Reed Electronics Group

MICROPROCESSOR REPORT

www.MPRonline.com

THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

EEMBC EXPANDS BENCHMARKS

New Digital Entertainment Suite Tests Audio, Video, Cryptography

By Tom R. Halfhill {2/22/05-01}

After two years of labor, the Embedded Microprocessor Benchmark Consortium (EEMBC) has delivered its largest new test suite since introducing its original benchmarks in 2000. Indeed, the new Digital Entertainment suite (DENbench) has more tests than all five suites of the EEMBC 1.0 benchmarks put together. In all, there are 69 new tests, though most are alternative datasets for a smaller number of basic tests.

Don't mistake the Digital Entertainment suite for a frivolous gaming benchmark. Actually, DENbench isn't the best choice for measuring game performance, because it has no tests for 3D graphics. Instead, DENbench consists of four smaller suites whose tests are useful for a broad range of applications, from consumer electronics to secure communications and digital rights management (DRM). EEMBC members can test their processors using all four of the minisuites to report an overall DENmark score, or they can run a subset of the minisuites to report a narrower score—a desirable alternative for benchmarking application-specific processors.

The four minisuites in DENbench measure how quickly and accurately a processor can compress, decompress, and change the colorspaces of still images; encode motion video; decode motion video and audio; and encrypt and decrypt a datastream. DENbench is relevant for cellphones, MP3 players, digital cameras, camcorders, DVD players/recorders, TV set-top boxes, secure routers, broadband modems, PCs, PDAs, and just about everything else that does audio, video, graphics, or cryptography—which these days is virtually everything. Table 1 summarizes all the benchmark tests in each DENbench minisuite.

In addition to releasing the new benchmark suite, EEMBC has published the first certified DENbench scores. The first brave companies to test their processors with DENbench and submit their results to the EEMBC Certification Labs (ECL) are Analog Devices, Freescale, and IBM Microelectronics. Also, with AMD's permission, ECL tested an AMD Geode NX processor to help verify and calibrate the benchmarks, not to achieve the highest possible scores. (See the sidebar, "Few Surprises In First DENbench Scores.")

As with all EEMBC benchmark suites, the DENbench source code is available only to dues-paying members of EEMBC, a nonprofit benchmark consortium founded in 1997 by EEMBC president Markus Levy. EEMBC currently has more than 50 members, including most leading vendors of microprocessors and processor-related intellectual property as well as vendors of software-development tools, operating systems, and Java virtual machines. Under EEMBC's strict bylaws, members can share benchmark scores with customers that sign a nondisclosure agreement, but members cannot make their EEMBC scores public unless ECL certifies the results. ECL verifies repeatability and enforces EEMBC's rules, to ensure accuracy and prevent cheating. (See *MPR 6/21/99-01*, "Embedded Benchmarks Grow Up.")

DENbench Builds On Consumer Suite

EEMBC already has a benchmark suite for the same kinds of applications as DENbench, but it's showing signs of wear. Although EEMBC revised it slightly once, the Consumer 1.1 suite has scarcely changed since EEMBC first released it in

1999. It has only five basic tests: JPEG compression and decompression for still images; RGB-to-CMYK conversion (a red-green-blue to cyan-magenta-yellow-black colorspace transformation); RGB-to-YIQ conversion (an RGB-to-chrominance-luminance colorspace transformation); and a high-pass grayscale filter. Noticeably absent from the consumer suite are benchmark tests for motion video, digital audio, and cryptography.

Much has changed in the past six years. In 1999, Apple's iPod didn't exist, MP3 players were a new experiment, DVD recorders were a twinkle in Sony's eyes, cellphones didn't have cameras and couldn't play video, digicams were ramping up their assault on silver halide, and Napster was inadvertently paving the way for DRM. In only six years, the need for efficient media processing in both line-powered and mobile consumer products has become nearly universal. The demand for rapid cryptographic processing to maintain security and enforce DRM in consumer-electronics products was hardly anticipated in 1999. EEMBC has recognized these trends, but revising the consumer benchmarks required years of painstaking committee work and software development. When it comes to defining benchmarks, EEMBC's broad membership is both a strength and a weakness, as is true with most consortiums. (See *MPR 8/30/04-01*, "Benchmarking the Benchmarks.")

