EMBEDDED x86 PROCESSOR PERFORMANCE RATING SYSTEM

The Problem

Clock speed has become a standard marketing tool for computer vendors and microprocessor companies seeking to gain mindshare with system designers. The benefit of clock speed as a measure of performance is that it's an easy to understand single number, where the higher the number - the better the performance.

However, our initial research showed that clock speed, while being simple to understand, can also mislead the design community. Instead of clarifying the market by providing meaningful information, it confuses the market by being only weakly correlated to real-world performance.

Synchromesh Computing decided to focus on the embedded market, which includes such diverse applications as Information Appliances, