

M I C R O P R O C E S S O R

www.MPRonline.com

THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

BENCHMARKING THE BENCHMARKS

Ever Controversial, Embedded-CPU Benchmarks Make Fitful Progress

By Tom R. Halfhill (8/30/04-01)

We live in a season of divisive partisan politics: endless bickering, blame games, finger pointing, strident propaganda, arguments over strategy, and embarrassing scandals. And that's just the politics of microprocessor benchmarking.

An outsider might reasonably wonder what all the fuss is about. Isn't measuring the speed of a microprocessor as easy as measuring the speed of an automobile? Just use the electronic equivalent of a speedometer and be done with it. Of course, it isn't that simple. If the marketing vice presidents from three different processor vendors were randomly selected and put in the same room, they probably wouldn't agree on whether the best instrument of measurement is a speedometer, tachometer, or barometer.

For years, the embedded-processor industry limped along with the Dhrystone MIPS (DMIPS) benchmark, the descendant of a relatively simple 1984 Ada program whose reference machine was a DEC VAX 11/780 minicomputer. Even its author, Reinhold Weicker, has disavowed the benchmark's usefulness for evaluating anything designed in the past 10 years. The greatest leap beyond Dhrystone has been the Embedded Microprocessor Benchmark Consortium (EEMBC, pronounced "embassy"). Founded in 1997 by Markus Levy, EEMBC currently has 58 member companies, including almost all the biggest names in the business. EEMBC revolutionized microprocessor benchmarking with its strict bylaws, democratic business model, application-oriented test suites, provisions for optimizing benchmark code, and rigorous score-certification process. (See *MPR* 5/1/00-02, "EEMBC Releases First Benchmarks.")

However, even EEMBC hasn't been a panacea. Seven years after the consortium was founded, fewer than half its members have published any benchmark scores for public

scrutiny. Most companies prefer to keep their scores private or share them with customers only, under an NDA. Two relatively new benchmark suites—one for measuring Java performance, another for testing 8/16-bit microcontrollers—have stagnated after garnering a mere handful of published scores, despite years of internal development. Creating new benchmarks and revising existing suites have proved to be arduous tasks, because everything must be decided by committees of member companies fiercely competing with each other in the marketplace. Some new suites have been under development or delayed for as long as three years.

Recently, a Texas entrepreneur connected with EEMBC decided to fill a different benchmarking niche, using a less cumbersome business model. Alan R. Weiss's Austin-based startup, Synchromesh Computing, has introduced a new benchmark suite for testing x86-compatible processors—specifically, x86 processors suitable for low-end PCs, thin clients, high-end set-top boxes, and Internet appliances. Those applications are at the crossroads between PC processors (which already have a plethora of popular benchmark programs) and embedded processors, which are EEMBC's traditional territory. Although Synchromesh Computing calls its benchmark suite the Embedded Processor Rating System (EPRS), its focus on system-level tests and higher-end applications differs from EEMBC's mission.

Synchromesh Computing has stitched together a composite suite by choosing off-the-shelf benchmark programs whose source code is publicly available and by writing some

new tests. As a for-profit private company, not an industry consortium, Synchromesh Computing is unimpeded by committee meetings and a multivendor board of directors. Weiss's connection to EEMBC is that he also owns EEMBC Certification Laboratories (ECL), EEMBC's exclusive lab for certifying EEMBC benchmark scores.

Weiss's first client for EPRS benchmarking was AMD, which funded the development of the new suite. (Synchromesh Computing also has some clients for its other services.) The lab's tests of AMD's x86-compatible Geode processors have provoked howls of protest from VIA, a rival x86 vendor chasing the same markets that AMD does. Controversy seems inseparable from the science and art of benchmarking, often derided as "benchmarking."

Even *Microprocessor Report* isn't immune from the politics. In the interest of full disclosure, note that both Alan Weiss and EEMBC president Markus Levy are members of the *MPR* editorial board. Also, the author of this article, when employed at ARC International, was a voting representative on the EEMBC board of directors.

Legitimizing Embedded Benchmarking

Despite some troubles, EEMBC has been an impressive success. Of course, almost anything would have been an improvement over Dhrystone, but EEMBC is unique in three ways: it focuses exclusively on embedded processors; it requires members to submit their test results to an independent certification lab before sharing the scores outside the

company; and it gives members broad leeway to optimize the benchmark source code for their processors.

By concentrating on embedded processors, EEMBC can fine-tune its benchmark tests for the most popular embedded applications. When EEMBC released its first benchmarks in 2000, there were five suites: auto/industrial, consumer, networking, office automation, and telecommunications. As Table 1 shows, the largest suite was auto/industrial, which had 16 individual "kernels" or test programs. The slimmest suite was networking, which had only three kernels.

Technically, the EEMBC tests are synthetic benchmarks, because they aren't real embedded applications. However, they contain algorithms and routines commonly used in real applications, so they are a better measure of performance than truly synthetic benchmark tests like the integer-math loops in Dhrystone. The kernels in each EEMBC suite are written from scratch or donated from actual applications by member companies on EEMBC's technical committees. Because the job of defining and developing new kernels is a committee process, achieving a consensus is tedious and time consuming, but the results are respected.

EEMBC's benchmark tests are useful for purposes other than evaluating embedded processors. For example, by compiling the kernels with two different compilers and running the programs on the same processor, software developers can compare the relative efficiency of the compilers. Depending on the developers' priorities, it's possible to compare execution speed (how fast the compiled code runs) or code density (the size of the executable files when compiled from the same source code). Among EEMBC's members are vendors of software-development tools, such as Green Hills, MetaWare, MetroWerks, Red Hat, and Wind River.

Another unique feature of EEMBC is its rulebook. EEMBC is governed by a constitution that itself took years to create, even before the first line of benchmark code was written. One unusual rule is that only EEMBC members have access to the benchmark code. Unlike many other benchmark programs, you can't simply download the EEMBC suites from the Internet. This rule keeps nonmembers from publishing uncertified or inaccurate scores. Along with the right to participate in the benchmark-definition process, access to the 1.5 million lines of source code is a powerful incentive for joining EEMBC. Annual dues are \$7,500 to \$30,000, depending on the level of membership. EEMBC uses its income for further benchmark development, marketing, and management.

EEMBC does publish datasheets describing the way the benchmark kernels work—anyone can download them from the EEMBC website—but the source code remains under lock and key. Some critics object that the secrecy surrounding EEMBC's source code stops outsiders (such as *MPR* analysts) from evaluating the kernels for technical relevance and susceptibility to cheating. Even at *MPR*, our analysts are divided over the importance of this issue. However, it's a valid criticism. EEMBC could enhance its already good reputation by allowing trusted outsiders to examine the benchmark source

EEMBC 1.0 Benchmark Suites	
Auto/Industrial Suite	
Angle-to-time conversion	Inv discrete cosine transform
Basic floating point	Inverse FFT filter
Bit manipulation	Matrix arithmetic
Cache buster	Pointer chasing
CAN remote data request	Pulse-width modulation
Fast-Fourier transform (FFT)	Road speed calculation
Finite impulse resp (FIR) filter	Table lookup and interpolation
Infinite impulse resp (IIR) filter	Tooth-to-spark calculation
Consumer Suite	
Compress JPEG	RGB-to-CMYK conversion
Decompress JPEG	RGB-to-YIQ conversion
High-pass grayscale filter	
Networking Suite	
OSPF / Dijkstra routing	Packet flow (1MB)
Lookup / Patricia algorithm	Packet flow (2MB)
Packet flow (512B)	
Office Automation Suite	
Bezier-curve calculation	Image rotation
Dithering	Text processing
Telecommunications Suite	
Autocorrelation (3 tests)	Fixed-pt complex FFT (3 tests)
Convolutional encoder (3 tests)	Viterbi GSM decoder (4 tests)
Fixed-point bit alloc (3 tests)	

Table 1. These are the original EEMBC 1.0 benchmark suites. EEMBC announced the first ECL-certified scores based on these suites in April 2000. The benchmark suites remain largely the same today.

EEMBC's Networking 2.0 Benchmarks Worth the Wait

EEMBC has significantly upgraded the benchmark tests in its networking suite. After years of labor by ECL and consortium members, EEMBC released the first certified scores for the Networking 2.0 suite on August 9. The revised suite greatly expands the coverage of the benchmarks and addresses an application category that has become much more important in the seven years since EEMBC was founded. It's also an application category that has frustrated other attempts at benchmarking.

The original suite had only three kernels: a packet-routing test that used the OSPF (open shortest path first) Dijkstra algorithm; a network-address lookup test that used the Patricia algorithm; and a packet-flow test that used datasets of three different sizes (512KB, 1MB, and 2MB). It wasn't a bad benchmark suite, but it clearly needed to cover more ground.

Networking 2.0 has nine kernels. Gone is the overall NetMark score of the original suite. It has been replaced with two figures of merit: TCPmark and IPmark. TCPmark describes performance on tasks related to the Transmission Control Protocol (TCP); IPmark describes performance on tasks related to the Internet Protocol (IP). TCPmark and IPmark scores cannot be directly compared with NetMark scores.

Holdovers from the Networking 1.0 suite are the packet-routing (OSPF) and Patricia lookup tests, which contribute to IPmark. The data-flow test is replaced by a similar packet-check test, which evaluates packet-switching performance by operating on the packet headers in four different datasets. The geometric mean of those four results is one element of IPmark. Other kernels that contribute to IPmark are tests for translating IP addresses, the fragmentation and reassembly of IP packets, and a quality-of-service (QoS) test that simulates the rules used by

bandwidth-management software in routers. IPmark is the geometric mean of the six IP-oriented kernels.

Three kernels contribute to TCPmark, another geometric mean. The Telnet kernel simulates the small, short bursts of bidirectional network traffic generated by a Telnet command-line session. The FTP kernel simulates the large amount of unidirectional data traffic when using the File Transfer Protocol. The HTTP kernel is somewhere in between—it simulates web browsing, with bursts of files downloading to a client using the Hypertext Transfer Protocol.

Freescale Semiconductor and IBM Microelectronics are the first EEMBC members to publish Networking 2.0 scores. Both processors are based on the PowerPC architecture, and both companies used Green Hills Multi 4.0 to compile the benchmark code, which makes their out-of-the-box scores more comparable. Freescale's 1.4GHz MPC7447A easily surpassed IBM's 1.0GHz 750GX in the TCP tests, scoring a TCPmark of 819.8 vs. 467.1. But IBM turned the tables in the IP tests, edging out the higher-frequency Freescale processor with an IPmark of 286.1 vs. 245.1.

The MPC7447A's AltiVec extensions and wider parallelism probably contributed to its stellar TCPmark, because it can execute up to four instructions per clock cycle (three plus a branch), compared with only two instructions per clock for the 750GX. In the IPmark tests, the 750GX compensated for that shortcoming and its 40% slower clock speed by leveraging a faster memory bus (200MHz vs. 167MHz) and a larger L2 cache (1MB vs. 512KB).

It took a long time for EEMBC to revise the networking suite, but the work was productive. We judge the upgraded benchmarks worth the wait. However, EEMBC will have to move faster in the future to keep up with this fast-changing field.

code under an NDA. Lately, the consortium is starting to loosen up somewhat; memberships have been offered to OEM companies, sometimes for free.

Foiling the Benchmark Cheaters

To prevent the kind of cheating that has caused scandals with many other benchmarks, EEMBC's rules require members to certify their test results at an independent lab (ECL) before showing the scores outside the company. ECL performs more than 50 checks to find anomalies and outright cheating, even to the point of eyeballing the compiled benchmark code and output data. For more than five years, EEMBC and ECL have avoided being victimized by the ploys that have made a mockery of some benchmarks.

Perhaps the cleverest way EEMBC discourages cheating is by encouraging a legal form of "cheating." EEMBC allows members to report two types of results: "out-of-the-box"

scores, based on unmodified benchmark source code, and "full-fury" scores, based on optimized source code. To obtain out-of-the-box scores, members can use any publicly available compiler and built-in compiler switches to compile the source code, but they cannot change the source code. These scores represent the best performance a real embedded-system developer could obtain from a processor without doing much work. However, EEMBC members must disclose the compiler and switches they used to obtain their out-of-the-box scores so anyone can duplicate the results.

Full-fury scores are often more interesting, especially when compared with out-of-the-box scores for the same processor. (A vendor must report out-of-the-box scores along with full-fury scores, but full-fury benchmarking is optional.) EEMBC members can rewrite the C/C++ benchmark source code, replace high-level code with assembly language, substitute whole sections of source code with calls to

special application-specific hardware in the processor, or—with configurable processors—create entirely new CPU instructions to accelerate the kernels. In effect, EEMBC legalizes almost all the tricks commonly used to subvert other benchmarks.

At first glance, it looks outrageous, but it makes sense. EEMBC merely recognizes the techniques that real-world developers use to design efficient embedded systems. Of course, there are still some rules. For instance, members must disclose their optimizations under an NDA to ECL, which verifies that the modified kernels can still perform their intended tasks on the target datasets. (For some tests that don't require an exact result, such as MPEG compression, EEMBC uses a statistical method to measure the deviation from theoretical perfection.) Each EEMBC benchmark suite is governed by a technical committee that sets its own rules for full-fury benchmarking, so the freedom to optimize varies somewhat. In the telecommunications suite, members may rewrite the algorithms; in the automotive/industrial suite, they may not.

With competing benchmark suites, many optimization techniques used to obtain EEMBC's full-fury scores would be considered out of bounds. Some compilers are known to have special "Dhrystone" switches that reduce whole subroutines in the Dhrystone program to one or a few CPU instructions. At least one graphics-chip vendor has been caught building special hardware into its processor to speed up a popular benchmark. Smart programmers have compressed benchmark routines so they fit entirely within a processor's caches, thereby eliminating memory accesses. Under EEMBC's full-fury rules, it's all legal, as long as the modified code can still perform the defined workloads and the member company discloses its optimizations to ECL. In the real world of embedded-systems development, this kind of "cheating" is not only legal, it is the mark of skilled developers.

Is EEMBC a Victim of Its Own Success?

Alas, nothing is perfect, and EEMBC's flaws have become apparent over the past five years. To begin with, of the consortium's 58 member companies, only 22 have published certified scores. Even though EEMBC's membership continues to grow—the latest addition is AMCC, which recently acquired some embedded PowerPC processors from IBM—the rate of publishing scores has slowed to its lowest level since 2000, when EEMBC released the first batch of benchmark results.

The fact that only 22 companies have published scores doesn't mean only 22 companies have run the benchmarks. To the contrary, almost every member has used the benchmarks internally; under EEMBC's rules, however, ECL certification and publishing are optional. Some companies have certified their results at ECL and share the scores with potential customers under an NDA, but they don't publish the scores for public consumption. Other companies don't even bother with ECL certification and keep their results private. The relative scarcity of published scores is frustrating to

those who hoped EEMBC benchmarks would be quoted as widely as Dhrystone benchmarks are.

There are several reasons why a company would pay thousands of dollars to join EEMBC, perhaps spend hundreds of hours participating in technical committees, maybe even donate source code or help write the benchmarks—and then withhold its benchmark results. One obvious reason is disappointing performance. If your test results revealed your processor is at the tail of the pack, would you publish the scores and trumpet your position to the world? Probably not. But there are less obvious reasons for sequestering results.

Some processors perform very well in a few benchmark tests and poorly in others; publishing the scores might expose the processor's strengths and weaknesses to competitors and customers. Even if a processor finishes at the front of the pack, it's only a matter of time (sometimes only weeks) before another processor surpasses it, which limits the marketing potential. Then, too, some companies believe that sharing their benchmark scores with prospects under an NDA is a good enough reason to join EEMBC; anybody unwilling to sign an NDA and meet with a salesperson probably isn't a serious customer.

In other cases, companies join EEMBC to use the benchmark suites for internal testing and development, not to publish competitive scores. Because EEMBC derives the test kernels from real embedded applications, the kernels are invaluable for evaluating CPU designs and architectures.

Yet another reason companies refrain from publishing scores is that EEMBC's benchmarks focus on relatively narrow aspects of performance, such as computation and data movement. Those are the same aspects of performance commonly tested by benchmark suites for PC and server processors. However, customers shopping for embedded processors tend to care more about other factors, such as power consumption, power-performance ratios, cost-performance ratios, on-chip peripherals, on-chip memory, and—in the case of synthesizable processor cores—ease of design integration.

EEMBC is painfully aware of its problems and has taken steps to address them. A full membership in EEMBC includes two free ECL certifications, normally costing \$3,000 to \$5,000 each. Over some opposition from the more engineering-minded members, EEMBC created summary scores (such as the ConsumerMark) that express the detailed test results for each suite as an easier-to-grasp single figure of merit. (The detailed results are still available in the certified benchmark report.) EEMBC actively promotes the benchmark results in several ways, such as posting hundreds of certified reports on the consortium's website (www.eembc.org) and frequently distributing press releases. Markus Levy is a tireless speaker at industry events and constantly urges EEMBC members to publish more benchmarks. Despite all those efforts and more, most members remain shrinking violets.

New or Improved Benchmarks Can Take Years

Another way EEMBC has sought to strengthen its position is by exploring new territory. Over the past four years, EEMBC has worked hard to expand its benchmark suites and introduce new or improved suites. These efforts, too, have yielded mixed results. Revising the existing suites is a slow process, and two new suites have largely fallen flat in the marketplace.

As an example of how difficult it is for EEMBC to revise benchmarks that members have laboriously developed and adopted, consider what happened with the office-automation suite. Almost immediately after EEMBC released the first certified benchmark scores in April 2000, a problem was discovered in one of the suite's four kernels, a test for calculating Bezier curves. Some C compilers were able to optimize the Bezier code to such a degree that the test became meaningless. EEMBC promptly dropped the Bezier kernel from the suite and removed those test results.

More than four years later, the Bezier kernel is still missing in action. It will probably reappear when the consortium releases the next version of the office-automation suite, by the end of this year. Meanwhile, the suite—primarily intended to measure the performance of embedded processors for printers, scanners, and fax machines—has soldiered on with only three kernels (image rotation, dithering, and text processing). Nevertheless, the suite has been useful for printer companies like Lexmark. The next version—developed by ECL under the guidance of IBM's Ron Olson, chair of EEMBC's office-automation subcommittee—will add a Ghostscript interpreter that emulates a real printer, a major improvement.

Seven years after EEMBC's inception, and four years after EEMBC published the first benchmark scores, only one suite has completed the revision process and can boast of certified scores. EEMBC announced the first benchmark results for version 2.0 of the networking suite on August 9. (See the accompanying sidebar, "EEMBC's Networking 2.0 Benchmarks Worth the Wait.") Two more suites, office automation and digital entertainment, are near adoption. The other suites in the EEMBC 1.0 lineup are still in various stages of revision.

No one expects EEMBC to revise its benchmarks every year. It's hard work to write code that's portable to different 16-, 32-, and 64-bit architectures (including general-purpose microprocessors and DSPs), with big-endian and little-endian memory addressing, and dozens of different tool chains. It's also understandable that members don't want to hasten the obsolescence of their existing benchmark scores. (Revised suites are so different from the suites they replace that the scores aren't directly comparable.) However, four or five years is too long an interval after introducing the version 1.0 suites.

Java and Microcontroller Suites Fall Flat

While EEMBC has struggled to revise its existing benchmarks, the consortium has also tried to establish completely new suites. Last March, EEMBC announced the first benchmark

results for its new Java 2 Micro Edition (J2ME) suite, which took more than two years to develop. The composite score based on the six kernels in this suite is called the GrinderMark.

Sun used the GrinderMark tests to measure the performance of its Connected Limited Device Configuration (CLDC) virtual machine and CLDC HotSpot virtual machine. (CLDC is part of a stripped-down Java runtime environment for small embedded systems—such as cell-phones, PDAs, and point-of-sale terminals—that have as little as 160KB of memory available for Java. The CLDC HotSpot is a more powerful Java runtime environment for embedded systems that have 512KB to 1MB of memory available for Java.) So far, however, no EEMBC member except Sun has published GrinderMark scores. The EEMBC website shows a GrinderMark score for Sharp's Zaurus SL-5500 PDA, which has an Intel StrongARM processor, but that's the system on which Sun benchmarked its CLDC environments.

Another disappointment is EEMBC's 8/16-bit microcontroller suite. Nine months in the making, it borrowed six kernels from EEMBC's auto/industrial suite and added two new kernels: a memory-access test and the "task-based test," a minisuite of nine separate tasks, mostly involving mathematical calculations and data moves. The composite score is called the MicroMark. The prime force behind this project was NEC Electronics, a leading microcontroller vendor. Other EEMBC members agreed the consortium needed 8/16-bit benchmarks to complement the regular benchmarks, which are for 32- and 64-bit embedded processors. After hundreds of hours of committee work and development, the result was a respectable suite of 8/16-bit microcontroller benchmarks. To date, however, only two vendors—NEC and Infineon—have published MicroMark scores.

EEMBC hopes the same fate won't befall a third new benchmark suite that measures performance for digital entertainment applications. This ambitious suite—under development for years by ECL under the guidance of Freescale's Sergei Larin—includes several popular multimedia codecs and three cryptographic algorithms. The difficult technical work is done, and all that remains is for the EEMBC board to formally adopt the suite and the nicknames of the composite scores. (One proposal is to express the results as an EncodeMark, DecodeMark, and CryptoMark, then combine those three minicomposites into an overall score called the DENmark—short for digital entertainment.) Table 2 shows all the new benchmark suites adopted since the EEMBC 1.0 benchmarks debuted in 2000.

Why have the Java and microcontroller suites been greeted with such a lack of enthusiasm? One explanation for the Java snafu is that embedded Java developers—and they are now legion—would rather see benchmarks for the Mobile Information Device Profile (MIDP), which includes the CLDC and is more popular on the latest cellphones and PDAs. Another possibility is that Java performance is less crucial than developers once feared it would be, especially

