 94025; 800-331-1763, 415-926-6300; fax, 913-599-8753

481 on reader service card

Preview: *Gupta SQLBase Server*

By Brian Butler and Thomas Mace

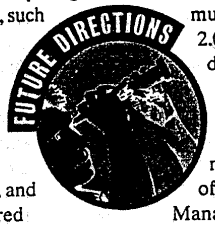
Gupta SQLBase Server for NetWare was the worst casualty of last year's testing. Version 5.12 took almost 60 hours to load our database—more than 10 times as long as the next-slowest competitor—and crashed repeatedly on index builds. Testing never got beyond this point.

As we go to press, Gupta is about to ship SQLBase Server for NetWare, Version 6.0—a major new release that will extend the server's feature set and target the problems we encountered. We invited Gupta to run through our Load and Index and Ad Hoc Query tests using a beta of this upcoming release. Testing was done at Gupta Corp. on a Compaq ProLiant configured similarly to our test-bed.

Loading and indexing ran without a hitch, even though our test database is more than four times as large as last year's. Total load-and-index time was also significantly faster, placing SQLBase within reach of other products in this story (although it would still have placed last). SQLBase also ran smoothly through our ad hoc queries and demonstrated times that were reasonable but again were not as fast as the times posted by the other tested products.

PC-CENTRIC

SQLBase was designed from the ground up as a small-footprint engine for PC-based client/server computing. Some of its slickest features, such as bidirectional scrollable cursors, are especially at home in Microsoft Windows–hosted applications.



The new release fleshes out the feature set with stored procedures, triggers, and timer events. SQLBase stored procedures are written in the SQLWindows Application Language (SAL), providing close ties with Gupta SQLWindows—Gupta's well-known front-end development tool. You can specify whether the new SQLBase triggers fire before or after the triggering operation. Timer events are stored procedures that can be set to execute at a specific time or at predetermined intervals.

SQLBase 6.0 shows enhancements to usability as well. New utilities can automate installation: the new SQLEdit utility, a particularly welcome breath of fresh air, automatically handles network configuration for both clients and servers. In previous versions of SQLBase, you had to edit the SQL.INI file manually—a confusing, tedious process.

Support for distributed databases has also been strengthened. SQLBase 6.0 will offer transparent two-phase commit to manage transactions across multiple servers. SQLConsole 2.0, an impressive new Windows-based remote-management utility, will be bundled with the server. This slick tool allows remote tuning, monitoring, and maintenance of multiple servers through its Manager modules.

The Scheduling Manager automates such maintenance tasks as backups. The Alarm Manager monitors the network for more than 20 definable events and automatically executes an appropriate response when necessary. If the event remains unresolved, the alarm can trigger further responses. The Database Object Manager lets you graphically manage every database component, including stored procedures and triggers. SQLTrace is a debugging tool that can trap SQL traffic between a client and the server. You can replay the SQL through SQLTrace's graphical debugger.

Gupta appears to have made great strides with SQLBase 6.0, addressing stability, performance, ease of use, and administration in a single release. □

An even newer release, Version 7.0, was codeveloped with Sequent Computer Systems and is already shipping on Sequent's Symmetry multiprocessing platform. This version adds parallel data query (PDQ) capability and the ability to optimize the partitioning of tables based on the contents of the data.

Informix has long been known for its strong Unix databases, but it has been less active in client/server products for the PC environment. This seems to be changing. The current release, while no screamer, shows major improvements over previous versions. The pending 6.0 release seems poised to add the missing element of top-flight performance.

Microsoft Corp.

• Microsoft SQL Server for Windows NT

For those who've wrestled with mix-and-match client/server environments, Microsoft SQL Server for Windows NT, Version 4.21, offers all the seductions of one-stop shopping. For a very competitive price, it delivers a powerful SQL engine, superb tools, strong networking components, and the benefits of close integration with the Microsoft Windows NT operating system—all in a single box.

There are some caveats, particularly for enterprise applications. The server is not ideally equipped for distributed environments and does not support replication. Its performance, while generally very fast, was surprisingly slow in a few areas critical for decision support. More fundamentally, Microsoft is clearly steering you into a total Windows NT solution, something that may be incompatible with larger enterprise strategies. But for workgroups and larger departments—even those with heavy transaction loads—Microsoft SQL Server is a compelling solution.

MAJOR REWRITE

Although Microsoft SQL Server had its genesis in Version 4.2 of Sybase SQL Server, Microsoft significantly rewrote

many important system components for the current release. While the changes are largely targeted at improving integration with Windows NT, they also fixed a few idiosyncrasies and made some important enhancements to the engine. While Microsoft has been careful to preserve full compatibility with the older Microsoft SQL Server for OS/2, Version 4.2, both Microsoft's and Sybase's SQL Server products are now clearly headed in different directions. The only official compatibility between them is at the 4.2 level of DB-Library.

The level of integration between Microsoft SQL Server and Windows NT is high. Microsoft discarded the internal threading engine used in its OS/2 product and has implemented Microsoft SQL Server as a single process using native Windows NT threads. Threads are preemptively scheduled and can be distributed over multiple processors, making Microsoft SQL Server SMP-ready out of the box.

The database also uses Windows NT's asynchronous I/O capabilities to handle physical inputs and outputs concurrently with other operations. As a whole, the database runs as an operating system service that can be started, stopped, or paused from the Windows NT Control Panel. Windows NT also lets Microsoft SQL Server simultaneously support multiple network protocols and connection types, including IPX/SPX, Named Pipes, NetBEUI, sockets, and TCP/IP. Server-based gateways to other databases can be written through the Microsoft Open Data Services (ODS) API. Backups are handled via Windows NT's backup facility. You can dump multiple databases to a single device and schedule on-line backups. Microsoft SQL Server supports any backup devices that are also supported by Windows NT.

Microsoft SQL Server integrates with the Windows NT Performance Monitor to provide a graphical display of database, network, operating system, and hardware performance data such as CPU utilization, I/O activity, cache hits and misses, database users, and network connections. This window into a client/server system takes much of the guesswork out of performance tuning and provides a firm guide when making hardware modifications. The Performance Monitor also lets you

Coming Soon: *A New DB2/2*

By Brian Butler and Thomas Mace

If anyone knows databases, it's IBM. That's why last year's look at IBM's OS/2 database—DB2/2, Version 1—was such a disappointment. DB2/2 proved stable in testing, but it was extremely slow and lacked basic amenities such as the ability to span a database across multiple logical volumes. NetBIOS was the only supported network protocol, and though you could connect to a DOS client, there were no available DOS development tools (all development had to be done in OS/2). A point revision, Version 1.2, improved network connectivity and added ODBC support and DOS client development tools but failed to address other shortcomings.

An important new release, Version 2.0 of DB2/2 is poised to energize this hitherto stodgy product. Enhancements will include a totally revamped query optimizer, flexible tablespace allocation, and a host of new engine features. Version 2.0 will be entering beta testing in the fall and should be generally available early next year. This release will be available for both OS/2 and Microsoft Windows NT.

STARBURST OPTIMIZER

New optimizer technology is high on IBM's list of enhancements. DB2/2's new cost-based optimizer, called Starburst, has been developed by some of the research-team members responsible for IBM's pioneering System R, the original prototype of DB2. IBM claims that Starburst will be the most advanced optimization technology on the market—more advanced than the mainframe version of DB2, with which it will share some technology. An accompanying visual tool will show a graphical representation of the access plan chosen by the optimizer for assistance in tuning. A graphical performance monitor will also be included.

Version 2.0 of DB2/2 will also address the previous one-volume limitation on database size by letting you divide a database into separately managed tablespaces. You will now be able to specify where tables or indexes are created by specifying the tablespace in which they reside. On-line backup capability at the database or tablespace level will also be added.

The new release will bring the engine feature set up to date by adding user-defined functions and data types, triggers, constraints, recursive SQL, and BLOB support. The DB2/2 engine has also been rewritten to support native operating-system threads, making it SMP-ready. While DB2/2 will not have its own server-based 4GL, you will be able to write 3GL stored procedures as DLLs. In addition, the new version will include Distributed Relational Database Architecture (DRDA) Server capability. (The previous release was a DRDA Requester only.) Two-phase commit will be supported.

IBM will also be releasing a set of data-replication products that support replication from multiple sources—including DB2/MVS, DB2/400, IMS, and VSAM—into DB2/2 and DB2/6000 databases. On the networking side, Version 2.0 will add support for TCP/IP.

DB2/2's upcoming feature set looks strong. If IBM delivers performance to match, this product will be a force to be reckoned with. □

Performance Tests: SQL Databases

How We Tested

Our demanding tests revealed improved performance across the board. Oracle7 took first place overall while Sybase led the pack in multiuser read transactions. Microsoft SQL Server delivered strong results everywhere except in ad hoc queries and load-and-index operations. Watcom SQL and XDB-Enterprise brought up the rear.

To evaluate the SQL relational database management systems in this roundup, we used a heavily modified version of the AS^{3}AP (ANSI-SQL Standard Scalable and Portable) Benchmark Tests for Relational Database Systems, originally developed at Cornell University by Dina Bitton and associates. This set of cross-platform performance tests covers a wide spectrum of typical database operations (although based on ANSI SQL, the AS^{3}AP tests are not an ANSI benchmark.)

For our database server, we used a Compaq ProLiant 4000 equipped with a single 66-MHz Pentium processor card, 128MB of ECC RAM, five 2.1GB Hewlett-Packard disk drives in an external cabinet, four Compaq EISA NetFlex-2 network adapter cards (configured for Ethernet), and a Compaq Smart SCSI-2 Array disk controller. Compaq's hardware RAID 0 striping was available to vendors if they chose to use it. On the client side, we used a network of 60 physical clients comprising a mix of 386- and 486-based machines. All clients were equipped with 8MB of RAM and an NE2000 network card. The network was divided into four segments (15 clients per segment); each segment communicated with a separate network card on the server. The Ad Hoc Query test workstation was a 486/33 PC equipped with 8M of RAM and an NE2000 network card.

All vendors were invited to Ziff-Davis Labs to observe testing and help us tune the database engines. Among the vendors whose products we reviewed, only Informix declined to send a representative. To give Informix equivalent representation during testing, ZD Labs hired Gregory D. Balfanz, an Informix consultant and Unix specialist from Open Systems Engineering of Boerne, Texas, to help us tune the Informix database.

Vendors were allowed to run their products under their choice of Intel-based operating systems and network protocols. Informix Software chose to run its Informix OnLine for SCO Unix 5.01 under Santa Cruz Operation's SCO Version 4.2 using TCP/IP. Originally, Informix chose to run Version 5.02 of the server, but after we encountered a memory-leak bug during our Load testing, the company substituted its 5.01 release. Microsoft ran Microsoft SQL Server for Windows NT 4.21 on Microsoft Windows NT Advanced Server 3.1 using Named Pipes on top of NetBEUI. Oracle Corp. ran Oracle7 Server for NetWare 7.0.16 on NetWare 3.11 using SPX/IPX. Sybase chose to run Sybase SQL Server for NetWare 10.01 on NetWare 3.12 using TCP/IP. Watcom International ran Watcom SQL Network Server for NetWare 3.2 on NetWare 4.01 using SPX/IPX. Finally, XDB Systems ran XDB-Enterprise Server 4 for Windows NT on Windows NT Advanced Server 3.1 using TCP/IP. The Windows NT products applied Service Pack 2 to the operating system. The client-side TCP/IP stack was FTP Software's PC/TCP Plus 2.3.

Our test database consisted of ten tables containing a total of 18.61 million rows. The breakdown of the table sizes was as follows: one table with 7 million rows, one table with 5 million rows, one table with 2 million rows, four tables with 1 million rows each, one table with 100,000 rows, one table with 10,000 rows, and one table with 5,000 .GIF images. We also created two empty tables used for inserts. The database size typically ran well over 2GB when fully loaded and indexed.

The raw data for our test database was generated using the AS^{3}APGen 2.0 program from Dina Bitton and Jeff Millman at DBStar of San Francisco, California. All the tables had the same structure, and each row was approximately 160 bytes long, although the exact values varied by vendor. The test data for each table was supplied in the form of an ASCII comma-delimited file. The data types in the database columns included integer, floating-point, and date, as well as fixed-length and variable-length character strings.

The multiuser tests were automated using the Benchmark SDK utility from Client/Server Solutions of St. Louis, Missouri. All of our multiuser tests measure total system throughput—the amount of work that the system is performing every second—calculated in transactions per second (tps). We generated tps scores using 11 different client-load levels ranging from 1 to 60 simultaneously active network clients.

Beginning with a single client, we ran each client level for 10 minutes. Scores for the first 3 minutes 45 seconds were discarded to allow the database cache to stabilize. During the next 5 minutes, we counted the number of transactions executed. This was followed by a rampdown interval of 1 minute 15 seconds, during which no measurements were made. Before moving to the next client level, we added a 30-second quiet period to allow the network to settle. This overall approach allows us to guarantee accurate and consistent scores. Transactions are processed as quickly as the database allows; test code does not include think time. This generates a workload far greater than 60 real-world clients would produce.

WEIGHING THE RESULTS

This year's testing was based on a significantly larger test database than last year's, and our test queries were considerably more demanding. As a result, this year's raw scores are considerably lower than last year's, despite the use of Pentium-level server hardware and 32MB additional server RAM. The best comparison with last year's results is provided by the Single Random Read test, which was not redesigned for this year's testing. Even assuming that the hardware used this year is twice as fast as last year's (and discounting the larger test database size), we still saw improvements of between 15 and 270 percent.

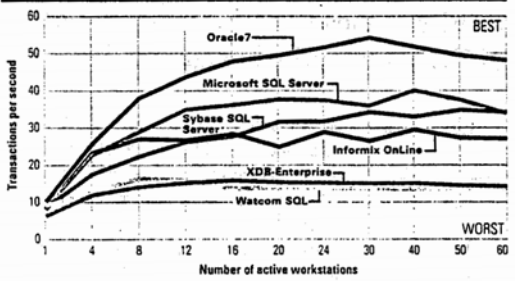
In general, it is important to realize that a benchmark testing scenario can bring optimizations into play that may not be fully exploited in real-world situations. A good example is the issue of manually striping the database across multiple disks versus using the hardware to stripe it. In a benchmark situation, a vendor can often achieve optimal performance by manually placing the database objects on the disk subsystem because the transactions and access methods are very well defined. Given enough time and intimate knowledge of the database engine, the vendor can find an absolutely optimal balance of inputs and outputs across the disk drives.

This type of optimization is usually achieved

through trial and error, which is costly in both time and resources. In the real world, it is very rare that the administrator has comparable knowledge of data access and the database engine—let alone the time for experimentation. Informix and Sybase chose to stripe the database manually, while Microsoft, Oracle, and Watcom used hardware-level striping. XDB chose not to use hardware striping but was able to achieve significant optimization by experimenting with placement of tables and indexes on the disk array.

Below we describe the results we observed and attempt to explain these results in terms of each product's features. SQL databases are extremely complex artifacts, and it may not always be possible to isolate the many interrelated factors contributing to observed behavior.

RANDOM WRITE TRANSACTION MIX



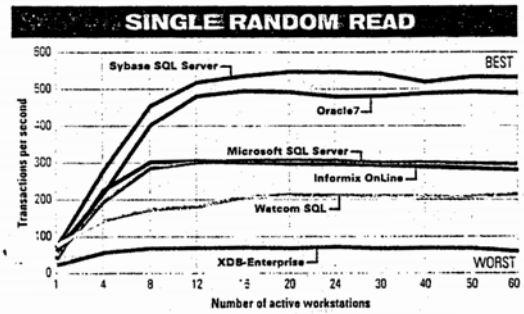
The **Random Write Transaction Mix** test, which accesses six tables in our database, simulates a heavy mixed workload of read and write transactions. This test simultaneously stresses Delete, Insert, Select, and Update functions of the database server. During the test, each station randomly selects and then executes a series of queries from a pool of five possible query types. The randomizer is constructed so that the frequency of execution for query types numbered one through five will be in a ratio of 6:4:4:3:3.

The first transaction updates an integer field in the 7-million-row table via the primary key using a Between operator. The second transaction is a two-way join between two 1-million-row-tables. The third transaction updates an integer field in a 1-million-row table and includes some in-line logic that stores the update in one of the blank tables. The fourth updates the 2-million-row table via an In clause, and the fifth moves a row from the 5-million-row table to a blank table. We used an extensive auditing script to ensure that all the products were actually performing these tests as specified.

The best performer on this extremely demanding test was Oracle7. Its high score is attributable to its record-locking scheme and efficient log management, features that have been part of the product for quite some time. The engine ran error-free and required very little tuning to achieve the measured performance level. Oracle7 only logs changes to the data, and its support for fast commit and group commits further reduces log-management overhead. The amount of data Oracle7 can write in a group commit is limited only by the operating system. Record-level locking provided optimal concurrency: No deadlocks occurred during the execution of the test. Oracle7 did not use its Discrete Transaction feature on this test.

Following close behind was Microsoft SQL Server. This test highlights how much work Microsoft has done to speed up the transaction-processing aspect of the database's engine. One new feature, *asynchronous checkpoints*, allows translation processing to continue during the checkpoint process. A Lazy Writer feature, also new to this release, lets the database engine clean up dirty database pages in the background, minimizing the work required during the checkpoint process. Microsoft SQL Server also supports group commits of up to eight 2K pages at a time. A factor working against the product may have been its page-level locking scheme.

SINGLE RANDOM READ



Since two of our transactions perform an Insert into an empty table, we created a clustered index to avoid contention on the last page. Without a clustered index, the last page of the table is consistently locked, forcing a serialization of operations. With a clustered index, Insert commands are distributed over the range of the table. The only drawback to this approach is the extra overhead for maintaining the index. We avoided deadlocks by using a fill factor on the affected indexes.

Sybase, which came in third, implemented this test with stored procedures accessed via remote procedure calls (RPCs). The company also coded several of our transactions using its newly added support for cursors within the stored procedures itself. Another new feature of the tested NetWare port is Sybase's Buffer Wash mechanism. This is a background process that cleans up dirty pages, guaranteeing a supply of free pages. While checkpoints are essentially unchanged since Version 4.2, the new Buffer Wash feature means that checkpoints must perform significantly less work. Sybase also supports group commits, but only up to a single 2K page at a time. The company used a fill factor to avoid deadlocks. For this test, we allowed Sybase to modify the database schema slightly to make the update columns Not Null. This allowed the company to work around the engine's limitation on update-in-place for nullable columns.

Informix, which came in fourth, used record-level locking on all tables that were updated. Informix performed well once the database and operating system were properly tuned, although the tuning process was not particularly intuitive. The database was stable in operation and we encountered none of the problems with failed checkpoints that plagued it in last year's tests.

Watcom SQL and XDB-Enterprise brought up the rear. Although Watcom SQL supports group commits and only logs changed data to the transaction log, its performance was only about 25 percent as fast as the fastest product. XDB-Enterprise does not support group commits and logs the entire before-and-after image of the row.

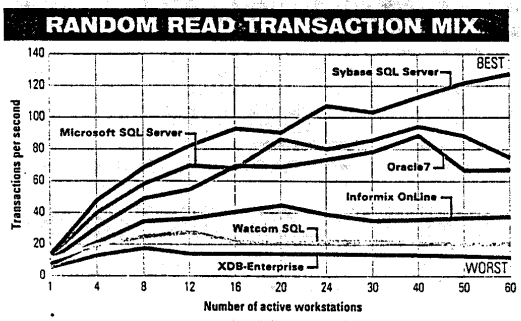
The **Single Random Read** test, based on a single-record read via the primary key, shows the maximum number of concurrent retrievals the system can handle. This test has not been modified since our last roundup and is included to show how far the products and hardware have come in the interim.

In this test, each workstation selects a random row from a single table that is then fetched across the network and discarded. All active workstations repeat this process at the maximum speed supported by the database. This scenario does not stress every component of a database engine and the results tend to exaggerate the engine's actual transaction-processing power. The small, quick transaction involved does put a significant stress on the network components of the operating system, however. For products tested under Windows NT and SCO Unix—both of which are true protected-mode operating systems—the overhead for privilege-level checking proved to be costly. Records are read in the lowest lock level each product supports, thus permitting the greatest degree of concurrency (we required that each vendor take at least a shared-level lock on the row or page). Since all locks are shared-level, no blocking occurs; multiple clients can access the same row or page without concurrency loss.

The best performer on this test was Sybase SQL Server. Contributing factors are the efficiency of its NetWare Loadable Module (NLM) architecture, Sybase's clustered indexes, and the use of stored procedures. Sybase called its stored procedures via an RPC instead of by using a straight stored procedure call. In an RPC, the function call is translated into a binary representation at the client; a normal stored procedure call is sent across the network as text and translated at the server. Sybase believes that the use of RPCs in CPU-bound situations such as this test can significantly improve performance.

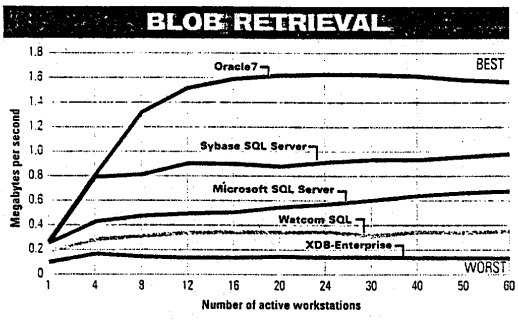
Oracle7, which came in second, did not use a stored procedure due to the simplicity of the transaction. It did open and maintain a cursor, however. Since Oracle7 has the ability to share cursors across multiple clients, this approach allowed clients to execute the transactions without having to reparse and optimize the SQL statement—in effect, the same advantage provided by a stored procedure. Oracle7 also supports a unique method of executing and fetching multiple rows in a single function call, thereby reducing network traffic. Most other products require several function calls to do the job, one to execute the query and others to retrieve the results.

Microsoft SQL Server placed third. It used clustered indexes and stored procedures but did not use RPCs. While clustered indexes usually help Microsoft SQL Server considerably, they were offset by the stress on the operating system's network communications layer.



Close behind Microsoft came Informix OnLine. We did not test Informix using its clustered indexes or stored procedures. The opinion of our Informix consultant was that stored procedures would be slower than a prepare/fetch mechanism, and the clustered indexes would have drastically slowed the index creation times.

Watcom SQL came in fourth—earning it the award for being the most improved product since last year. Improved performance was mostly due to the move to NetWare, a true 32-bit environment, and asynchronous I/O, which was not available in the DOS product we tested previously. Watcom does not support stored procedures at this time; a prepare/fetch mechanism was used to avoid the parsing/optimization phase of the query.



In last place was XDB-Enterprise, which did not use striping. The database table and index used in this test resided on a single drive, something that proved to be the biggest bottleneck. While performance could have improved by moving the index to a separate spindle, this approach would have hurt the product's multiuser test results. This highlights the problem with manually placing database objects: Tuning for one type of operation can hurt performance elsewhere.

The **Random Read Transaction Mix** test, which accesses five tables in the test database, simulates a mixed workload of read-only queries. This test was designed to stress the data-retrieval capabilities of the database engine, and proved to be extremely disk bound. During the test, each station randomly selects and then executes a series of queries from a pool of five possible query types. The randomizer uses the same ratios as for the Random Write test.

The first query is a single record read via primary key (this is identical to the Single Read Transaction test). The second is a Join on the primary key between two 1-million-row tables. The third is a Select on the 7-million-row table using a Between clause. The fourth query is a two-way join between a 1-million-row and a 2-million-row table using a Between as a restriction and a Join on a character field. The fifth query is a two-way Join between a 1-million-row and a 5-million-row table using an In clause; the Join is on a character field.

Sybase SQL Server was the clear winner. Its performance can be attributed to the use of clustered indexes and stored procedures (using RPCs), and the NetWare operating system's low overhead. Sybase SQL Server seemed able to satisfy the transaction requests with less I/O than other vendors, and its stored procedures cut down on network traffic. The package's clustered indexes must also be considered an important contributing factor since two of the queries used a Between clause on a clustered key. Since the data is physically arranged on the disk in clustered order, these queries could be typically resolved in fewer disk inputs and outputs than when using products that do not support clustered indexes. Sybase experimented with using SPX/IPX on this test, but TCP/IP, its original protocol of choice, proved to be slightly more efficient.

Microsoft SQL Server, which also used stored procedures and clustered indexes, came in second. While Sybase and Microsoft coded the test

ransactions in a similar manner, Microsoft chose ıot to use RPCs to call the stored procedures. The verhead of Windows NT may have also played a light role in the performance difference, although, as the transaction size increases, Windows NT overhead appears to decrease.

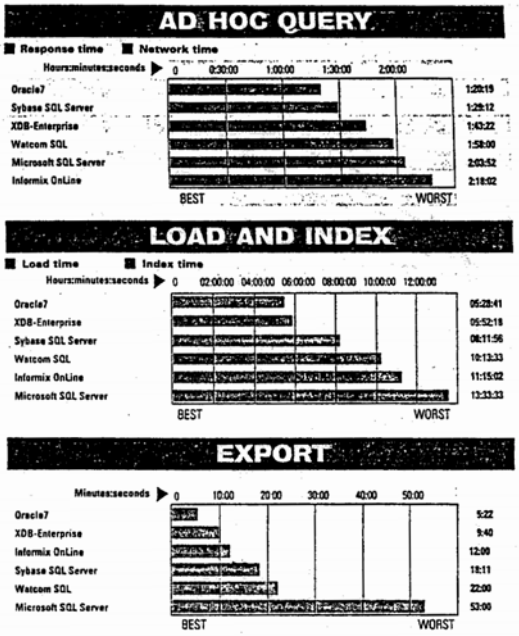
Third-place Oracle7 was able to hare the cursor among the clients, nd its ability to do an execute and ıtch in one statement also helped erformance. Oracle7 is also unique ı that it can retrieve multiple rows in single network fetch. But its performance was bottlenecked on this test y the disk I/O subsystem, something ıat could be attributable to nonclusered indexes and the nature of the ueries. As an experiment, we added nother five drives and saw tangible nprovements before the system ecame clearly CPU-bound. Other endors may well have achieved omparable improvement.

Informix came in fourth. We did ot use the clustered index feature of ıe database engine, because of its ffect on the index time. Our Informix onsultant also advised against using ıformix's stored procedures since he eels that they are inefficient for our ype of transactions. To achieve optimum performance, each station used a prepare/fetch mechaism, saving the processing overhead of a parsıg/optimization process. The overhead of SCO Inix may also have been a factor.

Watcom SQL came in fifth with XDB-Enterrise pulling up the rear. Watcom SQL did not use lustered indexes to avoid undue load times, and ıe product does not currently support stored rocedures. The transactions were coded using a repare/fetch mechanism. XDB-Enterprise's esults may be attributable to the product's ıanual distribution of database objects.

Binary large objects (BLOBs) are structures sed for storing images and other large binary elds in the database. The **BLOB Retrieval** test ıeasures how fast the client can retrieve these ırge structures—in effect, how well the database can utilize the network. All the tested roducts offer a method for fetching large blocks f data in a single network call, and many are able ɔ change the default network packet size dynamially.

This test used a database table containing 5,000 unique bitmapped images in .GIF format ranging in size from 20K to 150K, with a majority in the 70K range. During the test, clients randomly selected and retrieved a series of images. Images were not displayed, but we required that the full binary file be sent across the network. While the BLOB Retrieval test executes in a similar manner as our other multiuser tests, it measures sustained system throughput in megabytes per second (MBps).



Oracle7 was the clear winner with a maximum transfer rate of 1.61 MBps. No tuning of the network packet size was needed to achieve this result. While creating the BLOB table, however, we discovered that Oracle7 was the only product unable to load our BLOB images from a DOS client because the Oracle7 DOS client does not have a mechanism for sending BLOBs piecemeal to the server, and not enough memory could be allocated to load the entire image at once. We used an OS/2 client as a workaround. In experimenting with network protocols, we found that using TCP/IP gave Oracle7 about a 25 percent performance boost over the SPX/IPX protocol used for official testing (charted numbers show results for SPX/IPX).

Sybase SQL Server came in second with a maximum transfer rate of just under 1 MBps. The package's DB-Library does not support negotiated packet size, so the default packet size of 512 bytes was used. (The company's Open Client does support negotiated packet size, but it is currently available under Windows only.) We experimented with substituting SPX/IPX for the TCP/IP protocol Sybase chose for official tests. We observed a 45 percent performance degradation under SPX/IPX (charted numbers show results for TCP/IP).

Microsoft SQL Server came in third with a maximum transfer rate of 0.68 MBps. Microsoft tuned for this test by increasing the default packet size from 512 bytes to 4K. The negotiated packet size feature allows clients to configure the packet size at connection time. This feature is only supported under Named Pipes.

Watcom SQL placed fourth with a transfer rate of 0.35 MBps, and XDB-Enterprise was last with a transfer rate of 0.17 MBps. Watcom SQL also lets you specify the packet size when the DOS requestor is started. Watcom used a packet size of 1,450 bytes for our entire suite of tests (the default packet size is 512 bytes). While XDB-Enterprise used TCP/IP, the network was not an overriding factor in the package's performance. Since all of XDB-Enterprise's BLOBs resided on a single disk, the server remained strongly disk-bound throughout the test.

We could not obtain results for Informix OnLine due to time constraints. This was not due to any fault in the product.

The **Ad Hoc Query** test measures each product's effectiveness in a decision-support environment. The query mix is submitted from a single 486/33 client, and both the response time (the time for the first row to be returned) and the total elapsed time for each query are recorded. Response time is an important metric in a real-world environment in which the user is waiting to see results. Once the first row is returned, the user can begin scrolling through the data. Total elapsed time is more important in a batch-reporting environment in which large reports are being printed.

Because of the large number of rows returned by some of our queries, network overhead is in some cases the factor limiting performance. It is also difficult to separate engine processing speed from network overhead since many products

return rows to the user before the query is completely resolved.

The Ad Hoc Query test consists of 34 queries that stress six different types of server functions: selects, joins, projections, aggregates, sorts, and subqueries. Ten select queries measure the speed at which a database can selectively scan a table. Nine join queries show how well an optimizer can pick the fastest access path from the available indexes (the joins range from a two-way to a seven-way join). Two projection queries measure how fast a database can determine the number of distinct values in a table. Five aggregate queries calculate a variety of aggregates (minimum, maximum, average, and count). Five sort queries measure how fast the database can sort data sets ranging in size from 10,000 rows to 2 million rows. And finally, three subqueries show the effectiveness of the optimizer in resolving correlated subqueries and outer joins.

Oracle7 takes the top spot on this test with a total time of 1 hour 20 minutes. But when we first ran the test, Oracle7's optimizer made a mistake on the sort query, returning a score of over 10 hours, by far the worst score we saw. The optimizer chose to use an index when it should have performed a table scan. This entailed extra I/O in jumping between index pages and data pages and did not let the database take advantage of its read-ahead mechanism. This error was easily corrected using a Hint, a well-documented method of overriding the optimizer. Because all optimizers are based on statistics, there is always a probability of making a mistake. Consequently, an override mechanism is a must.

While the Oracle7 optimizer was not the most robust we saw on our 34 test queries, the currently shipping Oracle7 Server for SCO Unix, Version 7.1, was able to execute the same queries without hints, indicating that the problems have been addressed.

Interestingly, Oracle7 took second place in both response time and network time, yet the combination of the two made it the fastest. Oracle7 offers a way to tune the query for response time or total time. Due to the nature of our queries, the company chose to tune for total time.

Sybase SQL Server was close behind Oracle7 and was able to the run the queries untouched, something that demonstrates the strength of Sybase SQL Server's optimizer. While Sybase SQL Server only ranked fourth in terms of the response time for the queries, it was the fastest in terms of network time. We briefly substituted SPX/IPX for TCP/IP and saw that this made very little difference in the results. Third-place XDB-Enterprise sports a cost-based optimizer and a read-ahead mechanism. It was able to run the queries unaltered.

Watcom SQL ranked forth, and once again gets the most-improved award, having taken last place in last year's tests. The addition of read-ahead capability and the work done to improve the optimizer have clearly paid off. Watcom SQL had the highest score in terms of response time but the lowest in terms of network-transfer time.

Microsoft SQL Server, which placed fifth, required a little tuning to optimize performance (unoptimized results, not shown here, were in excess of four hours). But in many cases, the company found that tuning for response time hurt the product's total time, and vice versa. Microsoft also found that a smaller packet size (512 bytes) improved many of the smaller queries but slowed queries returning a large number of rows (larger 4K packets improved those). While Microsoft would have preferred to tune for individual queries, our benchmark testing specification did not allow for this.

Informix OnLine pulled up the rear despite their read-ahead mechanism and cost-based optimizer. We also encountered an optimizer bug that caused a server crash on two of the queries (the database was not corrupted by the crash). The problem was in the Update Statistics command that placed an invalid number in the statistics page. The Informix consultant was able to patch the statistics page to work around the problem.

The **Load and Index** test measure how quickly the database system can import 18.11 million rows, and create 33 indexes. This test is of particular interest for judging products used to implement decision-support systems, where the database must be loaded and indexed on a regular basis. Load times for our BLOB table were not included in the load score. The raw data was provided to the vendors in key order.

Vendors were allowed to choose the structure of the indexes, although we specified the columns on which indexes had to be created. Because load-and-index is typically an isolated operation, we allowed vendors to tune specifically for this test, whereas we required them to run all other tests with a single preselected set of runtime values. All tables were loaded serially. It should be noted that real-world load-and-index times can be reduced by using multiple sessions.

All vendors loaded the database directly from the server, thus eliminating network bottlenecks and optimizing load rates. In addition, all vendors provided a mechanism to bypass the transaction log for better performance. All vendors except Watcom also provided a utility or used SQL extensions to perform the load. Watcom's ISQL utility does include a feature to load data but it is not an NLM implementation. To optimize performance, Watcom took advantage of the engine's NetWare interface to write a custom NLM load module. While most users would probably not do this, we felt that this approach might make sense in a decision-support environment. Watcom SQL is the only database in this story that can directly interface with another NLM.

Oracle7 demonstrated its ability to load and index very quickly. While the actual load times lagged behind XDB-Enterprise, Oracle7 quickly made up for lost time with its efficient indexing mechanism. XDB-Enterprise was second overall, and takes top honors in load speed. This may be attributable to Windows NT's asynchronous I/O capabilities and the multithreaded nature of the load utility. Sybase SQL Server placed third, an impressive achievement considering that it created a clustered index on all the tables. While Sybase SQL Server did not have to perform a sort on the data, it did have to move the data physically to put it in a clustered structure. Watcom SQL took fourth place but with the second-fastest index time. Informix OnLine placed fifth, and Microsoft SQL Server placed last. While Microsoft SQL Server did place fourth on the data load, the overhead of creating a clustered index pulled the package to the rear.

The **Export** test measures how fast a database can export a 1-million-row table into comma-delimited ASCII text format. The export was made to a local disk on the server to avoid network overhead. Interestingly, several of the vendors actually took longer to export the table than to load it. This may be due to the overhead of a binary-to-ASCII conversion, which is typically more expensive than ASCII-to-binary. Also, when loading data, a database can cache multiple rows and write them as a single block. Export operations are typically dependent on the operating system's file-system cache. For most users, data export times will not be a significant issue.

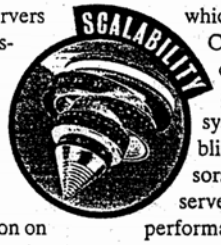
—Analysis written by Brian Butler

Intel-Based SMP: *How Strong?*

By Brian Butler and Thomas Mace

Even the fastest Intel-based servers may not be fast enough for massive client/server database applications. The classic gambit for the power-starved has been to forgo the Intel platform in favor of RISC-based symmetric multiprocessing (SMP) servers. (For information on comparative performance of Intel and RISC servers, see the sidebar "Competing with RISC.")

There is an alternative. The emerging class of Intel-based SMP servers delivers substantially improved performance over traditional single-CPU hardware. Scalability testing on the Intel-based Compaq ProLiant 4000, which can accept up to four CPUs on plug-in daughtercards, showed that with an appropriate operating system and database, doubling the number of processors and disk drives in the server can effectively double performance.

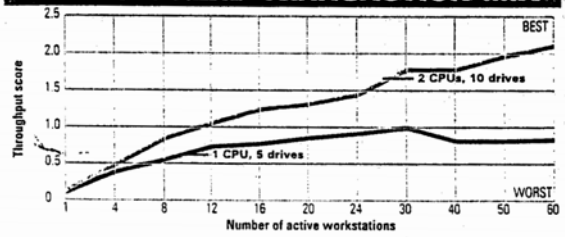


CLOCKING SMP ON INTEL

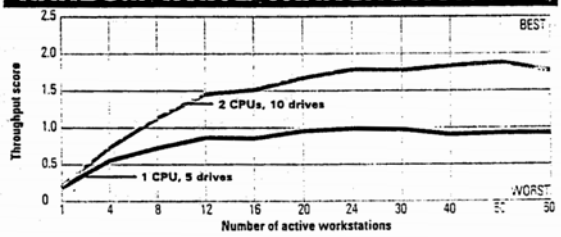
To investigate the benefits of Intel-based SMP database servers, we ran a series of scalability tests using Oracle7 Server for SCO Unix, Version 7.1. Under Oracle7, each client connection to the server is an independent process. The underlying SCO Unix operating system can distribute these processes symmetrically across multiple CPUs.

For testing, we used a subset of the AS^{3}AP database performance tests used for the main reviews. Results from the multiuser portion of the tests show throughput measured in transactions per second (tps). These results (see the accompanying graphs) are shown in normalized form based on the maximum throughput achieved by a single-CPU reference configuration. Query, Load, and Index are timed tests; the charts show the enhanced system's performance as a simple percentage of the reference system's

RANDOM READ TRANSACTION MIX



RANDOM WRITE TRANSACTION MIX



define conditions that can trigger an operating system script. You could use this feature to perform functions such as sending an administrator alert and initiating an automatic backup when a certain percentage of remaining log space is exceeded.

Other modifications to the database server and optimizer include a rewritten lock manager and loosened constraints on update-in-place. The optimizer can now use an available nonclustered index for queries containing an Order By clause. Microsoft has also implemented asynchronous checkpoints so that transactions can continue while a checkpoint is implemented. Dirty data pages are written to disk by a lazy-writer thread, reducing the overhead of the checkpoint operation.

While the server supports triggers and stored procedures, it also adds a powerful new feature called *extended stored procedures* aimed at leveraging workgroup technologies such as e-mail. Extended stored procedures are external Windows NT dynamic link libraries (DLLs) that can be dynamically loaded and executed on the server. For example, you could use this feature within a trigger to broadcast an e-mail message in response to a changed inventory level. Because every thread on the server is under structured exception handling, the server and database are protected from any errors arising from an extended stored procedure. Should a protection fault occur, only the thread would be terminated, not the process.

SUITABILITY TO TASK Microsoft SQL Server for Windows NT	
Production OLTP	EXCELLENT
Decision support	FAIR
Workgroup database	EXCELLENT
Connectivity & deployment	EXCELLENT

Although Microsoft has made improvements to Sybase SQL Server 4.2, it has not adopted some of the significant enhancements that Sybase introduced in its System 10. These include replication, declarative referential integrity, and ANSI cursors. Unlike Sybase System 10, Microsoft SQL Server cannot dump a single database to multiple backup devices.

Microsoft SQL Server also lacks transparent two-phase commit (this feature must be coded via the C interface), row-level locking, and built-in auditing. Remote procedure calls are outside of transaction management, a potential danger since consistency between remote databases cannot be physically guaranteed. While the Windows NT operating system is C2-level secure, the database provides only standard table-level security. Microsoft has committed to shipping a number of enhancements in future releases including declarative referential integrity, bidirectional scrollable cursors, parallel backup, and replication.

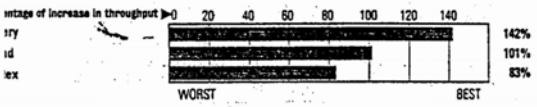
ire.
We began by establishing a refer-
ce score using the ProLiant 4000 in
standard test configuration for this
ry: an array of five 2.1GB hard disks
l a single 66-MHz Pentium CPU. In
s configuration, Oracle7's perfor-
nce is strongly I/O-bound so that
ıply adding a second CPU would
ve had little effect.
To scale performance while keeping
same balance between CPU and
k loads, we added five additional
ves and a second Pentium CPU and
an our tests. The results of the Ran-
m Read Transaction Mix test show
ıt throughput for the 2-CPU, 10-disk
tem was well over twice that of the
erence system. The Random Write
ansaction Mix test results show that
SMP system was just shy of twice as
t.
Version 7.1 of Oracle7 for SCO
ıix also supports parallelization of
ery, load, and index operations inter-
nally. The engine accomplishes this by dividing operations into separate tasks that are spread across multiple processors. Examples of operations that can be parallelized include table scans, joins, aggregations, and various sort operations.

To see how well the ProLiant 4000's SMP capabilities support these features, we selected queries from our standard Ad Hoc Query test. The results show better than a 140 percent improvement in query execution time on the double-CPU system. Not all queries benefit from parallelization, however. We tried several queries that do not perform table scans or large sorts and saw no performance improvement. The tests also show a doubling of load performance and a substantial improvement in indexing, in part because of the efficiency of parallel sorts.

While the lure of RISC-based servers remains strong, Intel-based database servers with SMP upgradability offer an alternative that is definitely worth investigating. □

QUERY, LOAD, AND INDEX

2 CPUs and 10 drives vs. 1 CPU and 5 drives



ICK TOOLS

ıe superb graphical administration tools
ındled with the server let the adminis-
ıtor manage the database, the operating
stem, and networking from a single lo-
tion. The SQL Object Manager is a
ck change-management application
at can be used to create stored proce-
ıres, triggers, tables, indexes, rules,
ews, and other database objects. It also
cludes a bulk-copy program that, unlike
e original command-line bulk copy pro-
am, provides postmortem information
r failed operations. The SQL Object
anager can also generate a transact
QL data definition language (DDL)
ript from existing database objects that
n be used to recreate a database on an-
her server or document an existing
ıtabase structure.
The SQL Administrator tool is target-
l at device and database management. You can use it to create databases, devices, and users and to implement security. The ISQL/Windows program provides a basic Interactive SQL (ISQL) server interface with the convenience of a few Windows navigation features. It also lets you create a *showplan*, a graphical display of the access plan for any given query, and displays I/O statistics graphically for tuning and optimization. Standard command-line ISQL is also provided.

All of the Microsoft tools can simultaneously connect to multiple databases, but they cannot administer multiple servers as a group. Like most competing toolsets, they also lack integration; you'll need to switch from one to the other depending on the task at hand. Microsoft plans to roll SQL Administrator and Object Manager into a single tool eventually. Future versions will also support OLE 2.0 drag-and-drop behavior and allow for remote management of server groups.

Although Microsoft SQL Server is positioned as a client/server computing solution for the masses, its overall performance—despite a few gaps—puts it in a league with the industry leaders. Its stability and strong administration tools are benefits in any applications. Except for the fading OS/2 release, Microsoft SQL Server is an NT-only solution and will ultimately be only as scalable as Windows NT itself. But if you are a believer in Windows NT, Microsoft SQL Server is a robust, well-oiled solution.

▶ FACT FILE

Microsoft SQL Server for Windows NT, Version 4.21

List price: Server software, one development system, 60 client connections, and client software: $8,690. **Requires:** Server: 386-based PC or better, 16MB RAM, 25MB hard disk space, Microsoft Windows NT 3.1 or later. DOS client: 286-based PC or better, 640K RAM, 1MB hard disk space. **In short:** Microsoft SQL Server offers a compelling combination of a powerful database engine, superb graphical administration tools, excellent connectivity features, and unmatched integration with the Windows NT operating system. Its performance goes beyond workgroup demands and puts it among the top products in our roundup. This product is a Windows NT–only solution, but if your organization can buy into a closed-shop strategy, the integration of server, operating system, and networking components is hard to beat.

Microsoft Corp., One Microsoft Way, Redmond, WA 98052; 800-426-9400, 206-882-8080; fax, 206-936-7329

482 on reader service card

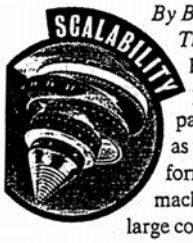
Oracle Corp.

EDITORS' CHOICE

Oracle7 Server for NetWare

Oracle7 Server for NetWare, Version 7.0.16, is a comprehensive, complex package that rolls together just about all the features you'll find in competing products. It is exceptionally fast, eminently stable, and very well suited to both multiuser and decision-support tasks. Oracle7 demands a sizable up-front investment and solid professional skills to get it up and running. But for mission-critical applications, especially in distributed environ-

Competing with RISC



By Brian Butler and Thomas Mace

From a PC-centric perspective, the Compaq ProLiant 4000 used as this story's test platform is a powerful machine. But large companies will naturally consider other hardware options when weighing a major client/server investment. This raises a basic question: How well does Intel hardware perform when compared with RISC?

We used ZD Labs' Ritesize IV test suite based on Sybase System 10 RDBMS to pit four RISC-based servers against a ProLiant 4000 equipped with two 66-MHz Pentium CPUs. The evaluated RISC systems were the Data General AViiON 8500 powered by six Motorola 88110 processors; the DEC 3000 Model 800S AXP Deskside Server with a single 200-MHz AlphaAXP 21064; the HP 9000 Series 800 Model G70 with two 96-MHz PA-7100 processors; and the IBM RISC System/6000 POWERserver 590 with a single 66-MHz IBM POWER2 CPU.

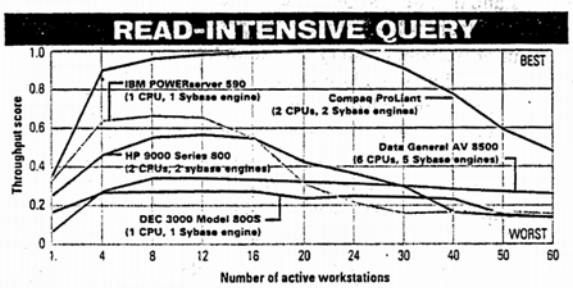
The Sybase System 10 database engine we used for testing the platforms supports SMP hardware by launching multiple instances of the database engine. The database server then binds the engines to a particular processor.

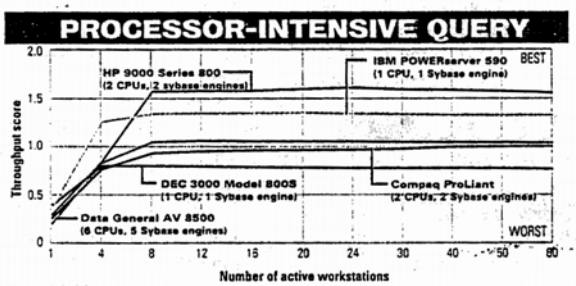
The 60-client test-bed was similar to the one used for review testing. In the accompanying graphs, the scores are shown in normalized form, with the maximum throughput achieved by the Compaq ProLiant in each test indicated as 1.0.

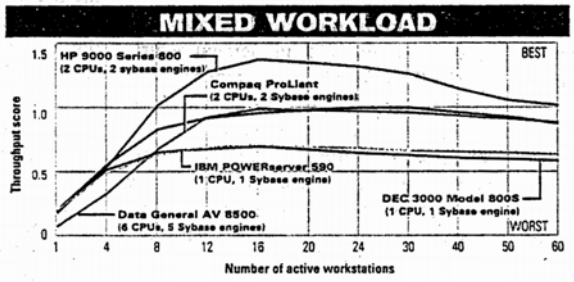
Our Processor-Intensive Query test selects a small number of cached rows. The Read-Intensive Query test performs a join on two tables that exceed the database cache size. The Mixed Workload test runs four transactions: the Processor-Intensive Query, the Read-Intensive Query, an Update transaction, and an Insert.

While the overall winner was the HP 9000, the Compaq ProLiant was a competitive midrange performer on both the Mixed Workload and Processor-Intensive Query tests. On the Read-Intensive Query test, it was fastest overall thanks to its strong disk-controller technology.

Examining performance is only part of the process of selecting a database server. Operating system maturity, hardware redundancy, service, and even intangibles such as the company's reputation will play a role. But in a straightforward speed comparison, Intel SMP hardware is clearly in the same league as the RISC-based heavyweights. □







ments, you'll be hard pressed to find a more robust solution.

Version 7.0.16 is little changed from the version we saw last year. But, even after a year's time, Oracle7 still looks extremely competitive. A newer release, Version 7.1, is already available on several platforms and is expected for NetWare by the end of this year. The same Oracle7 code base is currently available on about 80 hardware platforms, extending the product's reach from the Microsoft Windows desktop to the mainframe.

PERFORMANCE EDGE

While Oracle7 has adopted many of its competitors' best features, it owes some of its performance edge to technology that is not widely used by other products. For example, Oracle7 implements a multiversioning concurrency model, a unique feature in this roundup (Borland's upcoming InterBase Workgroup Server will offer a similar design).

In a multiversioning scenario, each transaction sees a consistent, unchanging view of the database precisely as it was when the transaction began. If the underlying data is changed by a later transaction, information from rollback segments

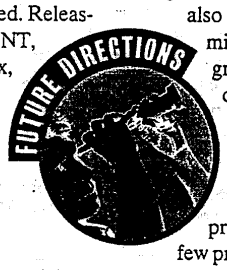
Ingres Server: *Still on Hold*

'y Brian Butler and Thomas Mace

.ast year's SQL roundup included a eview of Ingres Server for OS/2, Version 6.4, which at that time had just een acquired by The ASK Group. We ound this product to be intriguing but awed by serious bugs. This year we lanned a follow-up look at a point elease of Version 6.4 designed to ddress the problems we encountered. Vorking with technical representaves from The ASK Group, we put the pdated product through our standard ests in preparation for this story. ngres Server made it through our oad and Index and Ad Hoc query ests without a hitch but in our multiser tests, we ran into significant bugs nat made the product spontaneously rop clients. The performance numers we were able to generate put ngres Server at the bottom of the test neup.

In the middle of our tests, the ASK Group was acquired by Computer Associates International (CA), which mmediately withdrew all Intel-based ngres Server products from the market, ncluding the product we were testing. CA stated that the version we saw was a eta product not ready for release. Customers who received the product were told that they had received a beta version and that they would get the final release when it was ready.

CA plans to debug the existing Intel ports and release them once they are fixed. As we went to press, a ship schedule had not been announced. Releases for Microsoft Windows NT, NetWare, OS/2, SCO Unix, Solaris, and UnixWare are planned. CA also stated that it will continue to support the existing Ingres installed base.



TECHNOLOGY PIONEER

Ingres Server, which had its origins in the Berkeley Ingres prototype, has always had a reputation as one of the most academically strict relational databases. In the past, Ingres has served up impressive technology, pioneering cost-based optimization and many other now-standard database features, including triggers, which Ingres calls *rules*. Other high-end engine features of the version we tested include event alerters, user-defined data types, user-defined functions, stored procedures, and two-phase commit. Ingres also offers a distributed database strategy through its Ingres/Star server.

Ingres Server is unusual in that some of its most powerful features are managed by a server extension product called Ingres Knowledge Management. This component provides Ingres's rules and event alerters and also offers administration of permission levels by individual, group, or application. It also offers a resource-control feature based on the Ingres optimizer for preventing runaway queries.

Although the Ingres Server product we looked at suffered few protection-fault shutdowns during testing, most of the problems we encountered seemed to stem from the Ingres client libraries. CA agrees with our assessment and plans to concentrate its debugging efforts in this area. In the near term, CA plans to work on performance improvements within the existing engine and sees Ingres Server ultimately challenging Oracle and Sybase in mainstream OLTP markets. In the long term, they plan to rearchitect the database to support parallel operations and massively parallel hardware.

Ingres Server has clearly languished in the recent past, but its strong technology deserves a better fate. We look for future releases from CA to turn the product around. □

ed to maintain the first transaction's sistent view. The big advantage to a tiversioning model is that read transons do not need to acquire locks that k write transactions, improving overoncurrency.

When locks are needed, Oracle7 uses a -level locking scheme instead of the e conventional page-level locks. The base supports an unlimited number ow-level locks that never escalate to e or table locks. While the page-level ing schemes used in other products adequate for most applications (and oretically entail less management head), the page is not a "natural" unit orage. This can increase the difficulty of tuning operations. Row-level locking also provides optimal concurrency, indicated by Oracle7's excellent scores on our Random Write Transaction Mix test.

Oracle7's triggers are similar to those in other products except that the user can stipulate when a trigger executes relative to execution of the SQL statement that fires it. The database also supports declarative referential integrity, and automatic cascading deletes can be set to eliminate child rows when parent rows are deleted.

Since Version 6.0, the log manager has been optimized to support fast commits and group commits. Moreover, only changes to data are logged, not the entire before-and-after image of the row.

Oracle7's cost-based optimizer does not use histograms, but it does gather a number of statistics from tables and indexes. Using the Analyze command, you can update these statistics based on a subset of the data. This can be useful for huge decision-support databases where a complete table scan would be unduly long.

On our Ad Hoc Query test, the optimizer picked the wrong access method on our sort query. This resulted in a

SUITABILITY TO TASK	Oracle7 Server for NetWare
Production OLTP	EXCELLENT
Decision support	EXCELLENT
Workgroup database	GOOD
Connectivity & deployment	EXCELLENT

The Price of Performance

By Brian Butler and Thomas Mace

If all of the SQL database servers in this roundup cost the same, picking out the right one would only be a matter of weighing feature sets. But the wide range of prices here—from a mere $5,390 to almost $30,000—adds to the complexity of your decision. In order to help clarify pricing issues, we look at SQL database prices in two different ways: total cost and bang for the buck.

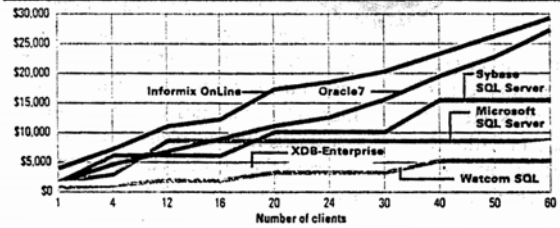
One good piece of news is that SQL database pricing has gotten noticeably simpler. SQL vendors used to be notorious for devising complex schemes with separate pricing for server connections and client libraries. This year, all with the exception of Informix are pricing on a per-user basis, where client software is now essentially free. Workgroup bundles from the major players will soon simplify pricing even further.

STRAIGHT COST

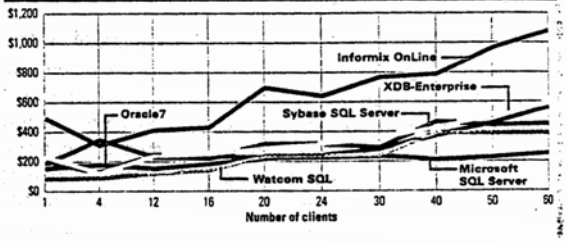
The simplest way to view the price of a database is by its straight deployment cost (see the chart "Price by Number of Clients"). This cost, shown for 1 to 60 users, includes the required number of user licenses, required client software, and one standard 3GL development kit (the cost of network protocol stacks and support is not included).

Prices vary widely and exhibit a

PRICE BY NUMBER OF CLIENTS



PRICE PER TRANSACTION



10-hour run on a test that ultimately took only one hour and 20 minutes to complete. The problem was fixed using Oracle's well-documented Hint mechanism for overriding the optimizer. A nice Hint subtlety is that the mechanism lets you tune queries for best response time (the amount of time required to return the first row of data) or best overall query time.

Stored procedures are available, although they cannot return result sets. But you can send an array of values to a stored procedure. This elegant trick could be used for problems such as inserting multiple line entries in an order table. Stored procedures can be logically grouped together into what Oracle calls a Package. This makes for easier administration, allowing the user to maintain all the stored procedures for a particular application as a single entity, for example. You can define global variables for an entire Package and also grant and revoke permissions at the Package level.

The current release of Oracle7 still lacks the GUI administration tools that recently premiered on Oracle's Workgroup Server product. It does ship with a full-featured character-based administration tool called SQL*DBA. This utility manages tasks such as opening and closing the server and lets you monitor system performance and use, perform backups, and interactively execute SQL statements. SQL*DBA provides two modes of operation: a screen-based interface complete with drop-down menus, and a command-line interface. You can also use SQL*DBA to monitor a variety of system statistics and to create and drop tablespaces and rollback segments.

The current release offers the convenience of role-based security administration. Users can be added to more than one role, and user privileges can be granted or revoked at the role level as needed. While this seems like a simple concept, it marks a big improvement over previous versions of the product in which privileges had to be granted individually for each user. The current version is B2-level secure. A separate product, Trusted Oracle7, is C2-level secure.

Oracle7 has strong support for distributed databases. It supports transparent two-phase commit and controls remote procedure calls (RPCs) as an integral part of transactions. Oracle7

▶ FACT FILE

Oracle7 Server for NetWare, Version 7.0.16

List price: Server software, one development system, 60 client connections, and client software: $27,400. **Requires:** Server: 386-based PC or better, 12MB RAM, 30MB hard disk space, NetWare 3.0 or later. DOS client: 286-based PC or better, 100K RAM, 100K hard disk space. **In short:** Excellent transaction processing speed and a rich feature set add up to one of the most sophisticated databases available. Oracle7 offers virtually all the features of competing products, and its multiversioning consistency model and row-level locking provide excellent concurrency. Despite its high price, Oracle7 still garners an excellent price/performance ratio, but it is not geared toward organizations with limited budgets. This is not a database for the meek, but for the most demanding applications, you'd be hard-pressed to find a better solution.

Oracle Corp., 500 Oracle Pkwy., Redwood Shores, CA 94065; 800-672-2531, 415-506-7000; fax, 415-506-7200

478 on reader service card

number of strategies. Informix On-Line is consistently the most expensive; its price increases in a smooth curve from $3,995 for one user up to $29,395 for 60 users. Oracle7, the second most expensive package, takes a similar approach to pricing. Sybase SQL Server is less expensive and shows a simpler tiered pricing structure. Microsoft SQL Server carries simplification even further, providing for 12 to 60 clients for the same price of $8,690. Watcom SQL, the least expensive package we tested, begins at $790 for one user and rises to a modest $5,390 for 60 users. XDB-Enterprise, while slightly more expensive, closely follows Watcom SQL's pricing.

PRICE/PERFORMANCE

A price/performance analysis shows the products in a very different light (see the chart "Price per Transaction"). To generate this graph, we divided each price by each product's transaction-per-second throughput on our Random Write Transaction Mix test.

The products look much more alike from this perspective. Oracle7, Sybase SQL Server, Watcom SQL, and XDB-Enterprise deliver very similar price/performance ratios across the range of client loads. Microsoft SQL Server is the leader above 20 clients, albeit by a narrow margin. The only standout is Informix OnLine, which offers the poorest mix of price and performance across the board. As with most product groups offering similar bang for the buck, your choice here will be dictated by the performance level you need.

As you calculate server prices, it's important to remember that servers are only a part of the equation. Client/server technology remains an expensive proposition to implement successfully due to the lack of turnkey systems, and by the time that you've factored in software development and support costs, the price of the most expensive server may not look so big. □

also initiates two-phase commit for any RPC outside the Oracle7 since the database has no way of knowing what the result of the RPC will be. Cost-based optimization is available for distributed queries based on statistics and available indexes in the distributed environment. Oracle7 also has a trigger-based replication scheme for making distributed read-only copies of tables, table subsets, or query results. Oracle has announced a more robust symmetric replication technology, which is expected to ship by the end of this year.

Access to non-Oracle7 data through RPCs or SQL is provided by the Oracle Transparent Gateway (formerly SQL* Connect), a set of gateway products for a variety of relational and nonrelational systems.

Its first-place Load and Index test scores are attributable in part to its SQL*Loader, one of the fastest, most functional loaders we used. It supports both direct-path and conventional-path loading. We used direct-path, in which records are written directly to the database block, bypassing most database processing. While direct-path loading has some restrictions, conventional-path loading is always available as a workaround.

The next version of the NetWare product should ship as Version 7.1 before the end of the year. This revision will add Oracle7 Symmetric Replication, and the impressive parallel query–execution, data-loading, and index-creation features already available in the Unix release we used to test CPU scaling (see the sidebar "Intel-based SMP: How Strong?"). Version 7.1 will also include support for user-defined SQL functions and dynamic SQL statements—statements whose contents are not known until runtime. Additional slated improvements include tweaks to the optimizer, encrypted network passwords, and faster database recovery.

NOT FOR THE MEEK

Oracle7 was the fastest database tested for this roundup, taking the lead on five out of seven tests. Almost no tuning was required to achieve these results.

At the same time, this is not a database for the meek. Exploiting its huge array of features demands expertise and time spent with the superb encyclopedic documentation. Oracle7's price/performance ratio is excellent, but its high price is targeted at users who need speed and functionality, not savings. But for bullet-proof operation in high-stress transaction-processing environments, Oracle7 is a winner.

Sybase Inc.

• Sybase SQL Server for NetWare

Sybase SQL Server for NetWare, Version 10.01, represents a solid upgrade to Version 4.2, which we reviewed last year. Sybase has made numerous enhancements, tweaks, and fixes to the engine, which proved to be stable and extremely fast in testing.

The jump from Release 4.2 to 10.01 brings Sybase SQL Server's version numbering in line with the company's System 10, an important family of add-on server products designed to address connectivity, replication, administration, and scalability.

NetWare makes an excellent showcase for the database engine's core features and performance, but a lack of add-ons leaves Sybase SQL Server in limbo as a product: While many System 10 components are shipping on other platforms, the only component available for the NetWare product we tested is the Backup Server. Our testing bundle only included the server, Backup Server, bulk copy program, and Interactive SQL (ISQL). Later this fall, Sybase plans to sweeten the offering by releasing NLM versions of some other System 10 components and repackaging the NLM server in separately priced workgroup and enterprise editions.

SUITABILITY TO TASK Sybase SQL Server for NetWare	
Production OLTP	EXCELLENT
Decision support	EXCELLENT
Workgroup database	GOOD
Connectivity & deployment	EXCELLENT

MATURE ENGINE

The NLM version of Sybase SQL Server we tested includes much of the advanced engine technology that first lifted Sybase to prominence. The list includes Sybase SQL Server's stored procedures (which

▶ SUMMARY OF FEATURES CONTINUES

SQL Databases

PC MAGAZINE

Products listed in alphabetical order
■ = YES □ = NO

	Informix OnLine for SCO Unix 5.01	Microsoft SQL Server for Windows NT 4.21	Oracle7 Server for NetWare 7.0.16	Sybase SQL Server for NetWare 10.1	Watcom SQL Network Server for NetWare 3.2	XDB-Enterprise Server 4 for Windows NT
List price*	$29,395	$8,690	$27,400	$15,590	$5,390	$8,490
Cost of standard one-year telephone support	$1,980	$7,500	$1,290	$4,000	$5,000	$2,000
SQL Implementation						
ANSI compatibility:						
ANSI Level 2 with Enhanced Integrity	■	□	■	■	■	■
ANSI Level 2	■	□	□	■	■	■
ANSI Level 1	■	■	□	□	■	■
Full DB2 compatibility	□	□	□	□	□	■
Binary large object (BLOB) data types	■	■	■	■	■	■
User-defined data types	□	■	□	■	□	□
User-defined range limits on data types	□	■	□	■	□	■
Advanced mathematical and statistical functions	■	■	□	■	□	■
User-defined functions and operators	■	□	□	■	□	□
Cost-based/rule-based optimization	■□	■□	■■	■■	■■	■□
Transaction Management						
Locking:						
Record-level	■	□	■	□	■	■
Page-level	■	■	□	■	□	□
Table-level	■	■	■	■	□	■
Adjustable for each table	■	□	□	■	□	□
Automatic lock escalation	□	■	□	■	□	□
Consistency levels supported:						
Cursor stability	■	■	N/A❶	■	■	■
Repeatable reads	■	■	■	■	■	■
Multiversioning	□	□	■	□	□	□
Release locks	■	■	□	□	■	■
Uncommitted reads	■	□	□	□	■	■
Read-only databases	■	■	□	■	■	■
Cost-based deadlock-detection schemes:						
Engine can abort transaction causing the deadlock	■	■	■	■	■	■
Engine can abort via a timeout option	■	■	□	■	□	■
Programming Interface						
Includes call-level interface	■	■	■	■	■	■
ODBC support included	■	■	■	■	■	■
Host-language interface:						
ANSI-compatible cursors	■	□	■	■	■	■
Included SQL precompilers	C	C, COBOL	C, COBOL, FORTRAN	ADA, C, COBOL	C, C++	C, COBOL, DL/T
Backward scrolling in result set	■	■	□	□	■	■
Preserves cursor context after Commit and Rollback	■	■	□	□	■	■
Supports result-set inserts	■	■	□	■	■	□
User can insert, update, and delete using an array of variables	■	□	■	■	□	□
Stores procedures in database	■	■	■	■	□	□
Embedded select, update, delete, insert	■	■	■	■	N/A❷	N/A❷
Supports control and flow logic	■	■	■	■	N/A❷	N/A❷
Supports message and error-code handling	■	■	■	■	N/A❷	N/A❷
Accepts variables and returns values or messages	■	■	■	■	N/A❷	N/A❷
Supports row-at-a-time processing	■	■	■	■	N/A❷	N/A❷
Supports remote stored procedures	■	■	■	■	N/A❷	N/A❷
Performs binding and optimization before runtime	■	■	□	□	■	■
Offers Wait and Nowait for lock to be released	■	□	■	□	■	□
Database Server Environment						
Database server architecture	Multithreaded	Multithreaded	Process-per-user	Multithreaded	Multithreaded	Multithreaded
Portability:						
DOS	□	□	□	□	■	■
Microsoft Windows	□	□	■	□	■	■
Microsoft Windows NT	□	■	■	■	□	■
NetWare	□	□	■	■	■	□
OS/2	□	■	■	■	■	■
Unix	■	□	■	■	□	□
VM	□	□	■	□	□	□
VMS	□	□	■	■	□	□
MVS	□	□	■	□	□	□

*The list price includes server software, 60 client connections with client software, and 1 development system.
N/A ❶—Not applicable: The product implements a multiversioning concurrency model.
N/A❷—Not applicable: The product does not have this feature.

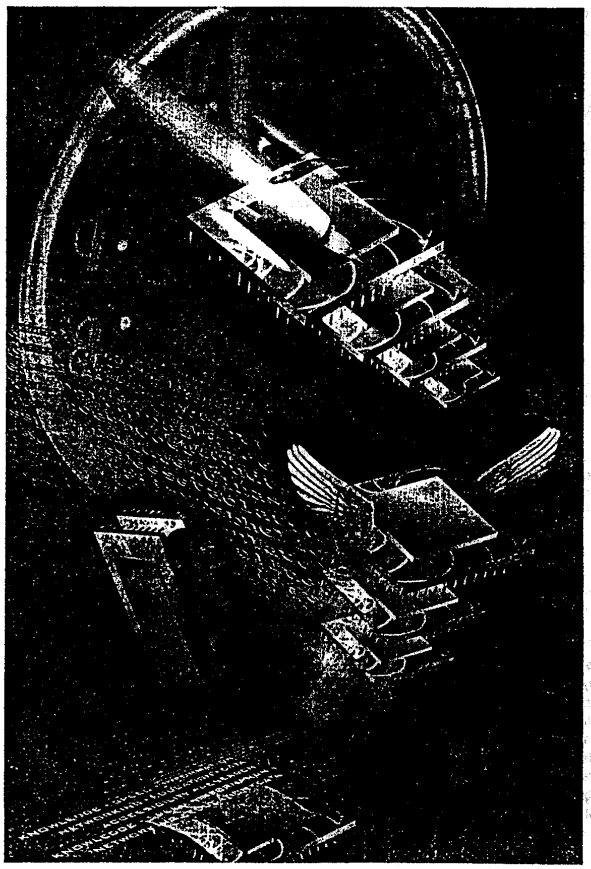
PRODUCT COMPARISON

AI

Symmetric multiprocessing servers

Scaling the performance wall

High-performance Pentium-based multiprocessors are catching up to RISC systems, giving multiprocessing network OSes, such as Windows NT, a foothold in what was formerly the sole domain of Unix.