Benchmark Tests	EEMBC Digital Entertainment Benchmarks (DENbench) Benchmark Suites and Test Descriptions
	DENmark (Geometric Mean of All Suites x 10)
Motion Images	MPEG-2 encode (integer & FP), MPEG-2 decode, MPEG-4 encode, MPEG-4 decode (25 datasets)
MP3 Audio	MPEG-2 Layer 3 (MP3) audio decoding (5 datasets)
Cryptography	AES, DES, RSA, Huffman decoding (4 datasets)
Static Images	RGB-to-YIQ, RGB-to-HPG, RGB-to-CMYK, JPEG compression, JPEG decompression (35 datasets)
	MPEG EncodeMark (Geometric Mean of 10 Tests x 1,000)
MPEG-2 Integer	MPEG-2 video encoding, integer (5 datasets)
MPEG-4	MPEG-4 video encoding (5 datasets)
	Optional Floating-Point MPEG Encode (Geometric Mean of 5 Tests x 1,000)
MPEG-2 FP	MPEG-2 video encoding, floating point (5 datasets)
	MPEG DecodeMark (Geometric Mean of 15 Tests x 1,000)
MPEG-2	MPEG-2 video decoding (5 datasets)
MPEG-4	MPEG-4 video decoding (5 datasets)
MPEG-2 Layer 3	MP3 player audio decoding (5 datasets)
	CryptoMark (Geometric Mean of 4 Tests x 10)
AES	Advanced Encryption Standard
DES	Data Encryption Standard
RSA	Rivest-Shamir-Adleman public-key cryptography
Huffman	Huffman decoding (data decompression)
	ImageMark (Geometric Mean of 35 Tests x 10)
RGB to YIQ	Convert red-green-blue colorspace to a luminance-chrominance colorspace (7 datasets)
RGB to HPG	Convert red-green-blue colorspace to Hewlett-Packard graphics (7 datasets)
RGB to CMYK	Convert red-green-blue colorspace to cyan-magenta-yellow-black colorspace (7 datasets)
JPEG Compress	Still-image data compression using JPEG standard (7 datasets)
JPEG Decomp	Still-image data decompression using JPEG standard (7 datasets)

Table 1. EEMBC's new Digital Entertainment benchmarks (DENbench) include four minisuites and an optional floating-point MPEG-2 encoding test. The four minisuites have their own mathematically derived benchmark scores. MPEG EncodeMark measures motion-video encoding; MPEG DecodeMark measures motion-video and digital-audio decoding; CryptoMark measures data encryption and decryption; and ImageMark measures still-image compression, decompression, and colorspace conversion. Each benchmark score is based on a geometric mean of the test scores (iterations per second) from the corresponding minisuite. EEMBC also calculates an overall DENmark score from the geometric mean of the MPEG EncodeMark, MPEG DecodeMark, CryptoMark, and ImageMark. The geometric mean of the optional floating-point MPEG-2 tests may be included in the overall DENmark score, too.

DENbench retains the JPEG and colorspace-conversion tests from the old consumer suite and adds a third colorspace conversion: RGB to HPG (a Hewlett-Packard graphics format). DENbench also introduces seven new datasets for each benchmark test in this minisuite. The new datasets include larger and more complex images for the data compression, decompression, and colorspace conversions. These expanded datasets are important, because as processors grow more powerful and incorporate larger caches and wider buses, the older, smaller datasets can paint an unrealistic picture of performance. Today's image processors are expected to handle much larger JPEG files than in the past—the one-megapixel digicams costing $1,000 in 1999 have been replaced with five-megapixel cameras costing $300.

Altogether, then, the still-image minisuite has five benchmark tests: RGB-to-YIQ conversion, RGB-to-CMYK conversion, RGB-to-HPG conversion, JPEG compression, and JPEG decompression. With seven datasets per test, the minisuite has 35 tasks. By multiplying the geometric mean of these 35 scores by 10, EEMBC derives a figure of merit called the ImageMark. It will be an important benchmark score for general-purpose processors as well as image processors, given the near-ubiquity of digital photography and 2D graphics.

Curiously, the ImageMark suite omits the high-pass grayscale filter found in the existing Consumer 1.1 suite. This test simulates a two-dimensional spatial filter used by digicams to sharpen a raw image from the CCD or CMOS sensor before applying JPEG compression. EEMBC says it dropped this test because the ImageMark suite was growing large (35 tasks), and because this particular sharpening algorithm is becoming less popular. However, we think EEMBC should have substituted another sharpening algorithm, because virtually all digital cameras automatically sharpen their raw images before compression. In addition, we think the ImageMark suite should add conversion tests for the Adobe RGB and sRGB colorspaces, which are widely used by digital cameras and image-editing programs.

At Last: Motion Video and Digital Audio

One of the biggest improvements in DENbench is EEMBC's first benchmarks for MPEG video and audio. If it seems incredible that any modern benchmark suite could have ignored those media types for so long, remember that EEMBC's mission is to benchmark embedded processors, not PC processors, and multimedia used to be less common in embedded systems than it is now.

DENbench has two multimedia benchmarks: the MPEG EncodeMark and the MPEG DecodeMark. Each score is derived from a geometric mean of the tests within their minisuites. For the MPEG EncodeMark, a minisuite implements MPEG-2 video encoding with integer math, plus an MPEG-4 video encoder. Each test uses five datasets, for a total of 10 tasks. For the MPEG DecodeMark, a minisuite performs MPEG-2 video decoding, MPEG-4 video decoding, and MPEG-2 Layer 3 (MP3) audio decoding. Each of those tests uses five datasets, too, for a total of 15 tasks.

Apart from the MPEG EncodeMark and DecodeMark minisuites, an optional MPEG-2 encoding test uses floating-point math. It has five datasets. The floating-point test is optional, because many embedded processors lack FPUs, and software floating-point libraries are too slow for video encoding. EEMBC doesn't factor results from the floating-point tests into the MPEG EncodeMark, but it does use the results (if available) to calculate the overall DENmark score.

The MPEG-2 encoding tests use algorithms adapted from the International Standards Organization (ISO). Both the integer and floating-point algorithms use Huffman compression and modified inverse discrete cosine transform (iDCT) routines. Input files are in the portable pixmap (PPM) format, and output files are in MPEG-2 format with optional UUencoding, so they're compatible with Microsoft's Windows Media Player and Apple's QuickTime Player.

MPEG uses lossy compression and doesn't mandate any particular level of audio or video quality. Without a way of judging quality, the MPEG benchmarks would be meaningless, because a vendor could severely compromise quality to achieve higher throughput. Therefore, EEMBC objectively measures the quality of MPEG encoding and decoding using a mathematical method called peak signal-to-noise ratio (PSNR). This method compares the output frames with reference frames of known quality and calculates a sum of squared distances for the pixels. PSNR analysis prevents vendors from cheating, because it ensures the processor has performed the required workload.

EEMBC, assisted by ECL, developed a special program for measuring PSNR. The program, PSNR.exe, runs on a PC, not on the benchmarked processor. (PSNR.exe runs on the same PC as EEMBC's special benchmark-control software, which is called the test harness.) Because the PSNR program runs outside the benchmark timing loop, it has no effect on the target processor's performance. PSNR.exe compares output frames from the MPEG encoder or decoder with YUV reference frames, and it can measure the quality of images with different pixel dimensions and bit depths. It can also measure the output quality of MP3 audio by comparing audio frames. The higher the PSNR, the better the quality of compression or decompression. (To learn more about using PSNR to measure video quality, follow the weblink in the "For More Information" box.)

Explaining the Odd PSNR Results

Although quality measurement is a vital part of DENbench, it would be inappropriate to require all media-playback systems to meet a single quality standard. Obviously, a cellphone can get by with lower-quality audio/video than a home theater system. For that reason, EEMBC allows members to tailor the quality of their MPEG encoding to the intended application. Any meaningful comparison of DENbench media-encoding and -decoding scores should pay attention to the associated PSNR numbers. When ECL certifies an MPEG EncodeMark or MPEG DecodeMark score, the lab's certification report lists the PSNR for each test run. ECL certification reports are available on the EEMBC website for all processors with certified and published DENbench scores.

Table 2 shows the PSNR measurements for all four processors benchmarked with DENbench so far. Our comparison revealed two odd results. First, the PSNR numbers for different processors running the same tests are remarkably similar—in several cases, the ratings are identical to a precision of six decimal points, which seemed impossible. Second, AMD's Geode NX1500 shows a PSNR of zero for two different runs of the MP3 audio-playback test. Because a higher PSNR is better, zero would seem to be the worst possible result, implying that Geode's audio playback consisted of either dead silence or sheer noise.

Intrigued by these odd results, *Microprocessor Report* turned to EEMBC and ECL for explanations. Why are the PSNR numbers in Table 1 nearly the same for all the processors? Answer: because they're all running the same MPEG source code, applying the same compression algorithms to the same datasets. Computers are deterministic, so, in theory, the output should always be the same, even with lossy compression algorithms. In practice, there are some slight differences, usually caused by different C compilers, flag settings, and rounding errors.

The determinism of these benchmarks would seem to make PSNR measurements superfluous, at least for the purpose of judging quality, if not for catching cheaters. However, all the DENbench testing to date has followed EEMBC's "out of the box" benchmark rules: vendors were allowed to use any publicly available compiler and documented flag settings, but they were forbidden to modify the benchmark source code. EEMBC allows another form of benchmarking, called "full fury" or optimized testing, which allows vendors to modify the source code or even replace it with hand-tuned assembly language. When vendors begin benchmarking their processors under the full-fury rules, the PSNR numbers might reveal larger differences.

Our second question—why Geode scored a PSNR of zero in two MP3 test runs—brought an interesting answer from ECL CEO Alan R. Weiss. In those test runs, Geode's decoded audio exactly matched the reference audio. Because a higher PSNR is better, and Geode's output was perfect, the PSNR numbers for those datasets are actually infinity. However, infinity would distort (to put it mildly) the calculation of an aggregate PSNR, so EEMBC decided that perfect results should be recorded as zero and not included in the aggregation. To correct the likely misimpression that zero is the *worst* possible result, or that the processor choked on those datasets, Weiss has recommended to EEMBC that future benchmark-certification reports note the number of perfectly matched frames produced during a test run.

To eliminate the untoward effects of using infinity in an aggregate PSNR calculation, ECL calculates PSNR values only for those frames that don't match perfectly. The main purpose of measuring PSNR is to ensure that the processor performs the workload correctly, and to give vendors a way of verifying that their implementations are reasonably accurate. ECL and EEMBC don't expect vendors to compete for the highest possible PSNR.

EEMBC Explores the Mysteries of Mars

It's interesting to see how EEMBC crafted the MPEG benchmarks to challenge the target processors with different tasks. The first MPEG encoder dataset, entitled "Graphic," has only seven 720- × 480-pixel frames. They consist of a ray-traced animation sequence with moving light sources, coronas, and reflections against a black background. This is the only MPEG encoder dataset with artificial images instead of photographic images, so it emphasizes certain kinds of MPEG artifacts.

The second MPEG encoder dataset, called "Railgrind," has a longer sequence of 30 frames at a lower resolution of 320 × 240 pixels. It shows a skateboarder sliding down a handrail and landing on the ground, a maneuver that creates a fast-moving color background as the video camera follows the subject. This dataset may cause "tearing"—a common MPEG artifact—in the lower background.

Another MPEG-2 dataset, "Sign," shows a person using sign language as the video camera zooms in against a complex color background. It's designed to reveal artifacts in the form of small color blocks near the bottom of the picture. The input frames are 352 × 256 pixels, but the MPEG-2 encoder must convert them to 352 × 240 pixels. Another dataset, "Zoom," has the video camera rapidly zooming out from a beach scene; it's a 30-frame sequence at 320 × 240 pixels. The last MPEG-2 encoder dataset, "Marsface," is a rotating black-and-white radar image of a topographical feature on Mars that conspiracy theorists believe is a giant sculpture of a face. (EEMBC disavows all knowledge of the true origin of this feature.) The Marsface sequence is 49 frames long and measures 192 × 192 pixels. Figure 1 shows a picture of the "face," located in the Cydonia region of Mars.

Benchmark Test	AMD Geode NX1500@6W	Analog Devices Blackfin ADSP-BF533	Freescale PowerPC MPC7447A	IBM PowerPC 750GX
MP3 Player—Data 1	7.798442	n/a	7.798442	7.798442
MP3 Player—Data 2	8.327399	n/a	8.326329	8.326329
MP3 Player—Data 3	0	n/a	22.029446	22.029446
MP3 Player—Data 4	26.676444	n/a	26.512523	26.512523
MP3 Player—Data 5	0	n/a	19.218792	19.218792
MPEG-2 Decode—Data 1	46.592435	n/a	46.592435	32.83448
MPEG-2 Decode—Data 2	55.913141	n/a	55.913141	36.696946
MPEG-2 Decode—Data 3	43.037027	n/a	43.037027	31.284806
MPEG-2 Decode—Data 4	36.722157	n/a	36.722157	24.799008
MPEG-2 Decode—Data 5	35.459819	n/a	35.459819	28.602152
MPEG-4 Decode—Data 1	32.83448	n/a	32.834480	32.83448
MPEG-4 Decode—Data 2	36.696946	n/a	36.696946	36.696946
MPEG-4 Decode—Data 3	31.284806	n/a	31.284806	31.284806
MPEG-4 Decode—Data 4	24.799008	n/a	24.799008	24.799008
MPEG-4 Decode—Data 5	28.602152	n/a	28.602152	28.602152
MPEG-2 Encode—Data 1	46.592435	46.548221	46.592435	46.592435
MPEG-2 Encode—Data 2	55.913141	57.110722	55.913141	55.913141
MPEG-2 Encode—Data 3	43.037027	42.918978	43.037027	43.037027
MPEG-2 Encode—Data 4	36.722157	36.738283	36.722157	36.722157
MPEG-2 Encode—Data 5	35.459819	32.057099	35.459819	35.459819
MPEG-4 Encode—Data 1	32.83448	32.83448	32.834480	32.83448
MPEG-4 Encode—Data 2	36.696946	36.696946	36.696946	36.696946
MPEG-4 Encode—Data 3	31.284806	31.284806	31.284806	31.284806
MPEG-4 Encode—Data 4	24.799008	24.799008	24.799008	24.799008
MPEG-4 Encode—Data 5	28.602152	28.602152	28.602152	28.602152
MPEG-2 Encode (FP)—Data 1	47.128665	n/a	47.131717	n/a
MPEG-2 Encode (FP)—Data 2	59.317249	n/a	59.301531	n/a
MPEG-2 Encode (FP)—Data 3	43.428701	n/a	43.360496	n/a
MPEG-2 Encode (FP)—Data 4	36.762583	n/a	36.764099	n/a
MPEG-2 Encode (FP)—Data 5	35.523269	n/a	35.525290	n/a

Table 2. ECL-certified DENbench scores include a report with quality ratings for each relevant dataset in the MPEG EncodeMark and MPEG DecodeMark minisuites. EEMBC uses its own peak signal-to-noise ratio (PSNR) program to measure the quality of encoded or decoded audio and video files. A higher PSNR indicates higher quality. Note that lower quality is acceptable for some applications, such as cellphones. (n/a = not available; vendor chose not to run those tests.)

Notice that none of the MPEG-2 encoding tests rises to the level of HDTV-quality video, and that none has more than a few seconds of video or animation. All are short snippets of video at relatively low resolution. EEMBC cites two reasons for these small datasets. First, the consortium's primary mission is to benchmark microprocessors, not complete systems. The MPEG video datasets are small enough to fit within manageable amounts of system memory alongside the MPEG benchmark code. Larger datasets—either in the form of higher-resolution video or longer video clips—would require streaming I/O, which would test the whole system and I/O connections as much as it would the processor.

EEMBC's second reason for keeping the video datasets small is to accommodate vendors running the benchmarks on processor simulators, not on actual silicon. Large datasets would be impractical on these software simulators, which are much slower than real chips. EEMBC says future versions of DENbench may expand the suite to include larger video datasets, an improvement we would welcome, even if it requires streaming I/O. Home theater systems, HDTV, and other high-end multimedia applications are fast-growing product categories eagerly courted by embedded-processor vendors.

MP3 Benchmark Was Long Overdue

Most welcome are the MP3 digital-audio playback tests in the DENbench suite. Portable digital-audio players are taking the consumer-electronics world by storm, rapidly displacing portable CD and cassette players, especially among young people. Apple's hugely popular iPod players and iTunes online music service have legitimized downloadable music, building on a revolution started by renegade peer-to-peer file-sharing networks like Napster and Kazaa.

Apple's preferred music format is Advanced Audio Coding (AAC), not MP3, but iPods can play MP3 files, too. (AAC delivers higher fidelity and smaller files.) MP3 remains the most popular format on digital-audio players from other vendors and on global file-sharing networks. Both AAC and MP3 are part of the MPEG-2 standard, and EEMBC's MP3 playback test is part of the MPEG Decoder-Mark minisuite. The MP3 benchmark consists of one basic test using five different datasets.

The test is an integer implementation of the ISO 13818-3 MP3 decoder at sampling frequencies of 16KHz, 22.05KHz, and 24KHz. Note that EEMBC omits the CD-standard 44.1KHz sampling frequency and the DAT-standard 48KHz frequency. The omission was deliberate, because EEMBC wanted to concentrate on the lower-fidelity sampling rates typically used by portable MP3 players, PDAs, audio-enabled cellphones, and file-sharing networks. For the same reason, EEMBC uses the baseline MPEG reference code without some optimizations that are part of the standard. The MPEG reference code relies heavily on Huffman decoding and iDCT routines, much like the MPEG-2 video decoding benchmarks.

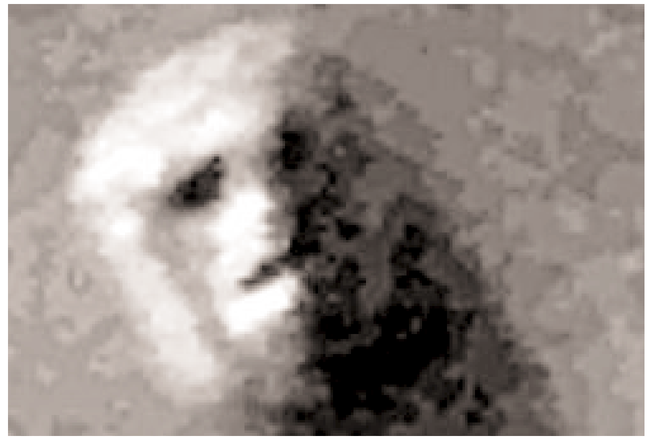


Figure 1. Is this "face" an ancient Martian sculpture or a suggestive optical illusion? EEMBC uses an animated video of this actual topographical feature on Mars to benchmark MPEG-2 video encoding on embedded processors. This photo, extensively computer-enhanced by NASA, was taken by one of the Viking orbiters in the 1970s.

EEMBC's five MP3 datasets are static files, not I/O datastreams. For all of them, EEMBC selected a rousing rendition of "Jupiter, the Bringer of Jollity" from *The Planets*, Gustav Holst's early–twentieth century orchestral suite. This piece of music isn't particularly popular among the teenagers who are the biggest customers for MP3 players, but it's free of copyright entanglements and has a wide range of instrumentation and dynamics, which is why audiophiles often use it to show off their stereo systems. (It also happens to be the school song at Westwood High in Austin, Texas, where the children of ECL's Alan Weiss are students. If Weiss were a little more hip, he might have selected rock guitarist Joe Satriani's "With Jupiter in Mind," from the album *Crystal Planet*.)

The five MP3 datasets are encoded at bit rates of 160Kb/s stereo, 128Kb/s stereo (the highest-fidelity common bit rate for portable MP3 players), a variable bit rate of 48Kb/s–128Kb/s stereo (the lower end of that range is common on cellphones), 64Kb/s stereo (a popular compromise between fidelity and file size), and 48Kb/s mono.

As with the MPEG video tests, the MP3 audio tests include a quality measurement using the PSNR method. In this case, a program developed by EEMBC compares each pulse-code modulation (PCM) frame with an Audio Interchange File Format (AIFF) reference file. Again, a higher PSNR is better. If an EEMBC member certifies the benchmark results through ECL and publishes the scores, EEMBC's website posts a report showing the PSNR for each MP3 dataset.

Cryptographic Benchmarks Vital for Security

Before long, a processor without security features will seem like a chip without contacts—not very useful. More and more data traffic is moving over networks in secure packets, and content providers of all kinds are eager to protect their copyrighted property with DRM. Unfortunately, security

Few Surprises In First DENbench Scores

On the same day EEMBC officially announced its new Digital Entertainment benchmark suite, the consortium also published the first certified DENbench scores. The first EEMBC members to publish certified scores are AMD, Analog Devices, Freescale, and IBM Microelectronics.

AMD's scores are for the Geode NX1500@6W, an x86-compatible embedded processor that's actually a relabeled Athlon XP desktop PC chip. However, in a departure from EEMBC's usual practice, AMD didn't perform the benchmarking. Instead, ECL—with AMD's kind permission—benchmarked the Geode NX1500 to help test and calibrate the DENbench suite. To produce solid baseline code, ECL deliberately used a generic GNU compiler (GCC 3.3.3) and didn't set the compiler's most important performance-optimization flags (–Ospeed and the microarchitecture-specific –cpu switch). Therefore, the current DENbench scores for the Geode NX1500 don't reflect the processor's potential performance. We include the scores here because they're published, certified scores on the EEMBC website and provide a baseline reference for future x86 benchmarking with this suite.

Analog Devices tested its Blackfin ADSP-BF533, a high-performance embedded processor and DSP. Freescale benchmarked its MPC7447A, a high-performance embedded PowerPC processor. IBM tested its PowerPC 750GX, another high-performance embedded processor. The Freescale and IBM processors are the most similar in this group, because they share a common RISC architecture and run at gigahertz clock frequencies. The AMD Geode is based on a different CISC architecture, but it runs at a similar clock speed and has similar capabilities. Analog Device's Blackfin, though based on a proprietary DSP architecture, has numerous features that qualify it for general-purpose embedded processing.

EEMBC members aren't required to run all the benchmark tests when they use the DENbench suite, so Analog Devices chose to run only the MPEG EncodeMark minisuite. Freescale ran all the

Processor	MPEG DecodeMark	MPEG EncodeMark	MPEG-2 Encode (FP)	CryptoMark	ImageMark	DENmark Total
AMD Geode NX1500@6W (1.0GHz)	785.1	587.4	30.2	509.3	918.9	131.7
Analog Devices Blackfin ADSP-BF533 (594MHz)	n/a	355.5	n/a	n/a	n/a	n/a
Freescale PowerPC MPC7447A (1.4GHz)	1,506.30	1,281.50	67.2	1,263.30	1,709.40	257.6
IBM PowerPC 750GX (1.0GHz)	1,054	967.9	n/a	903	1,090.40	173.6

These are aggregate scores for each of the four minisuites in the Digital Entertainment suite, plus the floating-point version of the MPEG-2 encoder tests. EEMBC members don't have to run all the tests, so scores for some minisuites are not available (n/a). Keep in mind that these aggregate scores represent results for as many as 69 individual test runs, and that specific tests are often more informative than aggregate scores are. In addition, note that the AMD Geode NX1500 was benchmarked without using the best compiler and performance-optimization flags.

tends to be compute intensive, placing additional burdens on the processor. That's why hardware support for cryptography and random-number generation is showing up on everything from x86-compatible PC processors to low-power embedded processors. EEMBC is striving to keep pace with these trends by introducing a wholly new minisuite of cryptographic benchmark tests within the DENbench suite.

EEMBC's CryptoMark is an aggregation of four benchmark tests for common cryptographic standards and algorithms: the latest Advanced Encryption Standard (AES), the older Data Encryption Standard (DES), the popular Rivest-Shamir-Adleman (RSA) algorithm for public-key cryptography, and Huffman decoding for data decompression. So far, EEMBC has prepared documentation for only the AES component of the minisuite, but all the components will be familiar to engineers working in this field, and datasheets for all the CryptoMark tests are in progress.

AES is a royalty-free cipher conforming to U.S. government standards and is intended to eventually replace DES, which is more than a quarter-century old. The AES cipher is the foundation for several cryptographic protocols, including Transport Layer Security (TLS), Secure Sockets Layer (SSL), Secure Shell (SSH), and Internet Protocol security (IPsec). Web servers and browsers use some of these protocols for secure connections and transactions.

ECL developed the AES benchmark using original source code from Vincent Rijmen and Joan Daemon, who created the Rijndael algorithm. Rijndael is an iterated block cipher with variable-length blocks and keys. EEMBC's benchmark implements all three key lengths (128, 192, and 256 bits) for each loop iteration. The test program performs 16 passes and then decrypts the encrypted data to verify correctness. The program also implements the so-called wide-trail strategy that prevents certain kinds of attacks on the AES cipher.

First DENbench Scores *continued*

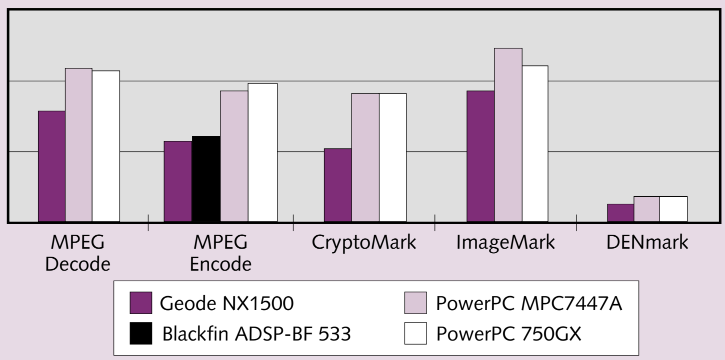
DENbench tests on the MPC7447A, and IBM ran all the tests on the 750GX except for the optional floating-point versions of the MPEG-2 encoder. The chip with the highest overall DENmark score was Freescale's MPC7447A, which at 1.4GHz also boasts the highest clock speed in this group. This result isn't surprising, because the MPC7447A also achieved the highest-ever ConsumerMark score in EEMBC's Consumer 1.1 benchmark suite, the predecessor of DENbench.

In fact, the DENmark scores roughly correspond to the clock speeds of the benchmarked processors. AMD's Geode NX1500 runs at 1.0GHz and scored 131.7 DENmarks. IBM's 750GX also runs at 1.0GHz but scored 173.6 DENmarks, probably thanks to its larger L2 cache (1MB vs. 256KB), a better compiler (Green Hills), and processor-specific compiler optimizations. Freescale's MPC7447A, at 1.4GHz, scored 257.6 DENmarks, which is 48% faster than the 750GX—not too surprising for a processor sharing the same RISC architecture and running at a 40% higher clock speed, even though it has a smaller L2 cache (512KB vs. 1MB). This table shows each processor's score for all the DENbench minisuites.

Of course, DENmarks are aggregations of all 69 DENbench tests, which isn't the best way to compare processors. Smart shoppers will delve deeper into the certified benchmark reports (available on the EEMBC website) to find the test results most relevant for their target applications. For instance, a comparison of these chips for performance per megahertz indicates that the PowerPC processors accomplished more work per clock cycle than the other processors, as this chart shows.

There are numerous ways to crunch benchmark scores, and EEMBC's benchmark reports can provide hours of spreadsheet entertainment. If power consumption matters—for mobile systems, power is everything—Blackfin reigns supreme among these processors. Although its MPEG EncodeMark was the lowest such score in this group, the Blackfin BF533 typically consumes only 0.5W, significantly less than 6W for the Geode NX1500, 8.3W for the 750GX, and 18.3W for the MPC7447A. In a performance-per-watt comparison, Blackfin is about seven times more efficient at MPEG encoding than the other processors are. (Note that ECL doesn't attempt to verify power consumption when certifying benchmark scores; the power numbers are provided by vendors. As a separate project, EEMBC is working on verifiable power benchmarks.)

For complete DENbench scores and benchmark reports, see this page on EEMBC's website: *www.eembc.hotdesk.com/digital_entertainment.html.*



PowerPC processors from Freescale and IBM fared better in a performance per megahertz comparison than AMD's x86-compatible Geode NX1500 and Analog Devices' Blackfin BF533. This comparison is a rough measure of architectural efficiency, as it reflects benchmark iterations per clock cycle. The Geode NX1500 would almost certainly do better if benchmarked with a better compiler and optimization settings.

Although EEMBC's benchmark source code is proprietary, the source code for the Rijndael algorithm is publicly available. *Microprocessor Report* examined the x86 assembly-language code and found that it makes extensive use of bit shifting, memory moves, and bitwise operations (especially XOR).

Although EEMBC hasn't released documentation for the other components of the CryptoMark minisuite—DES, RSA, and Huffman—they exercise similar capabilities of the processor. EEMBC derives the aggregate CryptoMark score from the geometric mean of all four cryptography tests. And, as with all the minisuite scores, CryptoMark is one component of the overall DENmark score.

DENbench Is Late, but Still Up to Date

EEMBC spent years debating and developing DENbench. Despite the arduous process, however, the finished benchmarks look almost as fresh as if they were developed last week. All the tests are relevant, useful, and immediately applicable to a variety of embedded processors designed for the red-hot consumer-electronics market.

Indeed, it would be illuminating to run many of these benchmark tests on PCs and other systems. Even servers and networking equipment would be relevant test beds for the CryptoMark tests. If some of the datasets are too small or otherwise unsuitable for more-powerful processors and systems, there's nothing to stop EEMBC members from creating their own private datasets. ECL wouldn't certify those benchmark results for publication, of course, but they would be useful for internal testing and performance optimization. EEMBC's lab, ECL, has already done the hard work of developing the benchmark code.

Our nitpicks are about what EEMBC omitted, not what it included. In the ImageMark suite, for example, we'd like to see EEMBC replace the missing high-pass grayscale

For More Information

EEMBC's Digital Entertainment benchmark suite (DENbench) is available now to EEMBC members. For more information about DENbench, including certified benchmark reports and datasheets, visit *www.eembc.org*.

For more information about using peak signal-to-noise ratios (PSNR) to measure video quality, follow this link: *www.broadcastpapers.com/sigdis/Snell&Wilcox QualityMeasure02.htm.*

filter with another sharpening filter used by digital cameras: the unsharp mask. Unsharp masking (ill-named after an old wet-darkroom technique) is a math-intensive iterative algorithm that "sharpens" an image by examining the pixels and exaggerating their differences along subject boundaries. Almost all digicams and scanners routinely apply unsharp masking to the images they capture, unless the user turns off the filter. Colorspace conversions for Adobe RGB and sRGB would be useful, too.

We would also like to see higher-resolution datasets for the MPEG video encoders and decoders, as well as benchmarks for H.264-standard video. Another improvement would be MPEG tests that simultaneously decode both an audio stream and a video stream. DENbench treats audio and video as separate datastreams with their own separate tests, even though multimedia streams are far more common. Perhaps EEMBC could include simultaneous audio/video encoding and decoding in its future multiprocessing benchmarks, which the consortium is now working to define.

On a positive note, we applaud EEMBC's decision to subdivide this large benchmark suite into minisuites with their own figures of merit. It gives EEMBC members the flexibility to run only the benchmark tests relevant for their purposes and still report meaningful scores. The PSNR measurements are a clever way to check the quality of audio/video transformations without penalizing EEMBC members for choosing a lower level of quality, if it's appropriate for their target application. We also like the large variety of datasets provided for all the benchmarks in this suite. Multiple datasets challenge the processors in different ways and produce more-detailed results.

Overall, then, we think DENbench hits a home run. The years of dogged committee work and tedious software development haven't gone to waste. There's still room for improvement, but that's always the case. Now we want to see more EEMBC members step forward and report certified DENbench scores. ◇