Benchmark Tests	Performance Measurements	Notes
Networking 2.0		
Telnet	Short, small bursts of bidirectional data from a terminal session	TCPmark test ("Bulk")
FTP	Large amounts of data in large unidirectional packets	TCPmark test ("Jumbo")
HTTP	Unidirectional bursts of files with bidirectional control and handshaking	TCPmark test ("Mixed")
NAT	Network address translation for Internet Protocol addresses	IPmark test ("NAT")
IP Processing	Fragmentation and reassembly of IP packets	IPmark test ("IPReass")
QoS	Quality-of-service processing to measure data-transfer rates; predetermined rules simulate bandwidth-management software	IPmark test ("QoS")
OSPF	Open-shortest-path-first protocol measures routing ability	IPmark test ("OSPF")
Route Lookup	Uses Patricia-tree algorithm to route IP packets based on lookup tables	IPmark test ("Rtlook")
Packet Checks	Branch- and pointer-intensive code for packet switching; operates on headers in four different data sets with different numbers of packets	Geometric mean is one component of IPmark
Java 2 Micro Edition (J2ME) Benchmarks		
PNG Decoding	Decodes a Portable Network Graphics (PNG) image file	GrinderBench test
Chess	Plays three games of 10 moves each using weighted, variable-depth tree searches; logic-intensive processing without graphics or file I/O	GrinderBench test
XML	XML parsing and DOM-tree manipulation using kXML package	GrinderBench test
Cryptography	Uses DES, DESede, IDEA, Blowfish, and Twofish encryption	GrinderBench test
Regular Expression	Text-string pattern-matching with GNU regular-expression package	GrinderBench test
ParallelBench	Runs two algorithms in parallel to test threading and synchronization	GrinderBench test
8/16-Bit Microcontroller Benchmarks		
Bit Manipulation	Move text into line buffer, convert to pixels, move to display buffer	MicroMark test
Memory Access	Add values from two different memory arrays and store results in a third array, all in different memory regions	MicroMark test
Pointer Chasing	Manipulates pointers by searching a linked list to match an input token	MicroMark test
PWM	Simulates an H-bridge motor driver using pulse-width modulation	MicroMark test
CAN	Simulates remote data requests in an automotive controller-area network	MicroMark test
Road-Speed Calc	Simulates automotive road-speed calculation using timer-counter values	MicroMark test
Tooth to Spark	Simulates automotive fuel injection and ignition	MicroMark test
Task-Based Test	Nine implementation-independent tasks, including memory fills, memory moves, checksum calculations, multiplication, division, UART loopbacks	MicroMark test
Digital Entertainment Benchmarks		
MPEG-2 Encode	Encode an audio/video stream using MPEG-2 specification	EncodeMark test*
MPEG-2 Decode	Decode an audio/video stream using MPEG-2 specification	DecodeMark test*
MP3 Decode	Decode an audio stream using MPEG-2 Layer 3 (MP3) specification	DecodeMark test*
MPEG-4 Encode	Encode an audio/video stream using MPEG-4 specification	EncodeMark test*
MPEG-4 Decode	Decode an audio/video stream using MPEG-4 specification	DecodeMark test*
Huffman Decode	Decode a datastream using a Huffman algorithm	DecodeMark test*
AES Encryption	Advanced Encryption Standard (AES) encryption and decryption	CryptoMark test*
DES Encryption	Data Encryption Standard (DES) encryption and decryption	CryptoMark test*
RSA Encryption	Rivest, Shamir, Adelman (RSA) public-key encryption and decryption	CryptoMark test*

Table 2. These are the new benchmark suites EEMBC has introduced since April 2000 or will introduce shortly. The networking 2.0 suite is now replacing the networking 1.0 suite, and the digital entertainment suite is nearing formal adoption. EEMBC introduced the Java 2 Micro Edition (J2ME) and 8/16-bit microcontroller suites over the past two years, but so few companies have published scores that both suites must undergo major revisions to survive. *Tentative names for components of the proposed DENmark summary score.

with the types of Java programs deployed on mobile consumer devices. Perhaps the overhead of the operating system and Java runtime environment obscures the processor's role in running Java software, making processor-level benchmarks less important. Yet another possibility is that vendors prefer Pendragon Software's less stringent Embedded CaffeineMark, sometimes criticized as the Dhrystone of Java. EEMBC still hasn't given up on its Java benchmarks and is considering ways to revive and revise them.

The 8/16-bit microcontroller committee is going back to the drawing board, too. Committee chairman David Lamar of NEC speculates that vendors may simply be more interested in promoting their 32-bit processors than their lower-priced 8- and 16-bit chips. There's certainly more profit in 32-bit

processors, although the smaller chips continue to outsell 32-bit chips by a wide margin.

Maybe the 8/16-bit benchmarks don't provide enough information that customers really want to know about the chips. Customers may care less about the speed with which a microcontroller can copy an integer array than about how much power the chip consumes, how much scratchpad memory it has, how many peripherals it integrates, which software-development tools are available for it, and which embedded operating systems it runs. Nevertheless, *MPR* thinks the processor-intensive tests in the microcontroller benchmark suite are worth keeping, because they can help a designer decide whether the target application needs an 8-, 16-, or 32-bit chip.

Configurable Processors Stretch the Rules

EEMBC has weathered some other storms. One controversy erupted when vendors of synthesizable processors lobbied for the right to benchmark their intellectual-property (IP) cores in simulation rather than on physical chips. Some EEMBC members objected that it's unfair to compare the performance of real chips with cycle-accurate simulations of processors that may be unable to achieve their target clock frequencies when fabricated in actual silicon. IP vendors argued that without simulator-based benchmarking, they wouldn't be able to certify the performance of their processors without making test chips or waiting for customers to spin their own silicon, which might take years after the introduction of a new core.

The IP vendors won that round. However, EEMBC reports simulator scores separately from chip scores and normalizes the simulator scores to a clock frequency of 1.0MHz. Extrapolating to the processor's target clock frequency is an exercise left to the vendor or customer. In this instance, EEMBC was able to reach a reasonable compromise without impugning the integrity of the benchmarks—although there was one minor incident when IP vendor ARC overestimated the clock speed of its ARCTangent-A4 processor by 25% and had to scale down the simulator-based benchmark results. (See the sidebar, “Wiggle Room in EEMBC's Simulated Benchmarks,” *MPR 9/16/02-01*, “Tensilica Xtensa V Hits 350MHz.”)

Configurable-processor vendors like ARC and Tensilica are special cases. They have pushed the boundaries of EEMBC's rule book to post astonishingly high benchmark scores, much to the chagrin of other EEMBC members, whose processors run at higher frequencies and have more sophisticated, but fixed, architectures.

Tensilica raised eyebrows in 2001 when it published the first certified EEMBC scores for a configurable processor core. The Xtensa III used specially designed custom instructions to accelerate the benchmark kernels, an extreme form of optimization that EEMBC didn't anticipate when it wrote the rules for full-fury benchmarking. (See *MPR 4/9/01-01*, “Stretching Silicon to the Max.”) Indeed, theoretically, it would be possible for a configurable processor to execute some benchmark kernels with a single custom instruction. ARC soon followed Tensilica with similarly impressive optimized scores.

Earlier this year, Tensilica went still further when it published benchmark results for its new Xtensa LX processor core. Once again, Tensilica designed custom instructions to accelerate the benchmark kernels, to great effect: Xtensa LX briefly set a new record for the ConsumerMark (since overtaken by Freescale's MPC7447A). But this time, Tensilica published the results as an out-of-the-box score, not a full-fury score, despite the blatant optimizations.

EEMBC had to overlook the optimizations because Tensilica didn't modify the benchmark source code, only the processor. In the past, Tensilica had to insert special intrinsic functions to invoke custom instructions, but the latest version

of Tensilica's C/C++ compiler can automatically use new instructions without any changes to the source code. (See the sidebar, “How Tensilica Busted the Benchmarks,” *MPR 5/31/04-01*, “Tensilica Tackles Bottlenecks.”) Nevertheless, Tensilica didn't break EEMBC's rules, and real developers can use the same techniques, so the benchmark results are a valid indicator of actual performance.

EEMBC Undeterred by Storms

Despite all the challenges EEMBC has faced, the consortium's outlook is bright. In the field of embedded-processor benchmarking, it has no direct competition. The only exception is Berkeley Design Technology Inc. (BDTI), which dominates DSP benchmarking. Some EEMBC suites contain signal-processing kernels (such as the FFTs in the auto/industrial and telecommunications suites), and DSP vendors such as Analog Devices and Texas Instruments are EEMBC members with published benchmark scores. However, EEMBC doesn't focus on signal processing to anywhere near the degree BDTI does. EEMBC's benchmark suites are application oriented, and the occasional DSP kernels are only a small part of them.

In addition, BDTI operates with a completely different business model than EEMBC's. It's a for-profit private company that writes its own benchmark tests and carefully optimizes the code in assembly language for each DSP (for a fee reportedly in the \$100,000 range). Because of this optimizing, BDTI's benchmark scores are similar to EEMBC's full-fury scores. BDTI doesn't certify its benchmark results through an independent third party, as EEMBC does, but the company has a good reputation and enjoys the trust of vendors and customers.

Looking toward the future, EEMBC has an opportunity to repeat its triumph of redefining embedded-processor benchmarking. The next frontier is power consumption. Power is a huge consideration in the design of embedded systems, and it can radically alter the evaluation of a processor. What good is high computational performance if the processor would bust the system's power budget or require an extensive redesign, such as the addition of active cooling or a larger battery? For decades, designers have been at the mercy of chip vendors, which have their own ways of measuring “typical” power under simulated workloads. A reliable, consistent, understandable power-consumption benchmark would be a tremendous boon to designers.

This opportunity hasn't escaped the attention of EEMBC, which began exploring the possibility of developing power-consumption benchmarks about three years ago. As might be expected, it's a project riddled with technical challenges, controversy, and competing corporate interests. Right now, the plan is to measure energy consumption while running the existing benchmark suites and then express the results in joules, perhaps summarized with an aggregate “PowerMark” score. There's still lots of work to do, so don't expect any power benchmarks to appear for at least another year. If EEMBC can find a path through this

thicket, the consortium has a chance to make up for its missteps with the Java and 8/16-bit suites.

A New Player for x86 Benchmarking

Meanwhile, a startup company whose founder has close ties to EEMBC burst on the scene this year with new benchmarks for x86 processors. Synchronesh Computing LLC—not to be confused with Synchronesh Limited, a visualization-software company in New Zealand—has introduced its Embedded Processor Rating System (EPRS), primarily intended for x86-based low-end PCs and high-end embedded systems. The sole proprietor of Synchronesh Computing is Alan R. Weiss, who also runs ECL, EEMBC's lab for benchmark certification. In fact, the staffs of Synchronesh Computing and ECL work on both EEMBC and non-EEMBC projects.

Despite the apparent connections, EEMBC, ECL, and Synchronesh Computing are separate legal entities with different business models and goals. EEMBC is a nonprofit industry consortium that concentrates on embedded-processor benchmarking for any CPU architecture. ECL is a for-profit private company that certifies benchmark scores and writes benchmark code for EEMBC. Synchronesh Computing is a for-profit private company that performs system-level x86 benchmarking, consulting, software development, technical writing, and other services independently of EEMBC. With Synchronesh Computing, there's no benchmarking organization to join, and anybody can use the benchmarks—for a negotiable fee in the tens of thousands of dollars. There is a little overlap between Synchronesh Computing and EEMBC, because the consortium's benchmarks are also useful for evaluating embedded x86 processors, albeit from a CPU-level perspective.

Both EEMBC and Synchronesh Computing require vendors to certify their benchmark results before sharing them with customers. But there's a notable difference. EEMBC uses a third party (ECL) for certification; Synchronesh Computing uses itself. Weiss explains that no one has more benchmark-certification experience than ECL, and the staffs of ECL and Synchronesh Computing are the same, so Synchronesh Computing is the logical choice to verify the EPRS results. It might be better if Weiss could find an equally qualified, disinterested third party for that task—if such a party were available—but BDTI has built a solid reputation on a similar business model, and it is self-policing, too.

However, within days after releasing its first test results, Synchronesh Computing became enmeshed in a heated controversy about its benchmark suite and testing methods. The controversy arose when the lab tested three of AMD's embedded x86 processors marketed under the Geode name and two of VIA's competing x86 processors. The scores tended to favor AMD's processors. This result hit a particular sore point with VIA, partly because AMD funded the benchmark suite's development and sponsored the comparison, and partly because the president of VIA's microprocessor subsidiary, Glenn Henry, long an outspoken

opponent of CPU benchmarking, rarely publishes scores. (See the accompanying sidebar, "VIA Disputes the Synchronesh Computing Benchmarks.")

Weiss fervently denies any implication that he skewed the benchmarking to favor his client. He points to the unblemished reputation of his other company (ECL) as the certification lab for EEMBC and his years of experience in microprocessor engineering and testing. Weiss says Synchronesh Computing and AMD are interested only in establishing a new x86 benchmark that's a better expression of system performance than processor clock frequency.

Analyzing the EPRS/PPR Performance Ratings

To create the EPRS benchmark suite, Weiss says he asked AMD to suggest some publicly available benchmark programs, then accepted some of AMD's suggestions while rejecting others. In addition, Weiss says he chose some publicly available benchmark programs he wanted, and Synchronesh Computing wrote new tests of its own. The result is a composite suite that measures many aspects of system, subsystem, and CPU performance.

An EPRS rating, expressed as a single number, is an unweighted geometric mean of seven benchmark scores derived from test runs of the composite suite. In turn, those seven scores summarize the results of about three dozen component tests. The final EPRS performance rating is a normalized number that represents clock-equivalent x86 performance.

However, *MPR* finds the EPRS performance ratings confusing, for two reasons. First, the ratings are normalized to embedded x86 processors from VIA—and not always to the same VIA processor. It's puzzling that processors from a vendor with less than 2% market share should be the yardstick by which all other embedded x86 processors are measured. And because Synchronesh Computing hasn't standardized on a single VIA processor as the baseline, the EPRS performance ratings vary from processor to processor.

Our second criticism of the EPRS performance ratings is that Synchronesh Computing hasn't reported its raw benchmark results. Unlike EEMBC and most other benchmarking organizations, Synchronesh Computing hides the raw numbers behind the normalized scores. This prevents anyone else from computing an alternative view of the data.

After its first round of benchmark testing for AMD, Synchronesh Computing's performance ratings for AMD processors have produced misleading results. The lab normalized the ratings for AMD's Geode GX processors to a 533MHz VIA Eden processor, and it normalized the ratings for the Geode NX to a 1.0GHz VIA Nehemiah processor. Because the ratings are normalized to different baselines, they aren't directly comparable. For instance, it's wrong to conclude—as a casual observer surely would—that the Geode NX 1500 is three times faster than the Geode GX 500, even though both of those EPRS performance ratings are derived from the same EPRS benchmark suite.

In reality, the Geode NX is certainly more than three times faster than the Geode GX. The NX is a recently rebranded Athlon XP processor incorporating every important microarchitectural innovation of the past 10 years, whereas the GX is little changed from the uniscalar MediaGX core Cyrix designed in 1995. Yet, the clock-equivalent EPRS performance ratings for those two processors suggest a 3× difference in performance—the same as their difference in core clock frequencies (333MHz vs. 1.0GHz).

Only by reading to the conclusion of a white paper Synchronesh Computing wrote for AMD would a customer learn that the EPRS ratings for the Geode GX and Geode NX are scaled to the baselines of two different VIA processors. Even then, customers would be unable to compute for themselves the performance difference between the GX and NX, because Synchronesh Computing hasn't reported the raw benchmark results, only the normalized scores. In sum, *MPR* finds the shifting EPRS ratings needlessly confusing and little more useful than clock speeds.

AMD has extended the EPRS rating with a power rating, expressed in watts, to obtain what AMD calls the Performance Power Rating (PPR). The power rating isn't an actual measurement by Synchronesh Computing. Instead, the processor vendor provides the power rating, based on idle power, total design power (TDP, a maximum power rating), and "typical" power, which is somewhere between idle power and TDP. Weiss says he didn't undertake the task of defining his own power-measurement benchmarks, because it's difficult to measure power on some processors and boards from different vendors. As mentioned above, this is a thorny problem that has already occupied EEMBC for quite a while. (Weiss is also involved with the EEMBC power benchmarking project.)

Synchronesh Computing's EPRS Suite

Although *MPR* finds fault with the EPRS/PPR clock-equivalent performance ratings, we have fewer objections to the makeup of the composite benchmark suite on which they're based. In general, it's a solid suite that could help observers discover useful information about the benchmarked processors—especially if the raw results were available.

Table 3 shows the major benchmark programs and component tests in the EPRS/PPR suite. The major programs are *HDBench 3.2.2*, a popular PC system benchmark from Japan; *HINT* (Hierarchical INtegration), a CPU/memory benchmark developed by the U.S. Department of Energy's Ames Laboratory; *IM Chat*, an instant-messaging benchmark created by Synchronesh Computing; *SANDRA* (System Analyzer/Diagnostic and Reporting Assistant), a broad system-level benchmark from SiSoftware; *Stream*, an integer and floating-point math benchmark created by Dr. John McCalpin, now at the University of Virginia; *Synchrobench*, a multimedia-intensive benchmark developed by Synchronesh Computing; and *Winbench '99*, a broad subsystem-level benchmark from Ziff-Davis, a publishing company.

The EPRS suite's system-level orientation is obvious. Unlike EEMBC, Synchronesh Computing measures the performance of PC disk drives, memory, graphics cards, and operating-system APIs. There are multiple levels of redundancy, including different tests of the same subsystems and tests derived from other tests in the EPRS suite (such as the memory tests in the *SANDRA* section, which are based on the *Stream* tests). *Winbench '99* was adopted instead of later versions because Microsoft uses it internally to benchmark its Remote Desktop Protocol (RDP)—a method for controlling thin clients over a network—and thin clients are a target application of the EPRS suite.

Details about the publicly available programs in the EPRS suite are widely available on the Internet (see the "For More Information" box for hyperlinks), but the two new minisuites created by Synchronesh Computing deserve special mention. One odd addition is the *IM Chat* (instant messaging) benchmark. At first, it seems trivial: Who cares how fast a teenager can gossip about Britney Spears? But according to the Yankee Group, 330 million businesspeople will be IM users by the end of 2005, compared with 65 million this year. Even after subtracting the fantasy-baseball trades and dumb-blonde jokes, that's a lot of business traffic over IM channels. And Weiss points out that IM chat is a rapidly growing application among young people on consumer PCs and mobile-computing devices.

Still, wider use of IM doesn't necessarily make *IM Chat* a relevant benchmark. The two most important factors governing the speed of IM are network throughput and typing dexterity. *IM Chat* eliminates both factors. To cancel the variable of network throughput, Synchronesh Computing sets up a small intranet consisting of a client PC and an IM server running open-source *Jabber* IM software. (The far more popular AOL, Microsoft, and Yahoo IM systems run on the public Internet, making it impossible to factor out the network overhead and control the response time of the servers.) On the client PC, two chat windows are open. The benchmark testers send messages from one chat window to the other by bouncing the text through the IM server on the intranet. To eliminate the variable of typing speed, Synchronesh Computing uses IBM's *Rational Visual Test* program, an automated keystroke-injection tool. It simulates someone typing text into the chat windows.

IM Chat measures the total time required to send an ASCII text message from one window to the other, and it also measures the amount of time the message spends traversing the network. Total time minus network time equals overhead time, which essentially is the time consumed by the IM client software and network stack. Weiss says AMD welcomed this benchmark test enthusiastically. However, *MPR* questions the relevance of *IM Chat*. We don't believe this data will influence a customer's buying decision or design choices.

Synchrobench Measures Web Surfing

Synchronesh Computing's other original benchmark program looks better. Known as *Synchrobench* (a white paper

Benchmark Tests	Performance Measurements
HDBench 3.2.2 (Japan)	
CPU	Integer and floating-point math
Memory	Read, write, and read/write operations
Graphics	Draw rectangles and ellipses, scroll text, bit-block transfers, DirectDraw
Hard Disk	Read, write, and file-copy operations
HINT (Ames Laboratory, U.S. Dept of Energy)	
HINT	Progressive mathematical calculations measure CPU and memory performance in NetQUIPS (quality improvements per second)
Instant Messaging Chat (SynchroMesh Computing)	
IM Chat	Measures total time and overhead time (total time minus network latency) to send simulated typed messages to and from an external instant-messaging server over a closed network
SANDRA (SiSoftware)	
Modified Dhrystone	CPU integer performance in nonstandard Dhrystone mips
Modified Whetstone	CPU floating-point performance in nonstandard Whetstone MFLOPS
Optimized Integer / FP	CPU integer and floating-point tests optimized for x86 extensions: MMX, Enhanced MMX, SSE, SSE2, 3DNow!, 3DNow! Enhanced instructions
Video	Multithreaded graphics engine uses DirectX 8.0 and OpenGL
File System	Large data file measures sustained read/write speeds of hard drive, CD-ROM, DVD
Memory (Uniprocessor)	Stream-derived integer and floating-point tests measure sustained memory bandwidth on uniprocessor systems with and without code prefetching and buffering
Memory (Multiprocessor)	Stream-derived integer and floating-point tests measure sustained memory bandwidth on multiprocessor systems with and without code prefetching and buffering
Cache and Memory	Stream-derived integer and floating-point tests for uniprocessor or multiprocessor systems measure sustained memory bandwidth without code prefetching or buffering
Network / LAN	Measures transfer rate of system's network interface by moving large packets
Stream (Dr. John McCalpin)	
Copy	Measures bytes per iteration of an array-copy loop
Scale	Measures bytes and FLOPS per iteration of an array-scaling loop
Sum	Measures bytes and FLOPS per iteration of an array-adding loop
Triad	Measures bytes and FLOPS per iteration of an array-adding and -scaling loop
Synchrobench (SynchroMesh Computing)	
HTML Rendering	CPU headroom while scrolling through static Web pages on a local hard disk
MP3 Playback	CPU headroom while playing back MP3 music files
MPEG Playback	CPU headroom while playing two MPEG video streams
Flash Playback	CPU headroom while playing a Macromedia Flash movie
Real-Time Clocks	CPU headroom while displaying multiple real-time clocks
Winbench '99 (Ziff-Davis)	
Disk WinMark	Disk-access patterns of seven PC applications, measured in megabytes/sec
Disk (Low Level)	Hard-disk access time (milliseconds), CPU utilization (percentage, lower is better)
Graphics WinMark	Graphics performance (unitless)
Graphics (Inspection)	Graphics performance in millions of pixels per second
CPUmark	CPU integer performance
FPU WinMark '99	Floating-point performance
Video (Visual Quality)	Number of frames dropped during video playback
Video (Audio Quality)	Breaks during audio playback
Video (Temporal Quality)	Percentage of nominal speed (ideal=100%)
Video (Frame Rate)	Maximum number of video frames per second
Video (CPU)	CPU utilization (percentage, lower is better)
DirectDraw	DirectDraw block-transfer speed in millions of pixels per second

Table 3. SynchroMesh Computing combined some popular off-the-shelf benchmark suites with two scratch-built minisuites to assemble this large composite suite, the basis of the company's new Embedded Performance Rating System (EPRS), also called Power Performance Ratings (PPR) by AMD. Note that many of the tests measure the performance of various PC subsystems, not just the CPU. Also, the suite runs on x86 systems only. Those two characteristics distinguish the suite from the EEMBC benchmark suites, which measure the performance of embedded processors and are portable to virtually any CPU architecture.

the lab wrote for AMD mistakenly refers to it as *Surfbench*), it measures the amount of performance headroom available on a processor while running various Internet-related tasks.

In an HTML rendering test, *Synchrobench* automatically scrolls through a number of web pages stored on a local hard disk. In the MP3 playback test, the benchmark program plays

MP3 audio files. In the MPEG playback test, the program plays two video clips encoded in MPEG-2 format. In the Flash playback test, the program plays a short movie (with audio and video) encoded in Macromedia Flash format. Finally, in the real-time clocks test, a program displays multiple graphical clocks in separate windows on the screen. During each test,

VIA Disputes the Synchronesh Computing Benchmarks

Synchronesh Computing's first benchmark tests have sparked the lab's first benchmarking controversy. It started when AMD commissioned the lab to test five embedded x86 processors: three from AMD and two from VIA. VIA wasn't involved in the benchmarking.

Alan R. Weiss, founder of Synchronesh Computing, describes the job as a study to prove that a processor's system-level performance may vary from the performance implied by its clock frequency and to help AMD express the performance differences among its Geode-brand x86 processors, which are based on two vastly different microarchitectures. Weiss says he benchmarked VIA's processors merely as baseline clock-frequency references, not to stage a head-to-head comparison. However, AMD's subsequent marketing of the benchmark results certainly makes it look like a contest. AMD posted the benchmarks on the Geode section of its website, along with a white paper written by Synchronesh Computing. The white paper makes numerous comparisons between the AMD and VIA processors.