COMPARED

Revolution Q-4SMP
Advanced Logic Research Inc.

Manhattan P5090
AST Research Inc.

ProLiant 2000 5/90
Compaq Computer Corp.

PowerEdge XE 590-2
Dell Computer Corp.

Poly 500EP2
Polywell Computers Inc.

If wimpy, single-processor performance has you climbing the walls, quit climbing and start scaling up with symmetric multiprocessing (SMP) systems. These Pentium-based machines are finally approaching the scalability and performance capabilities of RISC-based systems. And with multiprocessing network operating systems, such as Microsoft Corp.'s maturing Windows NT 3.5 and Novell Inc.'s upcoming NetWare MP, you can take advantage of this new processing power without having to switch to Unix.

Symmetric multiprocessing lets multiple CPUs share a server's memory, interrupts, and devices through a run-time algorithm. How much this boosts performance depends on the application, but you're likely to see at least some improvement across the board. SMP systems probably appeal the most to two groups — those downsizing mainframe applications to client/server systems and those who need to boost an already heavily loaded server. According to our survey of 1,000 *InfoWorld* readers, more than 80 percent of those who use SMP servers use them with a database engine, such as Microsoft's SQL Server or Oracle Corp.'s Oracle7.

Using Windows NT 3.5 as our NOS, we measured how much scalability the five Pentium-based SMP servers in this comparison provided by testing them with one processor and then two. The good news: If your network handles mostly CPU-bound applications, such as online transaction processing (OLTP), these servers offer a way up and out of the performance hole. Advanced Logic Research Inc.'s Revolution Q-4SMP and Compaq Computer Corp.'s ProLiant 2000 5/90 were the most scalable servers by far, performing almost twice as fast on two processors than on one. The Revolution was the upset winner of our speed tests, outperforming even the venerable ProLiant's multiprocessing server by nearly 20 percent in OLTP.

We chose Windows NT 3.5 as the multiprocessing NOS for our benchmark tests because of its focus on scalability. Most of the more than half of our readers using SMP servers are using two processors, but readers projected they might use as many as six processors per server. Because NT 3.5 supports as many as four processors right out of the box, it fit well with our readers' needs. If you need to harness the power of more processors, you can buy NT from a vendor like Sequent Computer Systems Inc., in Beaverton, Ore., which provides NT support for as many as eight processors. In this comparison, only the ProLiant and the Revolution were capable of using more than two processors. The Revolution can use as many as four 100-MHz Pentium chips, and the ProLiant can use as many as four 90-MHz Pentiums.

SCALABILITY IS REAL. Perfect scalability is a 100 percent performance increase between one and two processors. For example, a server with perfect scalability ran 50 transactions on a single processor in 1 hour, it would complete those transactions in 30 minutes using two processors.

None of the servers scaled perfectly, but the Revolution and ProLiant did quite well. The Revolution ran slightly more than 92 percent faster on two processors than one, and the ProLiant ran more than 97 percent faster.

Even at these speeds, these servers still ran about 25 percent slower than a MIPS Technologies Inc.-based machine running Windows NT 3.5, which we used as a point of comparison. (The NEC Technologies Inc. RISCserver 2200 was not yet shipping when we tested; see story, page 89.) But for the first time, Intel-based systems come close to RISC, and that's big news. It means that at least for a while, IS managers can cope with more demanding processing needs by simply moving their applications to Intel machines with more processors. The bad news: If you simply must have that remaining 25 percent speed increase, you'll have to port to a MIPS machine. And even though it runs NT, you'll have to port all of your data, not to mention buy new versions of your applications.

NOT A PANACEA. If you're dealing with I/O-bound applications — such as printing, file transfer, and, to a lesser extent, decision support — an additional processor won't help much. On such applications, we found only minor improvements, usually in the neighborhood of 15 percent. Traveling over the network appears to put a heavy dent in the effect extra processors have in such environments. Expecting them to make a difference would be tantamount to buying a Corvette and expecting it to make rush-hour traffic go away.

Computer manufacturers are well aware of this phenomenon; that's why they like to benchmark their multiprocessing systems on CPU-intensive activities such as database transactions — not on file and print services, which are much more I/O dependent.

Our readers were more concerned about transaction speed than about the more I/O-dependent transactions, according to our survey, so we tested accordingly. As expected, the servers' performance scaled much better in our OLTP test, which we designed to be CPU intensive (see "How we tested," page 85), than in responding to queries, a more I/O-intensive task.

While mulling over the lack of scalability in decision-support operations, we discovered a huge variance in the performance of each server's disk I/O subsystem — which ultimately determines the performance of your server.

Although we did not base any of our scores solely on these disk I/O results, you'll want to pay close attention to them (see chart, page 90) if your database servers perform both OLTP and decision-support operations (such as database queries) on a regular basis.

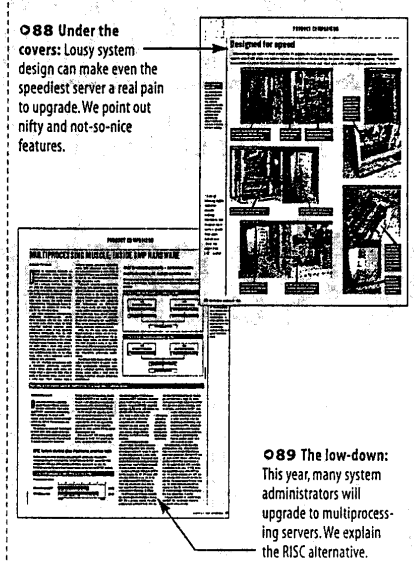
FOUR-PROCESSOR SUPPORT ON SQL SERVER 4.21A? NOT. Our testing turned up some other interesting results. With our scalability testing for two processors completed, we thought we'd fire up a few extra processors and see what even more could do. Using SQL Server 4.21a and Windows NT 3.5 on the Compaq ProLiant 4000 5/66, we tested three and then four processors. To our astonishment, three processors gave us virtually the same performance as two, and using four processors resulted in the same speeds we would have expected with three processors.

We rang Microsoft. It turns out that if you use SQL Server 4.21a's SMPStat parameter — not recommended by Microsoft — you tweak an internal parameter of the software, which in turn tells the software to use all available processors. If you don't, SQL Server treats a 4-processor system as a 2-processor system and a 4-processor system as if it uses only three. We didn't have any problems in our admittedly short tests, but Microsoft warns that using SMPStat could result in a deadly embrace, locking the database. Therefore, it doesn't support the use of SMPStat and won't help you out of any problems it causes. The upcoming Version 6.0 of SQL Server eliminates this quandary by using SMPStat as the default.

We had planned to include a Digital Equipment Corp. Intel-based SMP server (which it sells in addition to its own Alpha chip), but Digital was unable to provide us with one of its machines due to production schedule conflicts.

Also looming large in the SMP server race are multiprocessing systems based on the PowerPC chip. By June, sources expect versions of Windows NT and OS/2 to ship for PowerPC hardware, and both OSes will support SMP out of the box. This month, Microsoft formally released the beta version of Windows NT for the PowerPC, initially running on a Motorola system. (See "Playing with NT on PowerPC promises good times ahead for all users," March 13, page 108.) Regardless of which operating system proves most popular, Intel is feeling the heat from the PowerPC. We'll review the PowerPC servers as they ship.

A guide to this comparison



088 Under the covers: Lousy system design can make even the speediest server a real pain to upgrade. We point out nifty and not-so-nice features.

089 The low-down: This year, many system administrators will upgrade to multiprocessing servers. We explain the RISC alternative.

Contents

Results at a glance

If you can afford it, ask for a server with a dedicated Level 2 cache memory for each processor. The winner of our comparison, Advanced Logic Research Inc.'s **Revolution Q-4SMP**, was one of only two dedicated-cache Pentium systems. The super-fast RISC-based server we tested (see story, page 89) also uses dedicated caches.

There's a lot to like about the Revolution. It won both our online transaction processing (OLTP) and decision-support tests, showing its CPU performance and disk subsystem to be the best of any server we tested. It was the only machine capable of upgrading to four 100-MHz CPUs, aided by ALR's easily-plugged-in CPU boards. Its seven fans should keep the system plenty cool.

The Revolution didn't win in all categories, though. It's not as scalable as Compaq's ProLiant, and we were annoyed by the flimsy construction of its door. ALR was one of only two vendors that did not offer around-the-clock telephone support. The server comes with a five-year warranty, however.

Compaq Computer Corp.'s **ProLiant 2000 5/90** was the only other server capable of using four Pentiums (it uses 90-MHz chips). But you'll have to move the redundant array of independent disks (RAID) to an external drive cage in order to upgrade. The ProLiant is the most scalable of these servers — it showed a near-perfect 97 percent jump in OLTP performance when we added a second CPU. We also liked the system's handy SmartStart CD-ROM, which offered a choice of several network operating systems.

The ProLiant was the least expandable; it could only hold 10 gigabytes (GB) worth of additional hard drives without external drive cages.

Among systems designed for only two CPUs, the AST Research Inc. **Manhattan P5090** scored the highest, primarily because of its ease of use, including graphical utilities and some dealer preconfiguration. Only the Manhattan got an excellent rating for documentation. It also has two PCI slots. But the Manhattan was the only system with 256KB cache, not 512KB. This may be why it landed in last place in our OLTP comparison. Scalability and decision-support scores weren't impressive either.

The Score

7.3 ALR Revolution Q-4SMP

6.7 Compaq ProLiant 2000 5/90

6.3 AST Manhattan P5090

6.2 Polywell Poly 500EP2

5.8 Dell PowerEdge XE 590-2

The best thing about the Polywell Computers Inc. **Poly 500EP2** is its price. At $14,475, it's more than $7,000 less expensive than any other server in the comparison. Otherwise, the Poly is an average machine, with average performance numbers (it's a respectable transaction processor but had the worst scalability of the bunch) and terrible technical support. Let us say that again: Polywell's technical support was not only the worst in this comparison, it was the worst we've encountered in a while. Representatives were rude, practically hanging up on us even though we called during scheduled support hours.

Dell Computer Corp.'s **PowerEdge XE 590-2** is a mediocre performer. It was the slowest of all the servers at returning query results and performed almost as poorly in our transaction-processing tests. Compared to the other systems, its scalability was unimpressive.

We weren't particularly pleased with some aspects of the system's design, either. To gain access to the memory, you'll need a screwdriver and patience. And be careful where you leave the PowerEdge, because it doesn't have a lock or even a cover for the power switch. On the plus side, once we removed the case, we could easily swap cards and drives in and out without any problems. The PowerEdge has as many EISA slots as the ProLiant and supports two PCI slots as well. It's capable of storing as much as 24GB of data.

RELATED ARTICLES

March 6, page 37
IBM revamps server line
IBM prepares an entry-level dual-processor 90-MHz Pentium LAN server to challenge Compaq's dominant market share.

Feb. 27, page 66
No-fault insurance
We examine four RAID subsystems and tell you how to choose the right RAID.

Jan. 30, page 6
Novell SMP delayed until middle of year
Originally promised in 1989, the NetWare kernel is now due for SMP support this summer.

Dec. 19, 1994, page 1
NOS news is good news
We looked at the areas where SMP hardware helped NOS performance — and where it didn't.

CONTRIBUTORS

Introduction by **Lisa Stapleton**
Senior Editor, Enterprise Team
and **Laura Wonnacott**
Test Developer

Written by **Scott Mace**
Senior Editor, LAN Team
and **Ayse Sercan**
Assistant Editor

Tests developed by **Laura Wonnacott**

Testing by **Jeff Symoens**
and **Rod Chapin**
Technical Analysts

Edited by **Scott Mace**
and **Ayse Sercan**

Report Card

Symmetric multiprocessing servers

GUIDE

& Rating

Score in points

InfoWorld reviews only finished, production versions of products, never beta-test versions.

Products receive ratings ranging from unacceptable to excellent in various categories. Scores are derived by multiplying the weighting of each criterion by its rating, where:
Excellent = 1.0 - Outstanding in all areas.
Very Good = 0.75 - Meets all essential criteria and offers significant advantages.
Good = 0.625 - Meets essential criteria and includes some special features.
Satisfactory = 0.5 - Meets essential criteria.
Poor = 0.25 - Falls short in essential areas.
Unacceptable or N/A = 0.0 - Fails to meet minimum standards or lacks this feature.

Scores are summed, divided by 100, and rounded down to one decimal place to yield the final score out of a maximum possible score of 10 (plus bonus). Products rated within 0.2 points of one another differ little. Weightings represent average relative importance to *InfoWorld* readers involved in purchasing and using that product category. You can customize the Report Card to your company's needs by using your own weightings to calculate the final score.

The Test Center Hot Pick is *InfoWorld*'s new award for outstanding products we have evaluated in scored stand-alone reviews or product comparisons. To receive the Test Center Hot Pick seal, a product has to offer what *InfoWorld* deems to be a standout feature or technology that is unusually valuable or revolutionary compared to competitors. The product must also score at least satisfactory in all Report Card categories and receive a final score of 7.0 or more.

	Weighting	ALR Revolution Q-4SMP	AST Manhattan P5090	Compaq ProLiant 2000 5/90
		Advanced Logic Research Inc. Irvine, Calif. (800) 444-4257 or (714) 581-6770; E-mail: sales@alr.com	AST Research Inc. Irvine, Calif. (800) 876-4278 or (714) 727-4141; E-mail via CompuServe: GO ASTSUPPORT	Compaq Computer Corp. Houston, Texas (800) 345-1518 or (713) 370-0670; E-mail via World Wide Web: http://www.compaq/com/
Performance				
Scalability	150	& *Very Good* ● 112.50 92.85 percent gave the Revolution second place.	& *Good* ● 93.75 The Manhattan's 81.60 percent put it in third place.	& *Excellent* ● 150.00 97.34 percent won top honors.
Transaction processing	200	& *Good* ● 125.00 39.29 transactions per minute (tpm) made the Revolution a top on-line transaction processing (OLTP) performer.	& *Satisfactory* ● 100.00 28.88 tpm was the best the Manhattan could do for last place.	& *Satisfactory* ● 100.00 32.80 tpm put the ProLiant in second place.
Decision support (Times in hours:minutes:seconds)	125	& *Very Good* ● 93.75 By finishing in 1:39:58, the Revolution picked up another first place.	& *Satisfactory* ● 62.50 The Manhattan was in fourth place at 2:19:09.	& *Satisfactory* ● 62.50 2:15:05 earned the ProLiant third place.
Setup and ease of use	100	& *Good* ● 100.00 The included EISA utility is simple to use, but the disk configuration utility is more cumbersome. The two are not integrated and the disk can boot the system. The steps for creating a redundant array of independent disks (RAID) are not immediately apparent. The system does come preconfigured, though.	& *Very Good* ● 75.00 AST doesn't preconfigure the system, but some resellers may. The disk configuration utility is GUI-based and requires a mouse, so the current drive configuration is represented in a nice chart. The Manhattan's configuration utilities are not integrated. The system could not boot up from the disk.	& *Very Good* ● 75.00 The CD-ROM-based SmartStart system configuration utility is fully integrated, although it does not come pre-installed. SmartStart includes all supported operating systems with license activation codes for those purchased. The ProLiant system can boot from the CD-ROM and install the OS preconfigured to Compaq's specifications.
Expandability	100	& *Very Good* ● 75.00 The Revolution was the only server we tested that could upgrade to four 100-MHz processors. The 32-bit system memory can be increased to a whopping 1 gigabyte (GB) and the video RAM (VRAM) to 2MB. Three of its 10 EISA slots have VESA local bus extensions, but the system lacks PCI slots. The server supports as much as 22GB using half-height drives or 30GB using hot-swappable drives.	& *Good* ● 62.50 The Manhattan holds a maximum of two 90-MHz Pentium processors. It was the only system to offer only 256KB of cache RAM, instead of 512KB. It can hold 256MB of 32-bit system memory and 512KB of video memory. It has six EISA and two PCI slots (one slot is shared), but one PCI slot is used by the disk array controller that comes standard. The server supports as much as 32GB of storage (hot-swappable).	& *Very Good* ● 75.00 The ProLiant can hold four 90-MHz Pentiums, although going from two to four CPUs required that we move the RAID controllers used in our configuration to an optional ProLiant Storage System. The ProLiant can hold as much as 512MB of 32-bit system memory and 1MB of video memory. It has eight EISA slots but no PCI slots. The server only supports as much as 10GB of disk storage.
System design	75	& *Good* ● 46.88 This system had many of the features we found important, especially a dedicated cache for each CPU. It can use standard memory and includes error detection and correction. Its score was hurt by poorly designed doors and sticky hot-swappable drives. There are two conjunctive power supplies (redundant power supplies are optional) and seven fans. Gaining access to the memory is awkward. Access to drive bays is on the opposite side of the system board.	& *Good* ● 46.88 The Manhattan's design was average, with a shared cache and a single locking door, which was somewhat hard to open and close. The first 16MB of memory features error-correcting code (ECC); AST provides ECC in other components of the system, including cabling. When the single door is closed and locked, access to the side panel screws is restricted. The system has one 300-watt power supply with its own fan; each CPU has a heat sink with a fan; and there is a central cooling fan.	& *Good* ● 46.88 The ProLiant server was the only system besides the Revolution with a dedicated cache for each CPU. The first 32MB of system memory includes ECC. The locking front door swings open and can be detached from the system. The case is designed so the power cannot be on when the front door is open. The side slides off for access to the processors and slots. The system has one power supply, four cooling fans, and heat sinks on the processors. The drives have relatively easy access and glide smoothly in and out of the array chassis. The bus slots and SIMM banks are easy to reach.
Support and pricing				
Documentation	50	& *Very Good* ● 37.50 The Revolution's documentation is well laid out and easy to read, but it's missing some information. The manual has many diagrams and a nice troubleshooting guide.	& *Excellent* ● 50.00 AST's documentation is easy to read and well laid out. It includes an extensive troubleshooting section with a list of error codes, plus a section on system disassembly.	& *Very Good* ● 37.50 Compaq's written documentation is excellent. Although it didn't contain illustrations, step-by step instructions, or diagrams, it earned extra points for its on-line reference library.
Support policies	50	& *Excellent* ● 50.00 ALR was one of only two vendors without 24-hour, 7-day telephone support. A five-year warranty helps. ALR does offer support on the World Wide Web.	& *Excellent* ● 50.00 AST offers a three-year warranty, free on-site support, and an optional 4-hour response by Memorex Telex. Unlimited free support is available 24 hours a day, 7 days a week through the toll-free line, or via fax, private BBS, CompuServe, and Prodigy.	& *Excellent* ● 50.00 Unlimited free technical support is available. Compaq also offers a three-year, next-day on-site parts and labor warranty, with an optional 4-hour response. Support is available via on-line services. Under the Prefailure Warranty, Compaq will replace any part operating below par before it actually fails.
Technical support	50	& *Satisfactory* ● 25.00 The technical support staff was always available, with minimal hold times, but not all staff members would answer our questions. We sometimes had to call back and ask for someone with specific expertise.	& *Good* ● 31.25 Getting through to technical support was the only pitfall. We had busy signals on several calls and an average 3-minute hold time. Once we got through, the support staff was very pleasant, knowledgeable, and willing to help.	& *Satisfactory* ● 25.00 We waited several minutes on hold for a technical support person. The quality of support varied. Although some technicians were more knowledgeable than others, they all solved our problems. Some were patient and helped us through willingly; others were more abrupt.
Price	100	& *Good* ● 62.50 The price was $22,351 configured as tested with two 100-MHz CPUs. (A base system with one 100-MHz CPU, a 256KB cache, and 16MB of RAM costs $6,995.)	& *Good* ● 62.50 The price was $22,620 configured as tested. (A base system with one 90-MHz CPU, a 256KB cache, 16MB of RAM, and two 1GB drives costs $8,976.)	& *Satisfactory* ● 50.00 The price was $26,999 configured as tested. (Compaq builds to order and had no base price.)
Final score		**7.3**	**6.3**	**6.7**

SYMMETRIC MULTIPROCESSING SERVERS

HOW WE TESTED

TO TEST THE FIVE symmetric multiprocessing (SMP) servers in this comparison, we designed benchmark tests that would play to their primary strength — handling CPU-intensive work. We knew from prior testing that SMP servers could slow performance on a network that provides mostly file and print services. (See "Symmetric multiprocessing may not always boost performance," Dec. 19, 1994, page 77.) Additional research, based on discussions with vendors and results from our reader survey, confirmed that these SMPs are most effective when performing CPU-heavy processing chores such as transactions, CAD or CAM work, and statistical analysis.

We tested the SMPs using a tweaked version of the data used for testing the database servers' transaction-processing speed in our Nov. 14, 1994, comparison (page 128). To make the data as scalable and CPU-intensive as possible, we boosted the number of data lookups and calculations. For example, we increased the number of substring searches in our transactions, as well as line items per order, so the database server would not just fetch records but actually compare and manipulate them, processes bound by CPU performance. We eliminated "think" times — a few seconds between each transaction placed in the script to simulate users' pauses.

▸ Forty-eight percent of *InfoWorld* readers surveyed already own an SMP server.

SERVER CONFIGURATION. All of the servers we tested adhered to Intel Corp.'s 1.1 SMP specification. We asked each vendor to configure its server with two Pentium processors (dual 90-MHz or dual 100-MHz), 128MB of RAM, three network interface cards (NICs), and a CD-ROM. If a vendor failed to supply the NICs, we installed three of our own Microdyne Corp. NE3200s. We made sure the servers came with Pentium CPUs without the floating-point math error.

Each vendor's disk subsystem consisted of five 1-gigabyte (GB) drives (except Dell Computer Corp.'s — Dell could only configure its PowerEdge XE 590-2 server with five 2GB drives, because it had no 1GB drives at the time). Four of the five drives in each server were configured with RAID Level 5 to provide cost-effective fault tolerance for our database. In each server, we configured one drive without RAID outside the array to provide optimum performance for our workload.

Eighty-four percent of the respondents to our reader survey said they used a database engine with their SMP server. In addition, 70 percent either currently use or plan to implement Microsoft Corp.'s Windows NT 3.5 as their multiprocessing operating system. As a result, we chose NT 3.5 as the multiprocessing operating system for our benchmark tests and Microsoft's SQL Server 4.21a as our database engine. SQL Server is one of the best database servers we've reviewed, and it was a natural choice to easily test how well Windows NT scales.

If an SMP vendor typically installed the network operating system for its customers, we allowed it to install NT to our specification, which did not vary significantly from NT's default installation. We chose NetBEUI as our only network transport, because IPX's optimization for file-and-print services prevents it from fitting a database server-intensive test.

TEST WORLD CENTER

To optimize it for multiprocessing performance, we turned on two functions in SQL Server: Boost SQL Priority and Dedicated MP Performance, both of which let SQL Server know to use the second processor effectively. In addition, we allowed each vendor to tune one hardware-dependent parameter in SQL Server, called maximum asynchronous I/O, which determined the number of outstanding asynchronous requests at any one time in SQL Server. If that setting is too high or low, I/O performance suffers, according to Microsoft.

The RAID 5 array, which we formatted as an NT File System, housed our database test files. We placed the database's transaction log on the single drive outside the array to provide the best performance environment for our on-line transaction processing (OLTP) task. This optimization technique kept the activity of writing the transaction log from interfering with OLTP.

WORKSTATION CONFIGURATION. We configured 40 workstations in four racks of 10. Each rack consisted of four Gateway 2000 Inc. 486/33s, one Dell 386/33, one Dell 486/25SX, and four Hewlett-Packard Co. 486/66s. All workstations contained 8MB of RAM and a 3Com Corp. 3C509 NIC, except the Dells, which the vendor equipped with Standard Microsystems Corp.'s SMC8000 NICs. We installed Microsoft's Network Client 3.0 for DOS on each client, configuring NetBEUI as the network transport. We installed DOS SQL Utilities on each client, configuring Named Pipes as the TSR to communicate with the network layer.

NETWORK CONFIGURATION. The nature of the workload we chose for the servers meant network bottlenecks were highly unlikely. An analysis of our SMP test revealed less than 1 percent network bandwidth utilization when running transactions on 40 clients. In our Dec. 19, 1994 NOS comparison (page 1), we used four network seg-

▸ How we tested, page 90

MEMORY PITFALLS

When buying a server, be careful about the kind of memory you have to use; it can hinder expandability. Some vendors require you to use error-correcting code (ECC), which can end up being very expensive. This is why others — like Dell — don't require it.

▸ *According to Compaq, its customers' four most important server requirements are dependability, easy management, ease of service, and a good price-to-performance ratio.*

ell PowerEdge XE 590-2

ell Computer Corp.
ustin, Texas
00) 274-3355 or (512) 338-4400;
mail: sales@us.dell.com

Good ●93.75
.70 percent was fourth best.

Satisfactory ● 100.00
.50 tpm barely kept the PowerEdge out of last.
ace.

Poor ●31.25
lismal 3:04:41 on this test was the PowerEdge's
et to dead last place.

ood ●62.50
: PowerEdge comes unconfigured, but it can be
configured by the dealer at an extra cost. Switch
tings are fairly accessible. We had to run each con-
iration utility separately from the command line.
RAID configuration utility is not integrated with
, and the system configuration. The system could
boot from the configuration disk.

ood ●62.50
Dell can accommodate two 100-MHz Pentium
essors, with 512MB of 64-bit system memory
2MB of video memory. The server supports as
h as 24GB of storage. It has eight EISA slots and
PCI slots, but the two buses share the space
re only one card can go.

ry Good ●56.25
the PowerEdge's bus slots were totally accessible
the top of the machine, making up for its shared
and lack of ECC memory (which is an option).
owerEdge doesn't have a lock or a cover for the
r switch. Once the case is off, the bus slots are
sible from the top of the machine, and the
slide in and out of the array easily. The memory
ay be reached by dismounting the backplane
screws). The machine has two conjunctive
r supplies and four fans. The heat sinks on the
ack fans but monitor temperature.

ort and pricing

Good ●37.50
werEdge's manuals include lots of step-by-
structions, flowcharts, and diagrams, and they
arly written. Unfortunately, some features, such
Disk Configuration Utility, are undocumented.

llent ●50.00
arranty covers parts for three years and
s one-year, next-day on-site labor (provided
Dell), plus optional 4-hour turnaround
ire support is unlimited, free, and available
24 hours a day, every day. Support is also
e on-line.

Good ●37.50
o a support representative, we had to first go
an extensive voice-based menu. The techni-
re very friendly and knowledgeable. They
extra hints and tips and walked us through
es step by step.

ctory ●50.00
was $26,982 configured as tested. (A base
ith one 90-MHz CPU, 8MB of RAM, and a
ive costs $7,202.)

Polywell Poly 500EP2

Polywell Computers Inc.
South San Francisco, Calif.
(800) 999-1278 or (415) 583-7222;
E-mail: polywell@tx.netcom.com

Good ●93.75
77.68 percent left the Poly in last place.

Satisfactory ● 100.00
32.03 tpm put the Poly smack in the middle.

Good ●78.13
The Poly finished our I/O tests with a second-best
time of 2:07:17.

Good ●62.50
The Poly came with Windows NT Server 3.5 pre-
installed. Its disk configuration and EISA utilities are
not integrated, and the system cannot boot up from
the configuration disks. But the utilities themselves
are easy to use.

Good ●62.50
The Poly can accommodate as many as two 100-
MHz Pentium processors. It holds as much as 512MB
of 32-bit system memory and 4MB of video memo-
ry. It has four EISA and four PCI slots, though one of
those slots can only use one type at a time. The serv-
er can support as much as 36GB of storage.

Satisfactory ●37.50
The Poly has few refinements and the fewest fea-
tures of the systems we tested. It has the standard
shared cache and does not support ECC. The drive
bays are nicely designed, though. Facing out from
the front of the case, each bay has an individual
keyed lock for security on its hot-swappable drive;
there's a fan on each drive, as well. The drives slide
in and out easily. The machine has two system fans;
one is on the single power supply. The CPU heat
sinks also have fans.

Good ●31.25
Polywell's documentation is average. A provided
three-ring binder can hold all the manuals.

Excellent ●50.00
The warranty covers five years parts and labor. On-
site labor is $200 per year. Technical support is
unlimited and free, weekdays from 7 a.m. to 6 p.m.
Pacific time and Saturdays from 12 p.m. to 5 p.m.
Polywell has a next-day replacement parts program
and a money-back guarantee for first-time buyers.

Unacceptable ●0.00
We spent a lot of time playing phone tag with
Polywell's technical support. When we did get
through, the staff was less than helpful — even
rude — and referred us to component makers.

Excellent ●100.00
The price was $14,475 configured as tested.

6.2

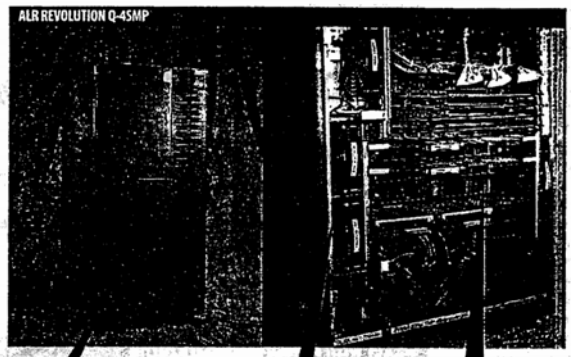
Designed for speed

▸System design can make or break a machine. To upgrade the ProLiant to more than two processors, for example, you have to remove your RAID array; you have to remove the cover from the PowerEdge for even the most routine operation. The Poly had easy accessibility in a plain box; the Manhattan echoed the Revolution's sleek black case, with a much higher quality door on the front.

DRIVE-SWAPPING MADE EASY

Hot-swappable drives can be replaced while the machine is running, saving the administrator the time involved in arranging downtime and maximizing the users' access time. One caveat: It may not be easy to get at the drive that needs swapping.

ALR REVOLUTION Q-4SMP

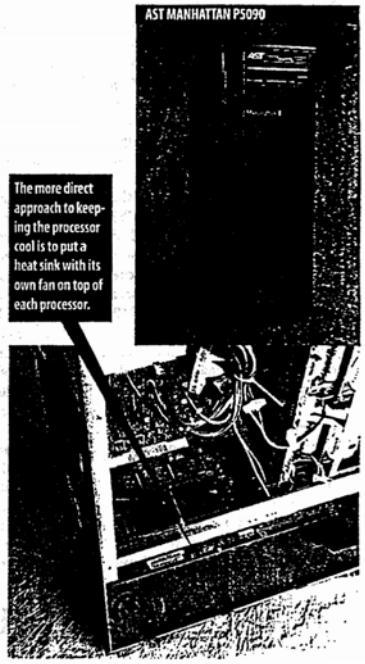


Front doors that lock make for a sleek-looking unit and also protect the server from accidental or unauthorized power-offs.

Plenty of fans keep the RAID array and two processors cool.

The Revolution, like the ProLiant, has room for a total of four processor boards.

AST MANHATTAN P5090



The more direct approach to keeping the processor cool is to put a heat sink with its own fan on top of each processor.

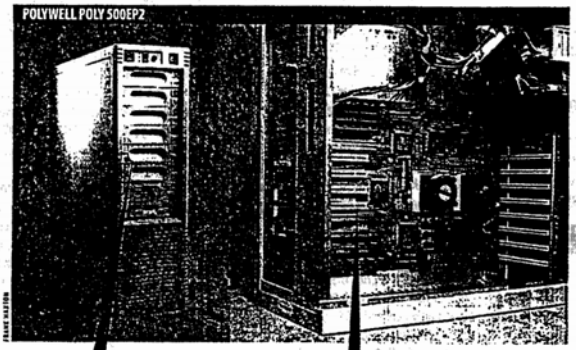
COMPAQ PROLIANT 2000 5/90

The ProLiant has room for two more processor boards if you take out this RAID array.

Opening the ProLiant's door makes the server shut off. Luckily, the door locks.

▸Lots of blinking lights make for snazzy-looking machines, but because most servers spend their days locked in a closet, we didn't find LEDs useful.

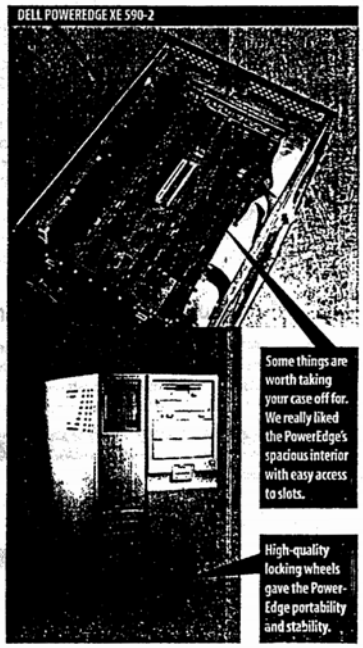
POLYWELL POLY 500EP2



Individually keyed hot-swappable drives are the Poly's only outstanding feature.

The Poly and the PowerEdge each have a shared bus slot, which can hold an EISA or a PCI card — but not both.

DELL POWEREDGE XE 590-2



Some things are worth taking your case off for. We really liked the PowerEdge's spacious interior with easy access to slots.

High-quality locking wheels gave the PowerEdge portability and stability.

MULTIPROCESSING MUSCLE: INSIDE SMP HARDWARE

By Laura Wonnacott

DRIVEN BY INCREASED DEMANDS on their networks, many system administrators are buying their first multiprocessing server. But it's not just a decision between Windows NT, OS/2 for Symmetric Multiprocessing, or the soon-to-be-released NetWare MP. Cache designs on multiprocessing servers are as different as condominiums, town houses, and ranch houses. Understanding fundamental design architectures can reduce a lot of the hassles for the first-time buyer.

Like single-processor Pentium PCs, symmetric multiprocessors (SMPs) come equipped with not only 16KB of on-board cache (on the chip), but also a secondary, external hardware cache called Level 2 cache to help eliminate processing bottlenecks. How the SMP uses this secondary cache varies, however, depending on whether it's a dual or multiprocessing machine. Dual-processor SMPs share a single large Level 2 cache, resulting in a less expensive system. Multiprocessor PCs, on the other hand, come with a dedicated Level 2 cache for each processor.

In a CPU-intensive environment, such as transaction processing, dedicated Level 2 cache allows more cache hits (times when a processor finds what it needs in the cache) than a shared Level 2 cache does. That's because when a cache is shared, processor contention (when both processors want access to the same information) is more likely to occur, resulting in one processor having to wait for the other processor to finish using the Level 2 cache.

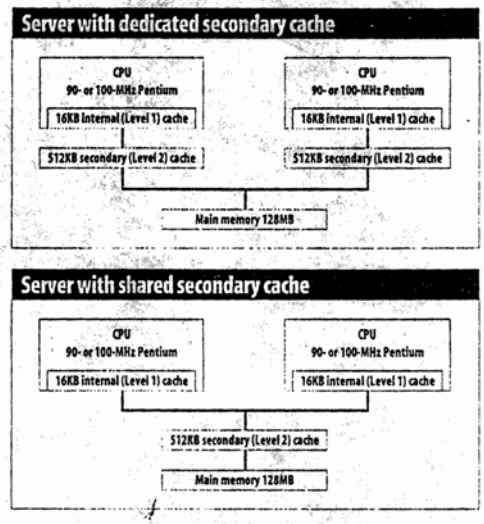
The results of our on-line transaction processing tests confirm the speed benefits of a dedicated Level 2 cache. Two of the five servers in this comparison (the Compaq ProLiant 2000 5/90 and ALR Revolution Q-4SMP) and a RISC server, the NEC Technologies Inc. RISCserver 2200, which we did not score (see story, below) offer a dedicated Level 2 cache. Not surprisingly, these servers out-performed all others in our CPU-intensive transaction processing test. In addition, these servers proved the most scalable in moving from one to two processors.

Given the two basic cache designs, there's still a lot a vendor can do to enhance processing performance. For example, larger caches are always helpful. The 2MB of Level 2 cache in the NEC RISCserver 2200 we tested no doubt was at least partially responsible for the machine's superior transaction processing performance.

The slower Intel-based servers typically had about 512KB of Level 2 cache. Other optimization techniques exist, such as Compaq's optional Transaction Blaster, which offers a third level of caching to further enhance processing performance.

Multiprocessing muscle – two symmetric multiprocessing PC design architectures

We tested servers containing two cache designs. One with a dedicated secondary (Level 2) cache for each processor, the other with a shared secondary (Level 2) cache for all processors.



MULTIPROCESSING WITH NOVELL

Multiprocessing on Intel architecture is not limited to Windows NT users. Novell's SFT III, with the appropriate NetWare Loadable Module, can use two processors on systems that adhere to its specifications. When SFT III has two processors, it off-loads some of the communications workload (which can cause a significant drop in performance) to the second processor, regaining as much as 15 percent of lost performance.

▸ *Intel's MP specification gives buyers some assurance that future MP operating systems will perform well on their servers, though vendors can enhance this scalability by going beyond the spec on their system designs.*

The need for more speed: MIPS RISC is a fast Intel alternative

By Laura Wonnacott

Do you need even more processing speed? If you're willing to consider a non-Intel architecture, take a look at the RISC symmetric multiprocessing (SMP) architecture based on the chip from MIPS Technologies Inc. We tested a MIPS-based machine using the same test bed and script used for the SMPs in this comparison, and we're impressed.

The machine we tested, NEC Technologies Inc.'s RISCserver 2200, completed 50 transactions on 40 workstations in an average of 8 minutes and 53 seconds, 24 percent and 12 minutes ahead of the fastest two-Pentium server in the comparison, the ALR Revolution Q-4SMP. In fact, the RISCserver was just as fast processing transactions on one processor as the slowest Pentium on two processors (AST's Manhattan P5090).

The RISCserver's disk subsystem performance was not as impressive as its on-line transaction processing performance. We suspect that the server was experiencing some problems with its disk subsystem, because the server reset its SCSI bus several times during our tests.

What makes MIPS RISC servers so fast? One reason for the NEC RISCserver's performance is probably that Microsoft NT, the only operating system it will run, was developed on MIPS architecture. It's not surprising that NT performs well on its native platform. The 64-bit MIPS RISC processors also help; the RISCserver 2200 contains two 200-MHz MIPS CPUs. Its dedicated cache design is similar to the classic SMP architecture in the Compaq ProLiant 2000 5/90 and the Revolution (see story, above). The RISCserver has more Level 2 cache memory, though (2MB per CPU vs. the 512KB in the Intel machines).

▸ **RISC vendors such as Sun provide customized C++ compilers for building multithreaded SMP applications.**

In addition to 2MB of Level 2 cache, the caching mechanism does something called "cache snooping," which allows a processor to look for data in the other processors' Level 2 cache before going to main memory. The processor gains speed without hurting overall performance, because it can "snoop" without locking out the other processor from its own cache.

So why aren't buyers flocking to MIPS? It can't be the price. The NEC RISCserver 2200's estimated street price is $32,195 for our test configuration (with RAID Level 5). That's only about $5,000 more than the most expensive Intel-based server we reviewed (Dell's PowerEdge XE 590-2).

The difference in architecture no doubt has something to do with buyers' shyness. MIPS RISC is certainly not Intel. The utility programs have a different flavor from the old-style Intel-based utilities. For example, you can't boot from a floppy. The RISCserver starts based on what's stored in nonvolatile RAM and comes up with something called an ark menu, as defined by the MIPS specification. To change the system's configuration, you choose an option called Run Setup from the ark menu after booting. Setup contains most of the system's configuration. To change the EISA configuration, you'll need to choose Run Program and then specify the A: drive and program name.

But moving from a single to multiple processors was a snap. Unlike with Intel, there's no need to load a different kernel when adding processors. Microsoft's NT and SQL Server on MIPS have an identical look and feel to that of their Intel versions. In fact, the support for Intel, MIPS, and Alpha is distributed on the same CD-ROM, so you may already have a copy of the MIPS version. The greatest difference with MIPS is in byte ordering, which you can't see from the interface. This means you can't merely move your SQL Server databases onto MIPS by copying the files. SQL Server includes an SQL Transfer Manager that lets you migrate data from one platform to another.

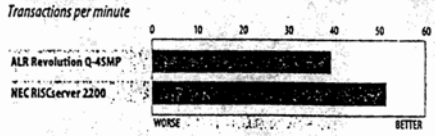
If you use Microsoft's NT and SQL Server, then MIPS RISC is a fast Intel alternative that doesn't cost much more.

RISC takers should give Pentiums another look

Pentium-based machines are finally approaching RISC performance. We pitted the strongest performer in this comparison, the ALR Revolution Q-4SMP, against NEC Technologies Inc.'s RISCserver 2200, which was still in beta when we ran our tests. The Revolution only did 24 percent fewer transactions per minute than the NEC RISC machine.

Fastest Pentium vs. NEC RISCserver 2200

Transactions per minute



PRODUCT COMPARISON

▸ How we tested (from page 85)

ments to remove the cable as a bottleneck and target overall server performance. In this comparison, we used three segments, because with less I/O and more CPU-intensive tests, we didn't have as much network traffic.

We distributed three standard racks per network segment, and the fourth rack was distributed across the existing three segments. This isolated server CPU performance from other network performance variables.

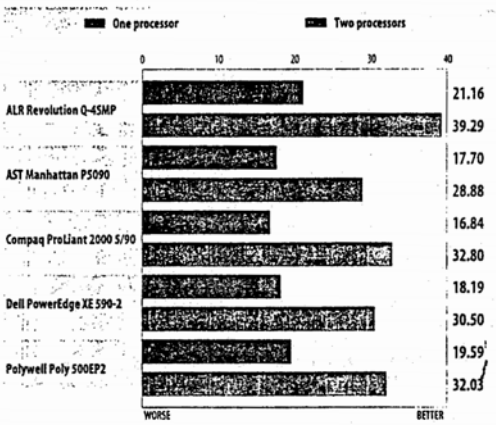
Each segment was supported by a Cabletron Systems Inc. Multi Media Access Center M8FNB concentrator.

Double your processors ...

... And nearly double your fun. The Revolution processed the largest number of transactions per minute on two processors – 39.29. But the ProLiant was the most scalable, nearly doubling its speed when we added the second processor.

Transaction processing

Transactions per minute



PARALLEL VS. SYMMETRIC

Machines are not limited to symmetric multiprocessing; parallel architectures provide even better performance, at the expense of both price and flexibility of platforms.

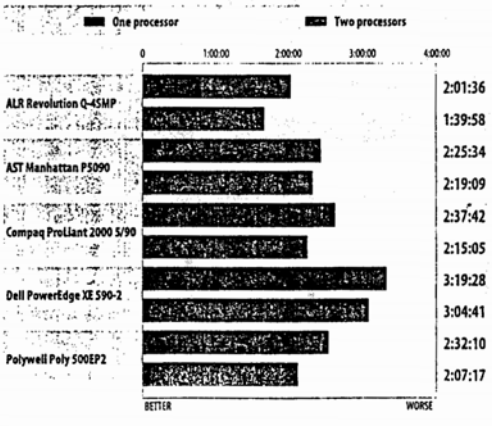
Distributed processing works over distributed networks but puts a heavy dent in communications overhead — not good news for a network that's already overloaded. Overall, symmetric multiprocessing is the easiest, most flexible, and most cost-effective way to add multiprocessing performance to your network.

Can't make up their minds fast enough

If decision support makes up the bulk of your network load, think twice before trying to improve performance through symmetric multiprocessors. The Revolution scaled the best when we went to two processors to run our database queries, but none of the servers scaled nearly as well as they did in our on-line transaction processing tests.

Decision support

Time to complete queries on four workstations. Times in hours:minutes:seconds.



THE TESTS. We based the scores for our performance categories (scalability, transaction processing, and decision support) on transaction-processing and query benchmark tests performed on the 40-workstation network.

The transaction-processing script simulated an on-line order-entry system by processing 50 transactions simultaneously across the 40 clients. The queries accessed the same database from four workstations after they processed transactions. We ran the transaction-processing script twice to measure scalability from one to two processors and averaged the results.

Each transaction looked up a customer by identification number or by searching for the name in our customer table. Next, we calculated the next invoice number and created an order by inserting a row into an orders table.