According to the EPRS/PPR benchmark scores, three of AMD's Geode GX and Geode NX processors performed much better than their clock frequencies might suggest when compared with two of VIA's x86 processors. VIA wasn't surprised that the Geode NX performed well, because it's a rebranded Athlon XP desktop PC processor with superscalar pipelines and other advanced features. VIA's processors have simpler, uniscalar microarchitectures designed for low power in embedded systems and bargain PCs.

VIA was surprised, however, by the other benchmark results, because AMD's Geode GX is based on the MediaGX processor announced by Cyrix at Microprocessor Forum 1995. (See *MPR 3/10/97-01*, "MediaGX Targets Low-Cost PCs.") While passing from Cyrix to National Semiconductor to AMD, the MediaGX has been improved with a deeper uniscalar pipeline and other enhancements (see *MPR 11/5/2001-02*, "National Polishes Geode"), but the newer VIA processors have more-sophisticated cores.

The two AMD processors in question are the Geode GX 466 and Geode GX 533. Their actual clock frequencies are 333MHz and 400MHz, respectively. Their product numbers are rounded clock-equivalent performance ratings derived from the Synchronesh Computing EPRS/PPR benchmarks, normalized to a 533MHz VIA x86 processor. For instance, the normalized rating for the Geode GX 533 is supposed to indicate that it performs on a par with VIA's 533MHz chip.

To obtain those performance ratings, Synchronesh Computing benchmarked two VIA processors identified as the 1.0GHz Nehemiah and 533MHz Centaur. Initially, this caused some confusion, because VIA doesn't sell any chips

under the Centaur brand. (It should be noted that even *MPR* finds VIA's processor nomenclature confusing.)

Adding to the confusion, the VIA motherboard that Synchronesh Computing's white paper associated with the "Centaur" processor isn't compatible with the memory the lab said it used for the tests. According to the white paper, Synchronesh Computing tested a 533MHz Centaur on a VIA EPIA-M motherboard with a 133MHz memory bus and 256MB of SDRAM. However, the EPIA-M board has a chip set with a 266MHz DDR memory bus, and it doesn't work with SDRAM.

Synchronesh Computing actually tested a 533MHz C3 or Eden processor (code-named Samuel-2) on a mini-ATX EPIA board, a two-year-old platform that supports SDRAM on a 133MHz memory bus. For that reason, VIA questions the validity of drawing conclusions about relative processor performance from the benchmark results, because Synchronesh Computing pitted the older VIA processor and board against the latest AMD processors and boards that support 266MHz DDR DRAM. In other words, the AMD processors had memory systems that were twice as fast, and several of the EPRS benchmark kernels are memory intensive. Synchronesh Computing says it bought the VIA processors and boards on the open market and couldn't find faster examples at the time.

Figure 1 is a graph from Synchronesh Computing's white paper on AMD's website. The graph shows VIA's processor matching or exceeding the performance of the Geode GX processors, except in the memory bandwidth tests (*SANDRA* and *Stream*). Almost certainly, the normalized performance ratings for the Geode processors would have been lower if the memory systems of all the test systems had been identical (either SDRAM or DDR).

To counter AMD's marketing of the benchmarks, VIA ran a variety of benchmark tests on its own processors and AMD's Geode GX 500. (VIA says the slightly faster GX 533 wasn't available from distributors at the time.) Among those benchmark tests were some found in the Synchronesh Computing suite. All systems in VIA's testing had 266MHz DDR memory, although the fixed clock multipliers on the AMD PR-500 motherboard reduced the GX 500's effective bus speed to 244MHz. In all, VIA ran about 175 test kernels—at least four times as many kernels as in the EPRS suite—and sent the results in a huge spreadsheet to *MPR*.

According to that data, VIA's 533MHz Nehemiah and Samuel-2 processors are at least twice as fast as AMD's Geode GX 500. The results are consistent with our assessment of the VIA Samuel-2 and Geode GX microarchitectures. Samuel-2 has a deeper pipeline (12 stages vs. 8), larger L1 caches (64K vs. 16K), an on-chip L2 cache (64K vs. none), and dual TLBs (two 128-entry TLBs vs. a unified

Continued on Page 12

64-entry TLB). Judging from their microarchitectures, we would be as surprised as VIA if any processor-intensive benchmarks rated a 366MHz Geode GX at nearly the same performance level as a 533MHz Samuel-2 with a similar memory system. Even the Synchronesh Computing white paper attributes the strong performance of the Geode GX to its unequal memory system.

With VIA's permission, *MPR* forwarded the benchmark results to Weiss at Synchronesh Computing for comment. (VIA has also made the information publicly available and is offering to provide customers and analysts with the hardware to duplicate the tests.) Weiss's response addresses three main points: the validity of VIA's scores, VIA's selection of processors, and the relevance of some of VIA's benchmarks.

First, Weiss says his benchmark scores are more trustworthy than VIA's because they have been certified by his lab. Weiss says he has asked VIA to submit its processors to the same benchmarking and certification process—for his standard lab fee, of course—but VIA has shown no interest.

Second, Weiss questions VIA's choice of the Geode GX 500 (actual clock speed 366MHz) instead of the GX 533 (400MHz) that was part of the original AMD-VIA comparison. As mentioned above, VIA says the GX 533 was unavailable at the time. VIA's substitution gave its 533MHz Nehemiah and Samuel-2 processors a 46% clock-speed advantage. However, that's not enough to account for VIA's much higher performance in the vast majority of its tests. In addition, the clock-speed advantage shrinks to 7% if the EPRS/PPR rating for the GX 500 is accurate.

Third, Weiss questions the relevance of some of VIA's benchmarks, such as the 3D graphics tests, Business Winstone, and OfficeBench. Those benchmarks are more suitable for higher-end desktop PCs, not the embedded systems for which the Geode GX is intended. Weiss points out that even under VIA's testing, the Geode GX fared better on some tests, particularly those involving floating-point math and memory throughput. VIA says it ran a variety of benchmarks to avoid criticism that it "cherry-picked" the tests to

Continued on Page 13

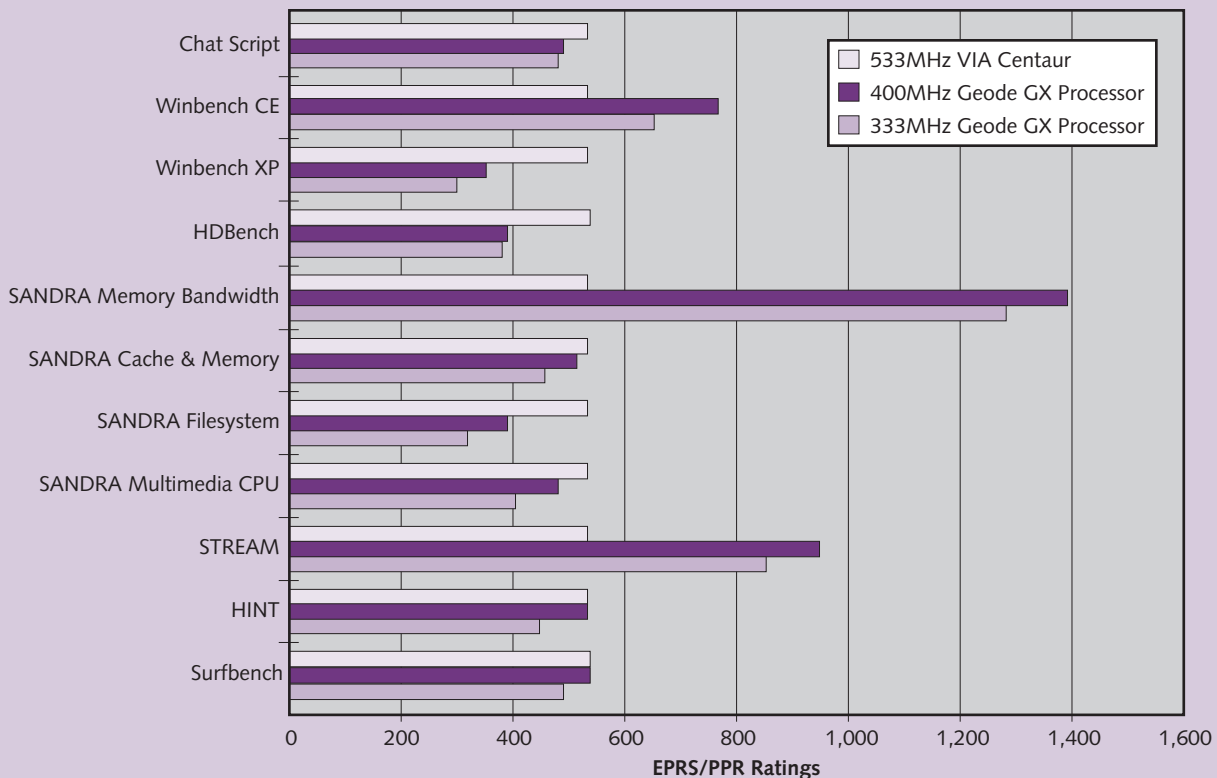


Figure 1. Synchronesh Computing published these benchmark results in its white paper written for AMD. The lab normalized all results to VIA's 533MHz Samuel-2 (mistakenly identified by Synchronesh Computing as a "Centaur"). VIA's higher-frequency processor matched or exceeded the performance of AMD's processors in almost all tests, with the notable exceptions of two memory-intensive benchmarks: the *SANDRA* memory-bandwidth test and *Stream*. Clearly, the VIA processor was handicapped by its 133MHz SDRAM memory system, whereas the AMD processors had 266MHz DDR memory. Synchronesh Computing says it couldn't find a VIA motherboard with DDR support in time for the testing, and that one goal of the exercise was to establish that processor performance is only part of total system performance.

favor its processors. VIA also notes that some embedded x86 applications, such as kiosks, use 3D graphics.

After examining the evidence, MPR has reached the following conclusions. Any attempt to judge relative processor performance from Synchronesh Computing's first round of AMD-VIA benchmarking is suspect, because the memory subsystems were unequal. System-level comparisons are more valid, but not very useful, because VIA's current DDR systems weren't tested against AMD's DDR systems. We think Synchronesh Computing's white paper and its choices of baselines for EPRS/PPR performance ratings (two different VIA processors) contradict its claims that the testing

wasn't a head-to-head comparison with AMD processors. Furthermore, we think an independent testing lab shouldn't scale its performance ratings to processors whose vendor wasn't involved in the testing and is a direct competitor of the vendor that commissioned the work.

Finally, we recommend that Synchronesh Computing consult with all vendors involved in future comparative benchmark tests, even if some vendors aren't paying clients. Vendors could make a case for which of their processors should be compared, and it would avoid the kind of reporting errors that made the AMD-VIA scores difficult to interpret.

Continued from Page 10

Synchrobench measures the percentage of CPU performance remaining. If no other tasks were executing, it would report 100% headroom available. The *Synchrobench* score is the geometric mean of the headroom percentages in each test.

Although the score describes CPU headroom, it's actually measuring much more, including the speed of the web browser and its multimedia plug-ins and the efficiency of the operating system, network stack, and graphics subsystem. It's really a system-level headroom test that would probably vary quite a bit with a different operating system or web browser. For now, the only operating systems compatible with the EPRS suite are Windows XP and Windows CE.NET, although Synchronesh Computing hopes to port the suite to GNU/Linux in the near future.

Other programs in the EPRS suite measure parameters similar to those in the *Synchrobench* tests. *HDBench*, *SANDRA*, and *Winbench '99* all test the graphics subsystem in various ways, although they usually work at a lower level than do the plug-in codecs in *Synchrobench*. From a technical standpoint, we believe *Synchrobench* adds value to the EPRS suite and is certainly more useful than the *IM Chat* benchmark. From a business standpoint, the most significant contribution of *Synchrobench* and *IM Chat* is that they require anyone who wants to use the EPRS suite to sign a deal with Synchronesh Computing, because the other benchmark programs in the suite are freely available.

Synchronesh Computing isn't directly competitive with EEMBC, but there is some common ground. High-end set-top boxes, Internet appliances, and some thin clients are embedded applications, so they fall within EEMBC's purview, at least at the processor level. Synchronesh Computing created the EPRS suite for the same applications (albeit at the system level) and for low-end PCs, which are beyond EEMBC's scope. Unless EEMBC or Synchronesh Computing significantly expands its respective territory, we don't fear a collision. In fact, Synchronesh Computing recently dropped plans to create a multicore, multiprocessing benchmark suite when it discovered EEMBC is working on a similar suite. Weiss (through his other company, ECL) and EEMBC are now working on the suite together.

How to Improve Embedded Benchmarking

Overall, the state of embedded benchmarking has never been better. EEMBC has brought an unprecedented level of legitimacy and integrity to the embedded-systems industry, and the PC industry is foolish not to emulate it. It wasn't easy to gather dozens of processor vendors together in the same room, much less to get them to agree on anything. Everyone involved with the consortium deserves credit for a job well done.

EEMBC has stumbled a few times, but it hasn't fallen. We don't believe Synchronesh Computing and other purveyors of proprietary benchmarks pose a significant threat to the consortium. Frankly, it's hard to imagine anything displacing EEMBC at this point, unless the consortium loses its vision and self-destructs from internal catfighting.

Nevertheless, there is always room for improvement. We hope EEMBC succeeds in its difficult attempt to define power-consumption benchmarks, which would be welcomed by all embedded-system developers. It's a tough job, but someone has to do it, and no one is better positioned to do it than EEMBC. We're also looking forward to the next revisions of EEMBC's existing benchmark suites, which are long overdue.

To satisfy outside critics, EEMBC should open its source code under an NDA to industry analysts, journalists, and other independent parties who want to evaluate the benchmark kernels for technical relevance and integrity. The relatively few people who would exercise this option could offer valuable input to the technical committees in charge of those kernels.

Obviously, encouraging more EEMBC members to publish certified scores would make the benchmarks even more useful and important. One idea is to allow any EEMBC member to test any other member's processors and publish the scores, as long as they are certified by ECL. Cross-testing would allow members to publicly compare their processors with those of the competition, even if the competitors are too bashful to publish scores. EEMBC does allow members to publish certified benchmarks for processors whose vendors don't belong to EEMBC, but, so far, no one has done so.

For More Information

- EEMBC: www.eembc.org
- HINT benchmark: <http://hint.byu.edu>
- SANDRA benchmark: www.sisoftware.co.uk
- Stream benchmark: www.cs.virginia.edu/stream
- Synchromesh Computing:
<http://synchromeshcomputing.com>
- Winbench '99:
www.veritest.com/benchmarks/winbench
- AMD's explanation of Performance-Power Ratings:
www.amd.com/us-en/ConnectivitySolutions/ProductInformation/0,,50_2330_9863_10848,00.html

We think Synchromesh Computing can improve its EPRS benchmarks by making the performance ratings more rational, reporting the raw test results, and tweaking the benchmark suite. The performance ratings should be normalized to

a single baseline, perhaps an older x86 processor, like the 286 or 386, that's relatively generic and won't ruffle any vendor's feathers by getting a low score. Benchmark reports should include raw test results, not just normalized scores, so independent observers can study the data and reach their own conclusions. The EPRS benchmark suite would be better without *IM Chat*, and Synchromesh Computing should solicit suggestions for a substitute from all embedded x86 vendors, whether or not they are paying clients.

Of course, *MPR* is biased in the sense that we prefer more benchmark results and better benchmark results. Benchmark scores help us evaluate and compare embedded processors. We recognize that processor vendors have their reasons for keeping scores private or for avoiding benchmarks altogether. Although we can sympathize with the vendors, we're more allied with the primary consumers of benchmarks—the hardware designers and software developers who crave something more substantial than datasheets and Dhrystone MIPS when facing a blank slate at the start of a development project. ♦

To subscribe to Microprocessor Report, phone 480.609.4551 or visit www.MDRonline.com