For each part or line item a customer wished to order, we searched our parts table by either an exact part number or a partial description of the part name as supplied by the customer. We updated several quantity fields, such as amount on hand and on back order, if applicable. We then inserted a row into our "parts ordered" table for each line item a customer wished to order. We created an invoice form and updated the sales commission for the appropriate sales representative.

▸ Some SMP servers can have more than 1 terabyte of disk space if you add external drive cages via SCSI.

After the 40 workstations completed their transactions, four began processing I/O-bound database requests — our queries. The first workstation processed two sales queries. The second workstation processed two ad-hoc queries. The third workstation selected a set of orders from the orders table and inserted it into a temporary table. The fourth workstation processed a large select query from our parts-ordered table and sorted the results by part number. All the workstations sent query results to the server's disk to test disk subsystem performance.

PERFORMANCE

Scalability

We determined how scalable each server was by measuring its performance after we added a second processor. We first ran our 50 transactions on the 40 clients with the server running NT's uniprocessor kernel. We then ran 50 transactions on 40 clients with the server running NT's multiprocessor kernel with two processors enabled.

We scored scalability as a percentage. Perfect scalability was 100 percent. For example, if a server that ran 50 transactions on a single processor in one hour scaled perfectly, it would complete 50 transactions on two processors in 30 minutes. A server that scaled more than 95 percent we rated excellent; 95 percent to 86 percent earned a very good score; 85 percent to 76 percent earned a good; 75 percent to 66 earned a satisfactory; 65 percent to 56 percent earned a poor; and a server that scaled less than 56 percent was unacceptable.

Transaction processing

To determine test results for each server, we first calculated the average time to run 50 transactions on 40 clients. We then determined the transactions per minute (tpm) for each client and multiplied that by 40 (total number of clients) to arrive at the server's tpm.

A server that processed greater than 49 tpm earned a score of excellent; 49 to 40 tpm earned a score of very good; 39 to 35 tpm earned a score of good; 34 to 25 tpm earned a score of satisfactory; 24 to 20 tpm earned a score of poor; and a server that processed less than 20 tpm was unacceptable.

Decision support

We designed this benchmark test to show how well each server could handle I/O-intensive work on a network that served both decision-support and CPU-intensive requests. We first calculated the average time it took each of the four workstations on our 40-client network to complete our queries. We then figured the averages.

A server that completed our queries in less than 1 hour and 15 minutes received an excellent; a time between 1 hour and 15 minutes and 1 hour and 45 minutes received a very good; between 1 hour and 45 minutes and 2 hours and 15 minutes earned a good; between 2 hours and 15 minutes and 2 hours and 45 minutes earned a satisfactory; between 2 hours and 45 minutes and 3 hours and 15 minutes earned a poor. Anything slower got an unacceptable.

Setup and ease of use

We attempted to capture the experience of setting up the server out of the box. We evaluated how easy it was to get the server running on the network, paying careful attention to both the EISA configuration and the RAID disk subsystem utilities. For a product to receive a score of excellent, the EISA configuration had to be completed, the disk subsystem initialized and operative, and the operating system installed. A server that we could set up by following a few uncomplicated tasks received a score of very good.

Expandability

We looked for expandable server components. The more system memory, cache, slots, drive bays, and different types of hardware buses the server could accommodate after our configuration, the higher the score.

For consistency, we defined a drive as external if we did not have to remove a case, even if it was protected behind a door on the server itself.

System design

We carefully examined each server to determine any significant design advantages or flaws. Servers that offered

more than one kind of bus, integrated hard drive interfaces on the system board, and patch-free system boards scored the highest. We also gave bonuses for easy access to parts, dedicated CPU cache designs, error-correcting code (ECC) memory, or hot-swappable drive arrays. We noted how easy it was to add processors or memory to the system. Cases that were hard to open, parts that were difficult to reach, or the lack of adequate fans hurt the score.

Compatibility

SMP servers typically support a variety of operating systems. The more operating systems a server supported, the higher the score.

SUPPORT AND PRICING

Documentation

We looked for clear and concise documentation. We awarded a score of satisfactory if the documentation explained how to set up and configure the server. We also required it to include accurate illustrations and diagrams.

Support policies

We gave a satisfactory score for unlimited free support and a one-year warranty. We gave bonus points for support via fax, on-line services, a money-back guarantee, extended hours, and a toll-free line. We subtracted points for limited or no support.

Technical support

We based technical support scores on the quality of service we received during multiple anonymous calls and on the availability of knowledgeable personnel. We awarded bonus points for extra helpfulness. We subtracted points for unreturned calls or long waits on hold.

Price

We based this score on the price of the server as configured for this comparison, except for the three server NICs (which we omitted because several vendors did not provide NICs). We used the vendor's suggested street price, when available, or suggested retail price. Servers that cost less than $15,000 received an excellent; those that cost $15,001 to $19,999 rated very good; those that cost $20,000 to $24,999 rated good; those that cost $25,000 to $29,999 rated satisfactory; and those that cost more than $30,000 received a score of poor.

SCALABILITY: YOU NEED MORE THAN JUST GOOD HARDWARE

by Laura Wonnacott

OUR TESTS SHOW that Windows NT scales well (at least as far as four processors; see story, right), but a scalable operating system and server hardware are not ough to guarantee scalability. Applitions must also be designed with scalility as an underlying objective. A orly designed application executes necessary code, wasting precious ocessing resources.

Writing scalable code begins by ndoning the traditional mindset, in ich code starts at the top and finishit the bottom. An application devel- r must analyze programs to deter- e which portions of code can be cessed simultaneously by different Js.

hreads, the basic unit of execution ultiprocessor applications, are the to this objective of parallel design. eads can run on any processor in a tiprocessor system. Splitting a sinthread into multiple concurrent ds is a great way to boost a multiessing server. Older, more tradil applications often require a ss to finish before they continue to ext one. In these older systems, cations and processors cannot the load on a single unit of work, large query.

gle-threading applications will rm the same no matter how many processors are used. The only way to realize a performance gain with a multiprocessor system is to use a multithreading application. Microsoft's SQL Server 4.21a, a multithreaded application designed to take advantage of additional processors, played a vital role in the scalability we saw in this comparison. We tweaked our on-line transaction processing application to better test scalability. Our original transaction wasn't a good test of multithreading, because its think times let the CPU sit idle.

Turning a single-threaded application into a multithreaded application requires a working knowledge of how threads work. Threads exist in three states — waiting (not ready to run), ready, or running. The number of runnable threads is limited by the system's resources; the number of threads running at once is limited by the number of processors in the system.

Many application developers (including us) still aren't familiar with all the programming techniques available for taking advantage of multiple processors. Windows NT provides a number of sophisticated synchronization objects, such as I/O completion ports, multiple synchronization objects, asynchronous I/O, and spinlocks. But as the demand for scalable applications increases, a working knowledge of these features will be essential to writing multiprocessing applications.

Microsoft's dirty little secret

By Laura Wonnacott

According to Microsoft Corp., SQL Server 4.21a supports as many as four processors, the same number the Windows NT 3.5 network operating system supports out of the box. But in some of our ad-hoc testing to see how well the two Microsoft products scaled beyond two processors, we discovered that SQL Server 4.21a's default configuration cripples it so that it's unable to use more than three processors — and Microsoft says changing the default is risky.

We had to conduct the scalability and speed tests in this comparison using only two processors, because that was the most supported by three of the five servers (AST's Manhattan P5090, Dell's PowerEdge XE 590-2, and Polywell's Poly 500EP2). Machines that share cache between two processors, instead of having dedicated cache for each CPU, can't be upgraded to four processors (see story, left). We wanted to see what NT 3.5 and SQL Server could do with the throttle open, so we fired them up on an available Compaq ProLiant 4000 5/66 with four 66-MHz Pentium CPUs and 128MB of RAM, and we ran our transaction processing benchmark test with one, two, three, and then four processors.

A server that scales perfectly doubles its performance by going to two processors. A perfect scale from two to three processors would improve performance 33 percent. Adding a fourth processor would increase performance over three processors only 25 percent but double the performance obtained with two processors, and so on. The scalability we witnessed using NT 3.5 and SQL Server beyond two processors was grim. As the graph below depicts, SQL Server seemed unable to find the third processor at all.

Puzzled, we checked our configuration carefully to make sure we had set SQL Server for dedicated multiprocessor performance. We had. It was only after several discussions with Microsoft that we found out about SMPStat, an undocumented and unsupported parameter set through NT's registry editor that declares the number of CPUs SQL Server can use.

In effect, SMPStat is SQL Server's throttle for multiprocessor performance. When we originally configured SQL Server for dedicated multiprocessor performance, the program automatically set SMPStat to zero, which tells SQL Server to use n-1 processors when the number of processors are greater than two. As a result, when we ran our test with three processors, SQL Server 4.21a did not take advantage of the third processor. When we ran our test with four processors, SQL Server used only three processors. We set SMPStat to -1, which tells SQL Server to use all available processors, and we reran our tests. With the proper setting, SQL Server came very close to perfect scalability.

According to Microsoft, fooling around with SMPStat could cause two program threads to eventually deadlock, bringing the database server to a halt. Had we run a battery of regression tests, we might have seen this happen, but we ran out of time to test, so we only have Microsoft's word to go on.

At any rate, SQL Server 6.0, now in beta testing and due the first half of this year, will come with SMPStat set to use all available processors without the risk of a deadlock, according to Microsoft. We tested a beta of Version 6.0 and were able to verify Microsoft's claim — the SQL Server 6.0 beta scaled similarly to SQL Server 4.21a with the SMPStat value set to -1. An interesting aside: The SQL Server 6.0 beta performed our test slightly faster on all multiprocessors than SQL Server 4.21a with SMPStat optimized for performance.

Who knows what CPUs lurk in the heart of SQL Server?

If you liked your Dick Tracy decoder ring, you'll like Microsoft SQL Server 4.21a. This version of SQL Server has a sneaky little undocumented tweak called SMPStat, which tells the machine to use all the available processors. If you don't alter SMPStat correctly, SQL Server 4.21a won't take full advantage of more than two processors. The catch: Microsoft doesn't want you to use SMPStat, which it says could result in deadly embraces. If you disobey and get into trouble, Microsoft won't help. SMPStat will be the default in SQL Server, Version 6.0.

SQL Server processing performance

Transactions per minute

KNOW YOUR RAID LEVELS

RAID 0 is simple data striping; RAID 1, simple disk mirroring. RAID 2 stripes data on mirrored disks; in RAID 3, bytes of data are striped across disks, with one drive storing parity information. With RAID 4, blocks of data are striped across disks, and one drive stores parity information; and in RAID 5, blocks of data and parity information are striped across all drives.

▸ Each channel on a SCSI controller is the equivalent of a complete controller. By putting two or three channels on one board, you can get the benefits of having several boards on the system without losing the actual slots.

METRIC
TIPROCESSING
ERS

PRODUCT COMPARISON

Support

	ALR Revolution Q-4SMP	AST Manhattan P5090	Compaq ProLiant 2000 5/90	Dell PowerEdge XE 590-2	Polywell Poly 500EP2
Free telephone support provided by vendor	Yes	Yes	Yes	Yes	Yes
Telephone support hours	Weekdays 6 a.m. to 6 p.m., Saturday 7 a.m. to 1 p.m. Pacific time	24 hours per day, 7 days per week	24 hours per day, 7 days per week	24 hours per day, 7 days per week	Weekdays 7 a.m. to 6 p.m., Saturday noon to 5 p.m. Pacific time
Warranty period	5 years on ALR motherboard, chassis, and parts; 3 years on third-party peripherals, drives, and parts	3 years with on-site support; optional 4-hour response by Memorex Telex	3 years next-day, on-site parts and labor; optional 4-hour response	3 years parts; 1 year next-day on-site labor; optional 4-hour turnaround	5 years parts and labor; 2 years on third-party components
Vendor-provided on-site service	1 year[1]	3 years	3 years	1 year	None
Money-back guarantee	No	No	No	No	Yes
On-line support	In-house BBS, CompuServe, World Wide Web server	In-house BBS, CompuServe, Prodigy	In-house BBS, CompuServe, America Online, World Wide Web server	In-house BBS, CompuServe, World Wide Web server, FTP server	In-house BBS, Internet E-mail
Fax-back support	Yes	Yes	Yes	Yes	Yes
Support policies score	Excellent	Excellent	Excellent	Excellent	Excellent
Technical support score	Satisfactory	Good	Satisfactory	Very Good	Unacceptable

1. $9.95 registration fee.

Features

	ALR Revolution Q-4SMP	AST Manhattan P5090	Compaq ProLiant 2000 5/90	Dell PowerEdge XE 590-2	Polywell Poly 500EP2
Setup and ease of use					
Additional preconfigured components	3 ALR E-Net 32 Ethernet controllers	None	3 Compaq NetFlex Ethernet controllers	5 2GB drives	Znyx 4-port PCI Ethernet controller
Power supply monitoring	No	Voltage monitoring software (requires NetWare); also an LED on the unit	Insight Manager software monitors voltage, temperature	Dell SafeSite software monitors voltage, temperature	No
On-line diagnostics	No	Yes	Yes	Yes	Yes[1]
Expandability					
Maximum number of Pentiums	4 90-MHz or 4 100-MHz	2 90-MHz	4 90-MHz[2]	2 90-MHz or 2 100-MHz	2 100-MHz
Maximum 32-bit system memory	1GB using 64MB SIMMs	256MB	512MB	512MB using 64MB SIMMs (not sold by Dell)	512MB using 64MB or 128MB SIMMs
Maximum external cache RAM	512KB per CPU	256KB shared	512KB per CPU	512KB shared	512KB shared
Maximum video memory	2MB	512KB	1MB	2MB	4MB
EISA slots	10	6	8	8[3]	4[3]
PCI slots	0	2	0	2[3]	4[3]
Proprietary slots (VESA, CPU, memory)	4 (3 EISA slots have VESA local bus extension)	None	4 proprietary EISA (2 CPU boards, 1 memory, 1 modem slot)	None	None
Levels of RAID supported	0, 1, 5	0, 1, 5	0, 1, 4, 5, 10	0, 1, 4, 5, 10	0, 1, 2, 5, 6, 10
Number of free slots in our configuration	5 EISA	3 EISA, 1 PCI	4 EISA and 3 proprietary upgrade slots (2 CPU boards, 1 memory)	4 EISA, 2 PCI[3]	2 EISA, 1 PCI[3]
Total number of external drive bays	11 half-height (with hot-swap option, every 3 hot-swap bays use 2 half-height bays), 2 3.5-inch	8 half-height (6 hot-swappable), 1 3.5-inch	2 half-height	4 half-height	8 half-height (6 hot-swappable)
Total number of internal drive bays	None	None	5 half-height (hot-swappable)	8 (hot-swappable)	2 half-height
Number of free drive bays after configuration	4 half-height, 4 hot-swappable bays[4], and 1 3.5-inch; all external	1 external half-height, 1 external hot-swappable bay	1 external half-height	2 external half-height, 2 internal hot-swappable bays	2 internal half-height
System design[5]					
Integrated hard drive interface	1 IDE	No	Embedded SCSI port	Embedded SCSI port	None embedded but ships with an additional controller
Type of video display	SVGA	SVGA	SVGA	ATI Mach 32 chip on PCI bus	Diamond Viper PCI card with 2MB video RAM 1,280 by 1,024
Number and type of serial ports	2 9-pin	2 9-pin	2 9-pin	2 9-pin	1 25-pin, 1 9-pin
Number of parallel ports	1	1	1	1	1 on add-on board
Types and locations of external SCSI ports	1 SCSI-2 port on controller on the back of the RAID caching controller	1 SCSI-2 port on controller	1 external SCSI-2 port off the motherboard, 1 external SCSI port off the RAID controller	1 embedded SCSI port on motherboard, 1 external SCSI-2 port on back of the system, running off the RAID controller	1 external SCSI-2 port off the RAID controller, 1 external SCSI-2 port off the CD-ROM controller
Keyboard security features	Keyboard disable button, password control	Password control	Password control	Password control	Keyboard lock, password control
Case lock	Yes; one for each side	Yes	Yes	Yes	Each hot-swappable drive has an individual keyed lock. The case itself has no locks.
BIOS manufacturer	Phoenix	Phoenix/AST	Compaq	Phoenix	Award Modular BIOS, Version 4.50G
Error-correcting memory type, amount	128MB error detection and correction (EDC) using standard DRAM	16MB error-correcting code (ECC)	32MB ECC	32MB ECC (optional)	None
Shared or dedicated L2 cache, size	Dedicated cache, 512KB per CPU	Shared 256KB cache	Dedicated cache, 512KB per CPU	Shared 512KB cache	Shared 512KB cache
UPS support	No	No	Yes (optional)	No	Yes
LEDs on system	Power, hard drive activity, memory, local bus slave and master, EISA slave and master	Power, hard drive activity, temperature, voltage, 4-digit post display (behind door)	Power and drive LEDs for each drive	Power, failed drive, diagnostics test	Power, drive for system and each bay

1. We could not get the supplied diagnostics to run.
2. You must remove the RAID array and house it in a separate drive cage in order to fit more than two Pentium boards.
3. One EISA and one PCI card share a slot.

ıtroduce
-bound
ivities into
ır appli-
ion and
tch scala-
ty run out
ras. Add
cessors to a
ver running
I/O-bound
olication,
d you'll
bably
ate even
re of an
) bottle-
:k.

AJ

VESA®

VUMA Proposal

(Draft)

Video Electronics Standards Association

2150 North First Street, Suite 440
San Jose, CA 95131-2029

Phone: (408) 435-0333
FAX: (408) 435-8225

VESA Unified Memory Architecture Hardware Specifications Proposal

Version: 1.0p

Document Revision: 0.4p

October 31, 1995

Important Notice: This is a **draft** document from the Video Electronics Standards Association (VESA) Unified Memory Architecture Committee (VUMA). It is only for discussion purposes within the committee and with any other persons or organizations that the committee has determined should be invited to review or otherwise contribute to it. It has not been presented or ratified by the VESA general membership.

Purpose

To enable core logic chipset and VUMA device designers to design VUMA devices supporting the Unified Memory Architecture.

Summary

This document contains a specification for VUMA devices' hardware interface. It includes logical and electrical interface specifications. The BIOS protocol is described in VESA document VUMA VESA BIOS Extensions (VUMA-SBE) rev. 1.0.

Scope

Because this is a draft document, it cannot be considered complete or accurate in all respects although every effort has been made to minimize errors.

Intellectual Property

Trademarks

All trademarks used in this document are the property of their respective owners. VESA and VUMA are trademarks owned by the Video Electronics Standards Association.

Patents

The proposals and standards developed and adopted by VESA are intended to promote uniformity and economies of scale in the video electronics industry. VESA strives for standards that will benefit both the industry and end users of video electronics products. VESA cannot ensure that the adoption of a standard; the use of a method described as a standard: or the making, using, or selling of a product in compliance with the standard does not infringe upon the intellectual property rights (including patents, trademarks, and copyrights) of others. VESA, therefore, makes no warranties, expressed or implied, that products conforming to a VESA standard do not infringe on the intellectual property rights of others, and accepts no liability direct, indirect or consequential, for any such infringement.

Support For This Specification

If you have a product that incorporates VUMA™, you should ask the company that manufactured your product for assistance. If you are a manufacturer of the product, VESA can assist you with any clarification that you may require. All questions must be sent in writing to VESA via:

(The following list is the preferred order for contacting VESA.)

VESA World Wide Web Page:	*www.vesa.org*
Fax:	(408) 435-8225
Mail:	VESA 2150 North First Street Suite 440 San Jose, California 95131-2029

Acknowledgments

This document would not have been possible without the efforts of the members of the 1995 VESA Unified Memory Architecture Committee and the professional support of the VESA staff.

Work Group Members

Any industry standard requires information from many sources. The following list recognizes members of the VUMA Committee, which was responsible for combining all of the industry input into this proposal.

Chairperson

Rajesh Shakkarwar	OPTi

Members

Jonathan Claman	S3
Jim Jirgal	VLSI Technology Inc.
Don Pannell	Sierra Semiconductor
Wallace Kou	Western Digital
Derek Johnson	Cypress
Andy Daniel	Alliance Semiconductor
Long Nguyen	Oak Technology
Robert Tsay	Pacific Micro Computing Inc.
Sunil Bhatia	Mentor Arc
Peter Cheng	Samsung
Alan Mormann	Micron
Solomon Alemayehu	Hitachi America Ltd.
Larry Alchesky	Mitsubishi
Dean Hays	Weitek

Revision History

Initial Revision 0.1p — Sept. 21 '95

Revision 0.2p — Oct 5 '95
- Added sync DRAM support
- Electrical Section
- Boot Protocol
- Reformatted document

Revision 0.3p — Oct 19 '95
- Graphics controller replaced with VUMA device
- MD[n:0] changed to t/s
- Modified Aux Memory description
- Added third solution to Memory Expansion Problem
- Synch DRAM burst length changed to 2/4
- Modified all the bus hand off diagrams
- Added DRAM Driver Characteristics section

Revision 0.4p — Oct 19 '95
- Sync DRAM Burst Length changed to 1/2/4
- DRAM controller pin multiplexing added
- Changed AC timing parameters

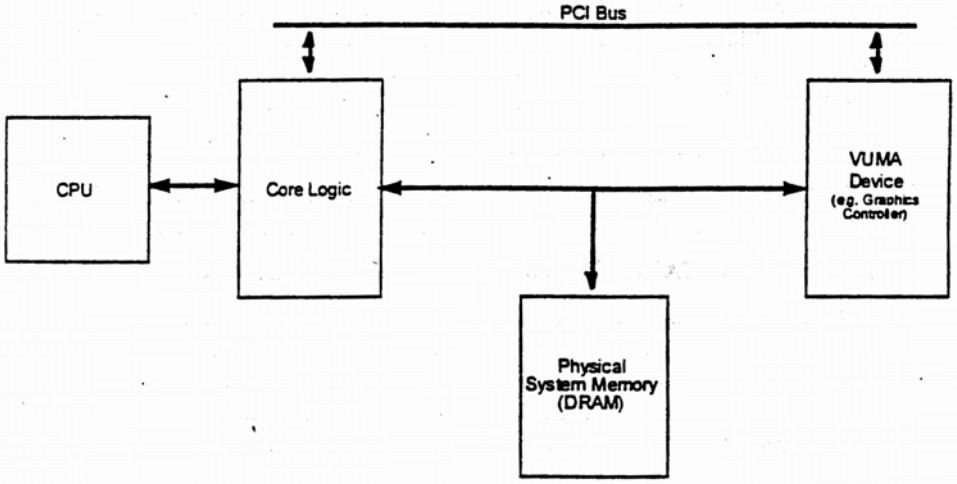
TABLE OF CONTENTS

1.0 Introduction

The concept of VESA Unified Memory Architecture (VUMA) is to share physical system memory (DRAM) between system and an external device, a VUMA device; as shown in Figure 1-1. A VUMA device could be any type of controller which needs to share physical system memory (DRAM) with system and directly access it. One example of a VUMA device is graphics controller. In a VUMA system, graphics controller will incorporate graphics frame buffer in physical system memory (DRAM) or in other words VUMA device will use a part of physical system memory as its frame buffer, thus, sharing it with system and directly accessing it. This will eliminate the need for separate graphics memory, resulting in cost savings. Memory sharing is achieved by physically connecting core logic chipset (hereafter referred to as core logic) and VUMA device to the same physical system memory DRAM pins. Though the current version covers sharing of physical system memory only between core logic and a motherboard VUMA device, the next version will cover an expansion connector, connected to physical system memory DRAM pins. An OEM will be able to connect any type of device to the physical system memory DRAM pins through the expansion connector.

Though a VUMA device could be any type of controller, the discussion in the specifications emphasizes a graphics controller as it will be the first VUMA system implementation.

Figure 1-1 VUMA System Block Diagram



2.0 Signal Definition

2.1 Signal Type Definition

in	Input is a standard input-only signal.
out	Totem Pole Output is a standard active driver
t/s	Tri-State is a bi-directional, tri-state input/output pin.
s/t/s	Sustained Tri-state is an active low or active high tri-state signal owned and driven by one and only one agent at a time. The agent that drives an s/t/s pin active must drive it high for at least one clock before letting it float. A pullup is required to sustain the high state until another agent drives it. Either internal or external pullup must be provided by core logic. A VUMA device can also optionally provide an internal or external pullup.

2.2 Arbitration Signals

MREQ#	**in** **out**	MREQ# is out for VUMA device and in for core logic. This signal is used by VUMA device to inform core logic that it needs to access shared physical system memory bus.
MGNT#	**in** **out**	MGNT# is out for core logic and in for VUMA device. This signal is used by core logic to inform VUMA device that it can access shared physical system memory bus.
CPUCLK	**in**	CPUCLK is driven by a clock driver. CPUCLK is in for core logic, VUMA device and synchronous DRAM.

2.3 Fast Page Mode, EDO and BEDO DRAMs

RAS#	s/t/s	Active low row address strobe for memory banks. Core logic will have multiple RAS#s to support multiple banks. VUMA device could have a single RAS# or multiple RAS#s. These signals are shared by core logic and VUMA device. They are driven by current bus master.
CAS[n:0]#	s/t/s	Active low column address strobes, one for each byte lane. In case of pentium-class systems n is 7. These signals are shared by core logic and VUMA device. They are driven by current bus master.
WE#	s/t/s	Active low write enable. This signal is shared by core logic and VUMA device. It is driven by current bus master.
OE#	s/t/s	Active low output enable. This signal exists only on EDO and BEDO. This signal is shared by core logic and VUMA device.

		It is driven by current bus master.
MA[11:0]	s/t/s	Multiplexed memory address. These signals are shared by core logic and VUMA device. They are driven by current bus master.
MD[n:0]	t/s	Bi-directional memory data bus. In case of pentium-class systems n is 63. These signals are shared by core logic and VUMA device. They are driven by current bus master.

2.4 Synchronous DRAM

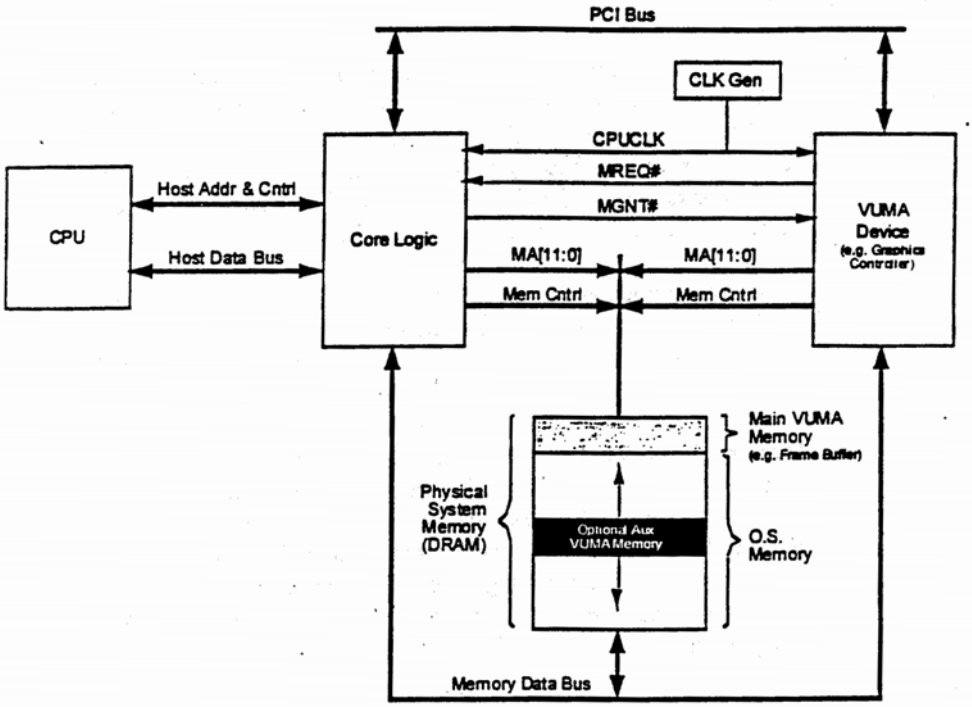
CPUCLK	in	CPUCLK is the master clock input (referred to as CLK in synchronous DRAM data books). All DRAM input/ output signals are referenced to the CPUCLK rising edge.
CKE	s/t/s	CKE determines validity of the next CPUCLK. If CKE is high, the next CPUCLK rising edge is valid; otherwise it is invalid. This signal also plays role in entering power down mode and refresh modes. This signal is shared by core logic and VUMA device. It is driven by current bus master.
CS#	s/t/s	CS# low starts the command input cycle. CS# is used to select a bank of Synchronous DRAM. Core logic will have multiple CS#s to support multiple banks. VUMA device could have a single CS# or multiple CS#s. These signals are shared by core logic and VUMA device. They are driven by current bus master.
RAS#	s/t/s	Active low row address strobe. This signal is shared by core logic and VUMA device. It is driven by current bus master.
CAS#	s/t/s	Active low column address strobe. This signal is shared by core logic and VUMA device. It is driven by current bus master.
WE#	s/t/s	Active low write enable. This signal is shared by core logic and VUMA device. It is driven by current bus master.
MA[11:0]	s/t/s	Multiplexed memory address. These signals are shared by core logic and VUMA device. They are driven by current bus master.
DQM[n:0]	s/t/s	I/O buffer control signals, one for each byte lane. In case of pentium-class systems n is 7. In read mode they control the output buffers like a conventional OE# pin. In write mode, they control the word mask. These signals are shared by core logic and VUMA device. They are driven by current bus master.
MD[n:0]	t/s	Bi-directional memory data bus. In case of pentium-class systems n is 63. These signals are shared by core logic and VUMA device. They are driven by current bus master.

3.0 Physical Interface

3.1 Physical System Memory Sharing

Figure 3-1 depicts the VUMA Block Diagram. Core logic and VUMA device are physically connected to the same DRAM pins. Since they share a common resource, they need to arbitrate for it. PCI/VL/ISA external masters also need to access the same DRAM resource. Core logic incorporates the arbiter and takes care of arbitration amongst various contenders.

Figure 3-1 VUMA Block Diagram



As shown in Figure 3-1, VUMA device arbitrates with core logic for access to the shared physical system memory through a three signal arbitration scheme viz. MREQ#, MGNT# and CPUCLK. MREQ# is a signal driven by VUMA device to core logic and MGNT# is a signal driven by the core logic to VUMA device. MREQ# and MGNT# are active low signals driven and sampled synchronous to CPUCLK common to both core logic and VUMA device.

Core logic is always the default owner and ownership will be transferred to VUMA device upon demand. VUMA device could return ownership to core logic upon completion of its activities or park on the bus. Core logic can always preempt VUMA device from the bus.

VUMA device needs to access the physical system memory for different reasons and the level of urgency of the needed accesses varies. If VUMA device is given the access to the physical system memory right away, every time it needs, the CPU performance will suffer and as it may not be needed right away by the VUMA device, there would not be any improvement in VUMA device performance. Hence two levels of priority are defined viz. low priority and high priority. Both priorities are conveyed to core logic through a single signal, MREQ#.

3.2 Memory Regions

As shown in Figure 3-1, physical system memory can contain two separate physical memory blocks, Main VUMA Memory and Auxiliary (Aux) VUMA Memory. As cache coherency for Main VUMA Memory and Auxiliary VUMA Memory is handled by this standard, a VUMA device can access these two physical memory blocks without any separate cache coherency considerations. If a VUMA device needs to access other regions of physical system memory, designers need to take care of cache coherency.

Main VUMA Memory is programmed as non-cacheable region to avoid cache coherency overhead. How Main VUMA Memory is used depends on the type of VUMA device; e.g., when VUMA device is a graphics controller, main VUMA memory will be used for Frame buffer.

Auxiliary VUMA Memory is optional for both core logic and VUMA device. If supported, it can be programmed as non-cacheable region or write-through region. Auxiliary VUMA Memory can be used to pass data between core logic and VUMA device without copying it to Main VUMA Memory or passing through a slower PCI bus. This capability would have significant advantages for more advanced devices. How Auxiliary VUMA Memory is used depends on the type of VUMA device e.g. when VUMA device is a 3D graphics controller, Auxiliary VUMA memory will be used for texture mapping.

When core logic programs Auxiliary VUMA Memory area as non-cacheable, VUMA device can read from or write to it. When core logic programs Auxiliary VUMA Memory area as write through, VUMA device can read from it but can not write to it.

Both core logic and VUMA device have an option of either supporting or not supporting the Auxiliary VUMA Memory feature. Whether Auxiliary VUMA memory is supported or not should be transparent to an application. The following algorithm explains how it is made transparent. The algorithm is only included to explain the feature. Refer to the latest VUMA VESA BIOS Extensions for the most updated BIOS calls:

1. When an application needs this feature, it needs to make a BIOS call, <Report VUMA

- core logic capabilities (refer to VUMA VESA BIOS Extensions)>, to find out if core logic supports the feature.
2. If core logic does not support the feature, the application needs to use some alternate method.
3. If core logic supports the feature, the application can probably use it and should do the following:

a. Request the operating system for a physically contiguous block of memory of required size.
b. If not successful in getting physically contiguous block of memory of required size, use some alternate method.
c. If successful, get the start address of the block of memory.
d. Read <VUMA BIOS signature string (refer to VUMA VESA BIOS Extensions)>, to find out if VUMA device can access the bank in which Auxiliary VUMA Memory has been assigned.
e. If VUMA device can not access that bank, the application needs to either retry the procedure from "step a" to "step c" till it can get Auxiliary VUMA Memory in a VUMA device accessible bank or use some alternate method.
f. If VUMA device can access that bank, make a BIOS call function <Set (Request) VUMA Auxiliary memory (refer to VUMA VESA BIOS Extensions)>, to ask core logic to flush Auxiliary VUMA Memory block of the needed size from the start address from "step c" and change it to either non-cacheable or write through. How a core logic flushes cache for the block of memory and programs it as non-cacheable/ write through is implementation specific.
g. Use VUMA Device Driver, to give VUMA device the Auxiliary VUMA Memory parameters viz. size, start address from "step c" and whether the block should be non-cacheable or write through.

3.3 Physical Connection

A VUMA device can be connected in two ways:

1. VUMA device can only access one bank of physical system memory - VUMA device is connected to a single bank of physical system memory. In case of Fast Page Mode, EDO and BEDO VUMA device has a single RAS#. In case of Synchronous DRAM VUMA device has a single CS#. Main VUMA memory resides in this memory bank. If supported, Auxiliary VUMA Memory can only be used if it is assigned to this bank.

2. VUMA device can access all of the physical system memory - VUMA device has as many RAS# (for Fast Page Mode, EDO and BEDO)/CS# (for Synchronous DRAM) lines as core logic and is connected to all banks of the physical system memory. Both Main VUMA memory and Auxiliary VUMA Memory (if supported) can be assigned to any memory bank.

4.0 Arbitration

4.1 Arbitration Protocol

There are three signals establishing the arbitration protocol between core logic and VUMA device. MREQ# signal is driven by VUMA device to core logic to indicate it needs to access the physical system memory bus. It also conveys the level of priority of the request. MGNT# is driven by core logic to VUMA device to indicate that it can access the physical system memory bus. Both MREQ# and MGNT# are driven synchronous to CPUCLK.

As shown in Figure 4-1, low level priority is conveyed by driving MREQ# low. A high level priority can only be generated by first generating a low priority request. As shown in Figure 4-2 and Figure 4-3, a low level priority is converted to a high level priority by driving MREQ# high for one CPUCLK clock and then driving it low.

Figure 4-1 Low Level Priority

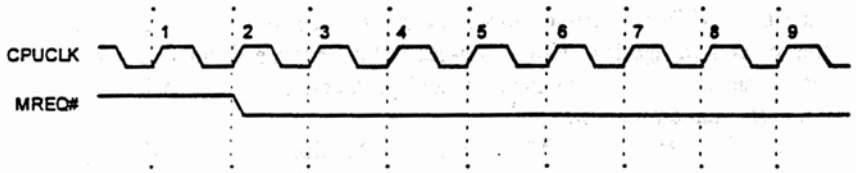


Figure 4-2 High Level Priority

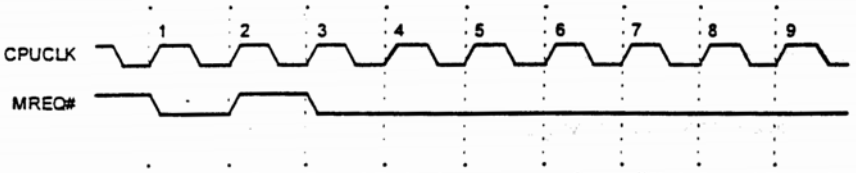
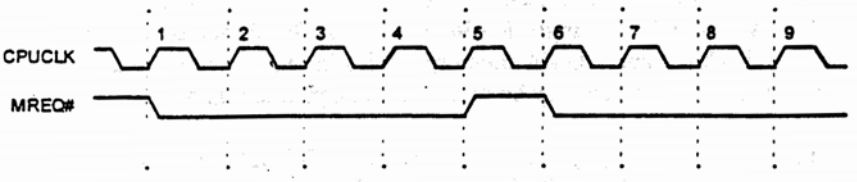


Figure 4-3 A Pending Low Level Priority converted to a High Level Priority



4.2 Arbiter

The arbiter, housed in core logic, needs to understand the arbitration protocol. State

Machine for the arbiter is depicted in Figure 4.4. As shown in Figure 4.4, the arbiter State Machine is resetted with PCI_Reset. Explanation of the arbiter is as follows:

1. HOST State - The physical system memory bus is with core logic and no bus request from VUMA device is pending.

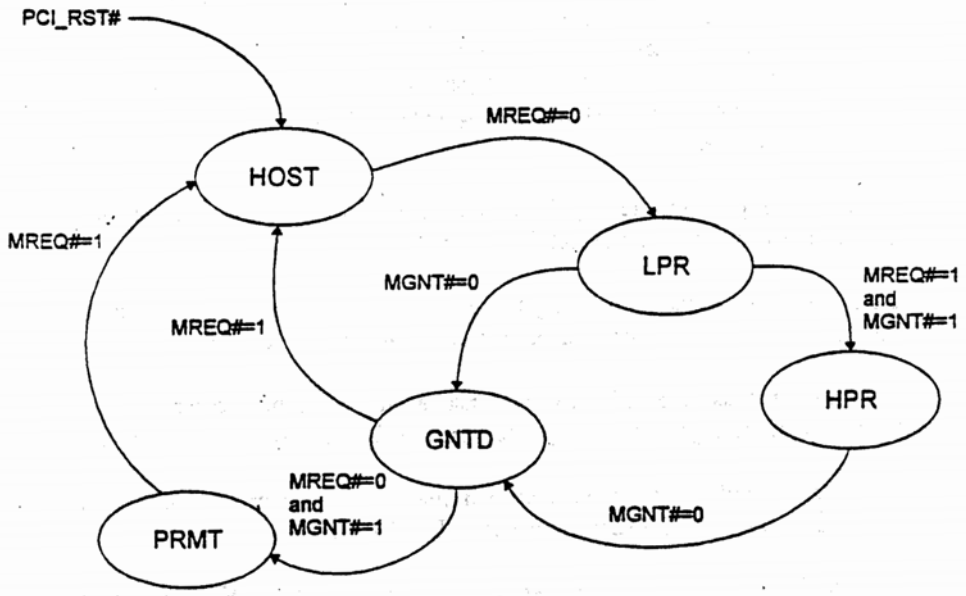
2. Low Priority Request (LPR) State - The physical system memory bus is with core logic and a low priority bus request from the VUMA device is pending.

3. High Priority Request (HPR) State - The physical system memory bus is with core logic and a pending low priority bus request has turned into a pending high priority bus request.

4. Granted (GNTD) State - Core logic has relinquished the physical system memory bus to VUMA device.

5. Preempt (PRMT) State - The physical system memory bus is owned by VUMA device, however, core logic has requested VUMA device to relinquish the bus and that request is pending.

Figure 4.4 Arbiter State Machine



Note:

1. Only the conditions which will cause a transition from one state to another have been

noted. Any other condition will keep the state machine in the current state.

4.2.1 Arbitration Rules

1. VUMA device asserts MREQ# to generate a low priority request and keeps it asserted until the VUMA device obtains ownership of the physical system memory bus through the assertion of MGNT#, unless the VUMA device wants to either raise a high priority request or raise the priority of an already pending low priority request. In the later case,

   a. If MGNT# is sampled asserted the VUMA device will not deassert MREQ#. Instead, the VUMA device will gain physical system memory bus ownership and maintain MREQ# asserted until it wants to relinquish the physical system memory bus.

   b. If MGNT# is sampled deasserted, the VUMA device will deassert MREQ# for one clock and assert it again irrespective of status of MGNT#. After reassertion, the VUMA device will keep MREQ# asserted until physical system memory bus ownership is transferred to the VUMA device through assertion of MGNT# signal.

2. VUMA device may assert MREQ# only for the purpose of accessing the unified memory area. Once asserted, MREQ# should not be deasserted before MGNT# assertion for any reason other than raising the priority of the request (i.e., low to high). No speculative request and request abortion is permitted. If MREQ# is deasserted to raise the priority, it should be reasserted in the next clock and kept asserted until MGNT# is sampled asserted.

3. Once MGNT# is sampled asserted by VUMA device, it gains and retains physical system memory bus ownership until MREQ# is deasserted.

4. The condition, VUMA device completing its required transactions before core logic needing the physical system memory bus back, can be handled in two ways:

   a. VUMA device can deassert MREQ#. In response, MGNT# will be deasserted in the next clock edge to change physical system memory bus ownership back to core logic.
   b. VUMA device can park on the physical system memory bus. If core logic needs the physical system memory bus, it should preempt VUMA device.

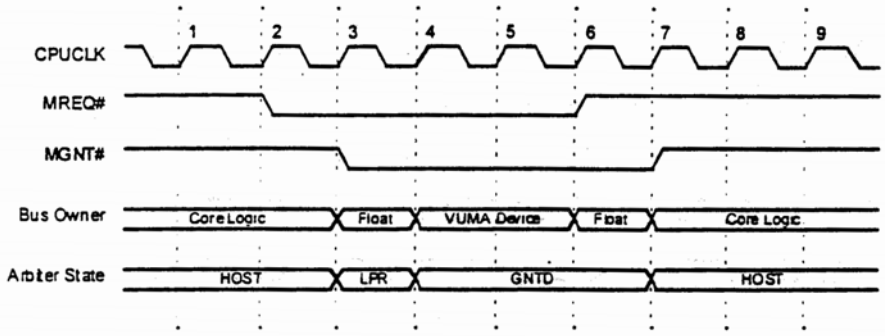
5. In case core logic needs the physical system memory bus before VUMA device releases it on its own, arbiter can preempt VUMA device from the bus. Preemption is signaled to VUMA device by deasserting MGNT#. VUMA device can retain ownership of the bus for a maximum of 60 CPUCLK clocks after it has been signaled

to preempt. VUMA device signals release of the physical system memory bus by deasserting MREQ#.

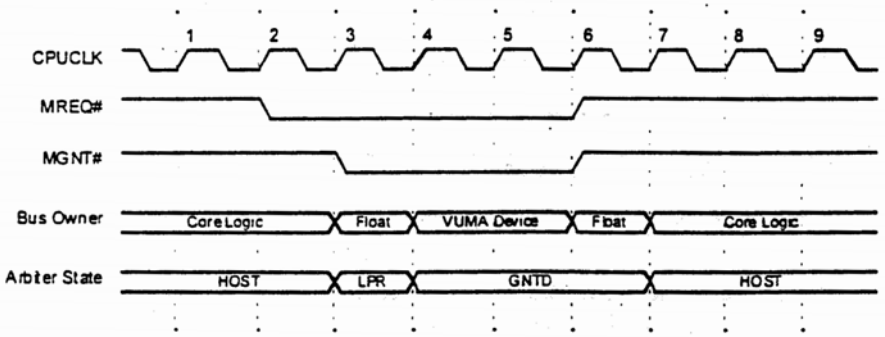
6. When VUMA device deasserts MREQ# to transfer bus ownership back to core logic, either on its own or because of a preemption request, it should keep MREQ# deasserted for at least two clocks of recovery time before asserting it again to raise a request.

4.3 Arbitration Examples

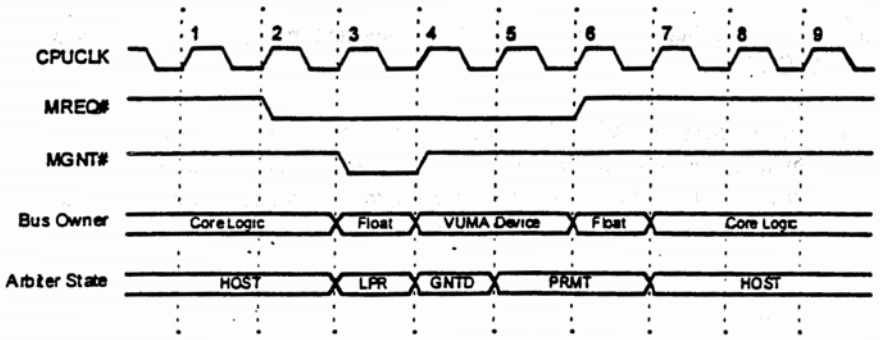
1. Low priority request and immediate bus release to VUMA device



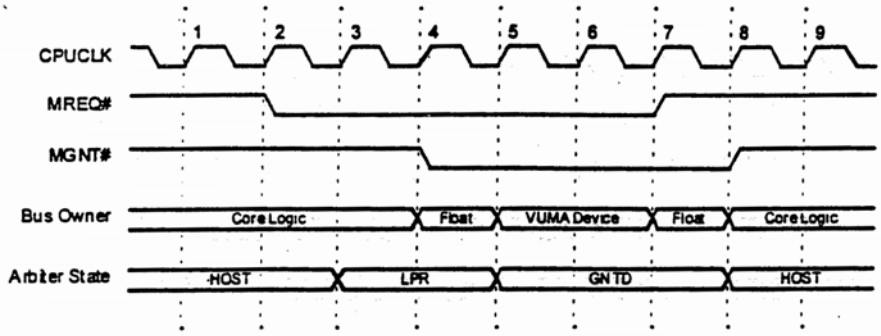
2. Low priority request and immediate bus release to VUMA device with preemption where removal of MGNT# and removal of MREQ# coincide



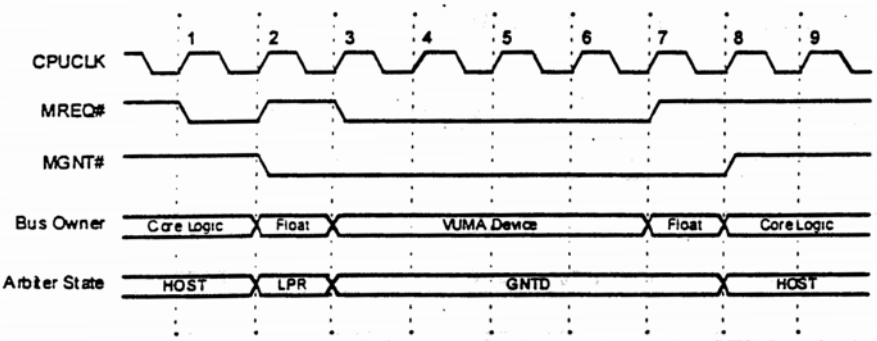
3. Low priority request and immediate bus release to VUMA device with preemption where MREQ# is removed after the current transaction because of preemption



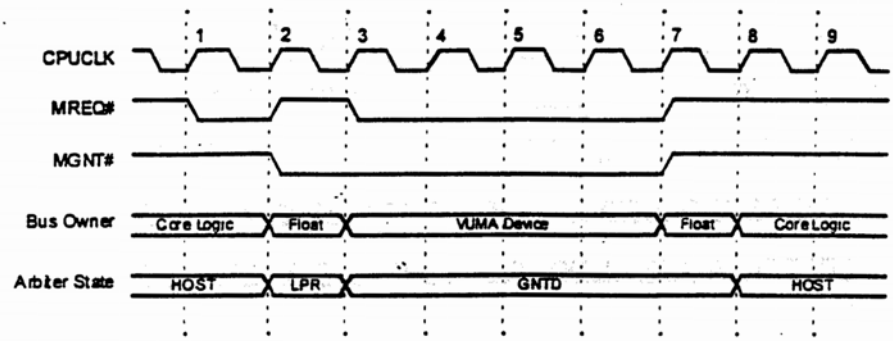
4. Low priority request and delayed bus release to VUMA device



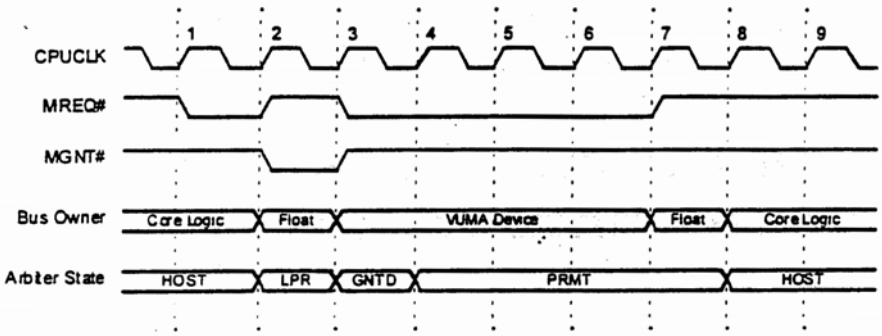
5. High priority request and immediate bus release to VUMA device



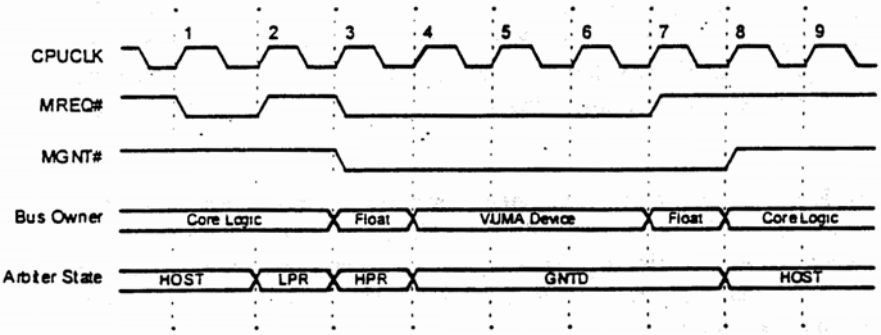
6. High priority request and immediate bus release to VUMA device with preemption where MGNT# and MREQ# removal coincides.



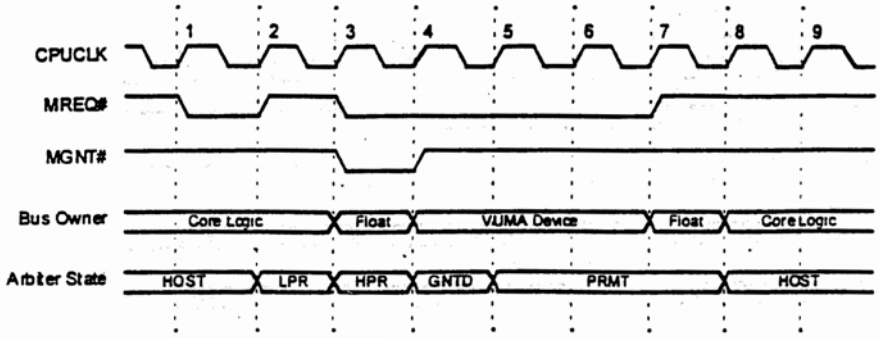
7. High priority request and immediate bus release to VUMA device with preemption where MREQ# is removed after the current transaction because of preemption.



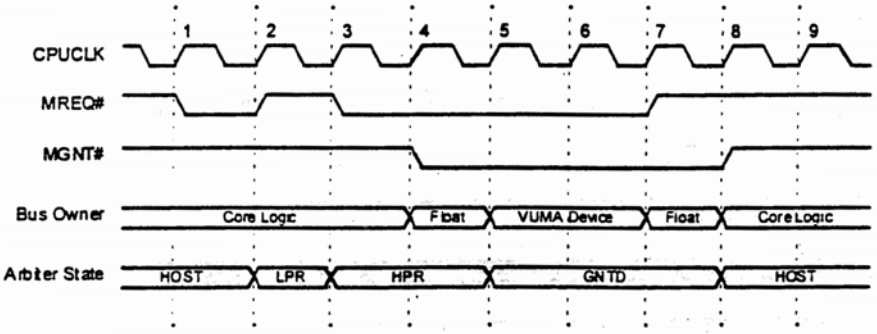
8. High priority request and one clock delayed bus release to VUMA device



9. High priority request and one clock delayed bus release to VUMA device with preemption where MREQ# and MGNT# removal do not coincide



10. High priority request and delayed bus release to VUMA device



4.4 Latencies

1. High Priority Request - Worst case latency for VUMA device to receive a grant after generating a high priority request is 35 CPUCLK clocks, i.e. after arbiter receives a high priority request from VUMA device, core logic does not need to relinquish the physical system memory bus right away and can keep the bus for up to 35 CPUCLK clocks.

2. Low Priority Request - No worst case latency number has been defined by this specification for low priority request. VUMA devices should incorporate some mechanism to avoid a low priority request being starved for an unreasonable time. The mechanism is implementation specific and not covered by the standard. One simple reference solution is as follows:

   VUMA device incorporates a programmable timer. The timer value is set at the boot time. The timer gets loaded when a low priority request is generated. When the timer times out, the low priority request is converted to a high priority request.

3. Preemption Request to VUMA device - Worst case latency for VUMA device to relinquish the physical system memory bus after receiving a preemption request is 60

CPUCLK clocks, i.e. after core logic requests VUMA device to relinquish the physical system memory bus, VUMA device does not need to relinquish the bus right away and can keep the bus for up to 60 CPUCLK clocks.

Design engineers should take in to consideration the above latencies for deciding FIFO sizes.

5.0 Memory Interface

The standard supports Fast Page Mode, EDO, BEDO and Synchronous DRAM technologies.

DRAM refresh for the physical system memory including Main VUMA Memory and Auxiliary VUMA Memory is provided by core logic during normal as well as suspend state of operation.

If VUMA device uses only a portion of its address space as Main VUMA Memory or Auxiliary VUMA Memory, it should drive unused upper MA address lines high.

5.1 Memory Decode

The way CPU address is translated in to DRAM Row and Column address decides the physical location in DRAM where a particular data will be stored. In the conventional architecture this could be implementation specific as there is a single DRAM controller. In unified memory architecture, multiple DRAM controllers (core logic resident and VUMA device resident DRAM controller) need to access the same data. Hence, all DRAM controllers should follow the same translation of CPU address into DRAM Row and Column address. The translation is as shown in Table 5-1.

Table 5-1 Translation of CPU address to DRAM Row and Column addresses

Symmetrical x9, x10, x11, x12

	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn	A26	A24	A22	A11	A10	A9	A8	A7	A6	A5	A4	A3
row	A25	A23	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12

Asymmetrical x8

	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn					A10	A9	A8	A7	A6	A5	A4	A3
row	A22	A21	A11	A20	A19	A18	A17	A16	A15	A14	A13	A12

Asymmetrical x9

	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn				A11	A10	A9	A8	A7	A6	A5	A4	A3
row	A23	A22	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12

Asymmetrical x10

	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn			A22	A11	A10	A9	A8	A7	A6	A5	A4	A3
row	A24	A23	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12

Asymmetrical x11

	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn		A24	A22	A11	A10	A9	A8	A7	A6	A5	A4	A3
row	A25	A23	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12

Synchronous 16Mb

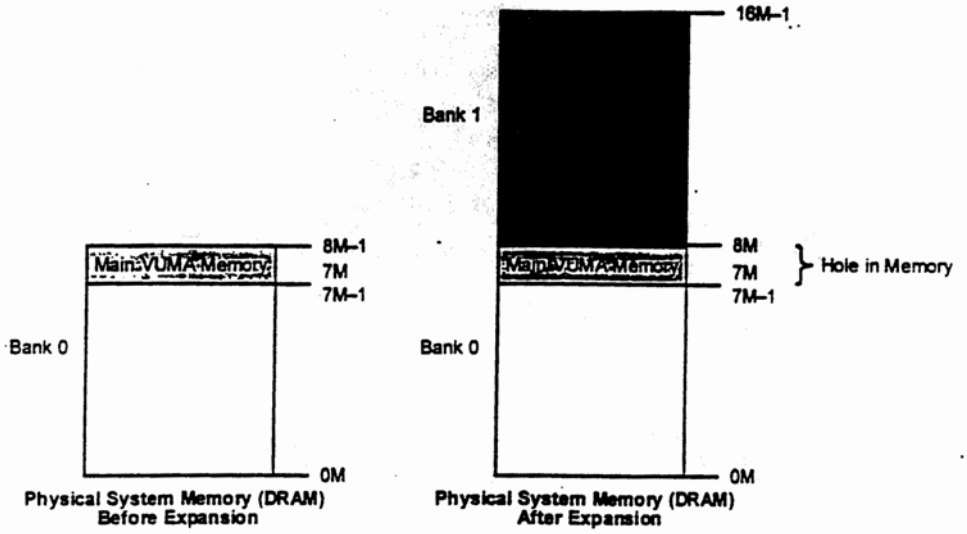
	MA11	MA10	MA9	MA8	MA7	MA6	MA5	MA4	MA3	MA2	MA1	MA0
clmn	A11		A24	A23	A10	A9	A8	A7	A6	A5	A4	A3
row	A11	A22	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12

5.2 Main VUMA Memory Mapping

When physical system memory (DRAM) is expanded, unified memory architecture poses a unique problem not existing on the conventional architecture. The problem and three different solutions are described below:

Problem: Main VUMA Memory needs to be mapped at the top of existing memory for any given machine. When physical system memory (DRAM) is expanded, this would cause a hole in the physical system memory as shown in Figure 5-1. The example assumes an initial system with single bank 8MB memory (1MB allocated to Main VUMA Memory) expanded to 16MB memory (1MB allocated to Main VUMA Memory) by adding a bank of 8MB memory. All the numbers mentioned in this discussion are just examples and do not imply to be a part of the standard.

Figure 5-1 Memory Expansion Problem

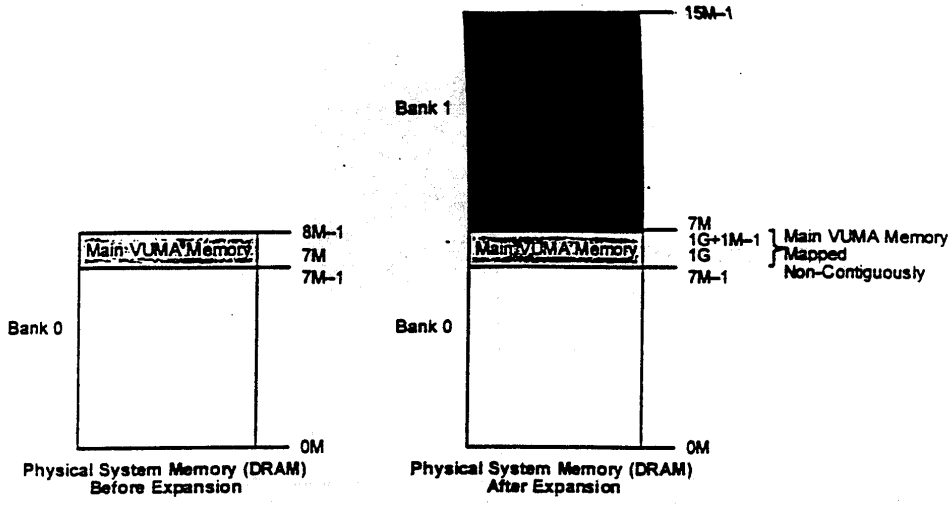


Three solutions are suggested for this problem. BIOS calls defined in VUMA VESA BIOS Extensions support all the three solutions. The BIOS calls are designed in such a way that a VUMA device can find out which of the three solutions is implemented by core logic and can configure the VUMA device appropriately.

Solution 1:

As depicted in Figure 5-2, core logic maps Main VUMA Memory to an address beyond core logic's possible physical system memory range. Main VUMA Memory is mapped non-contiguous to the O.S. memory. As shown in Figure 5-2, Main VUMA Memory is mapped from 1G to (1G+1M-1) and hence even if physical system memory is expanded to the maximum possible size, there will be no hole in the memory. As shown in Figure 5-2. Bank 0 is split with two separate blocks of memory with different starting addresses. If the VUMA device is a graphics controller, and if it wants to look at Main VUMA Memory also as a PCI address space, it can allocate a different address than what has been assigned by core logic (1G in this example).

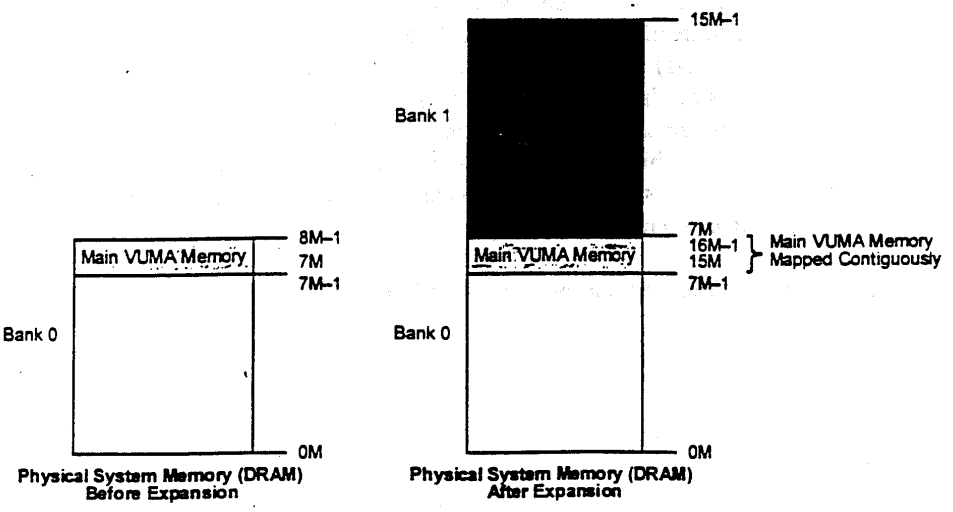
Figure 5-2 Main VUMA Memory mapped non-contiguously



Solution 2:

As depicted in Figure 5-3, core logic maps Main VUMA Memory to the top of memory. Main VUMA Memory is mapped contiguous to the O.S. memory. As shown in Figure 5-3, Main VUMA Memory is mapped from 15 M to (16M-1). As shown in Figure 5-3, Bank 0 is split with two separate blocks of memory with different starting addresses.

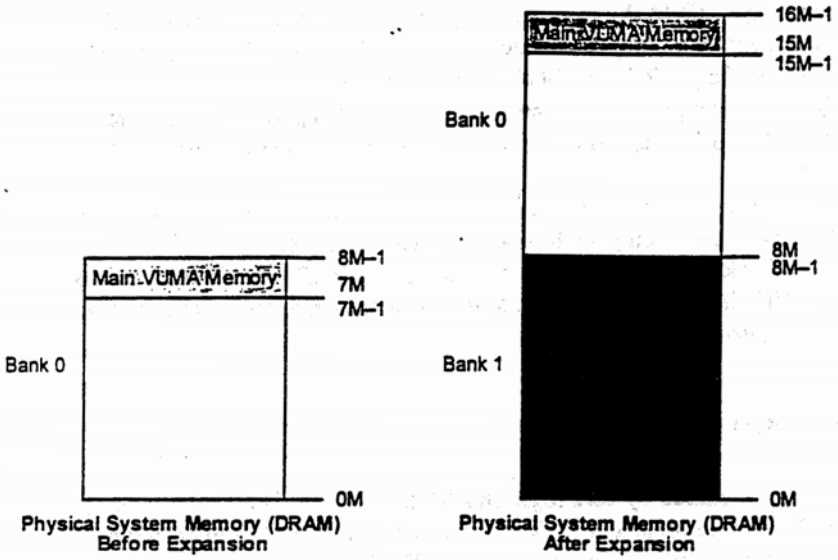
Figure 5-3 Main VUMA Memory mapped contiguously



Solution 3:

As depicted in Figure 5-4, core logic swaps the bank containing main VUMA Memory to the top of memory. As shown in Figure 5-4, Bank 0 is not split with two separate blocks of memory with different starting addresses like in solution 1 and solution 2.

Figure 5-4 Main VUMA Memory bank swapped



5.3 Fast Page EDO and BEDO

The logical interfaces for Fast Page, EDO and BEDO DRAMs are very similar and hence are grouped together. If no specific exception to a particular technology is mentioned, the description in this section applies to all the three types of DRAMs.

BEDO support is optional for both core logic and VUMA device. Various BEDO support scenarios are as follows:

1. Core logic does not support BEDO - Since core logic does not support BEDO, there will not be any BEDO as the physical system memory and hence whether VUMA device supports BEDO or not is irrelevant.

2. Core logic supports BEDO - When core logic supports BEDO, VUMA device may or may not support it. Whether core logic and VUMA device support BEDO or not should be transparent to the operating system and application programs. To achieve the transparency, system BIOS needs to find out if both core logic and VUMA device support this feature and set the system appropriately at boot. The following algorithm

explains how it can be achieved. The algorithm is only included to explain the feature. Refer to the latest VUMA VESA BIOS Extensions for the most updated BIOS calls:

a. Read <VUMA BIOS signature string (refer to VUMA VESA BIOS Extensions)>. Check if VUMA device supports BEDO.
b. If VUMA device does not support BEDO, do not assign BEDO banks for Main VUMA Memory. Assign Main VUMA Memory to Fast Page Mode or EDO bank. Also, if Auxiliary VUMA Memory is assigned by operating system to BEDO banks, do not use it. Either repeat the request for Auxiliary VUMA Memory till it is assigned to Fast Page Mode or EDO bank or use some alternate method.
c. If VUMA device supports BEDO, read <VUMA BIOS signature string (refer to VUMA VESA BIOS Extensions)> to find out if VUMA device supports multiple banks access.
d. If only single bank access supported on VUMA device, exit, as the Main VUMA Memory and Auxiliary VUMA Memory bank is fixed.
e. If multiple banks access is supported and if the RAS for BEDO bank is supported on VUMA device, assign the Main VUMA Memory to obtain the best possible system performance and exit.

5.3.1 Protocol Description and Timing

All the DRAM signals are shared by core logic and VUMA device. They are driven by current bus master. When core logic and VUMA device hand over the bus to each other, they must drive all the shared s/t/s signals high for one CPUCLK clock and then tri-state them. Also, they should tri-state all the shared t/s signals.

The shared DRAM signals are driven by core logic when it is the owner of the physical system memory bus. VUMA device requests the physical system memory bus by asserting MREQ#. Bus Arbiter grants the bus by asserting MGNT#. Also, as mentioned above, before VUMA device starts driving the bus, core logic should drive the s/t/s signals high for one CPUCLK clock and tri-state them. Core logic should also tri-state all the shared t/s signals. The float condition on the bus should be for one CPUCLK clock, before VUMA device starts driving the bus. These activities are overlapped to improve performance as shown in Figure 5-5.

Figure 5-5 Bus hand off from core logic to VUMA device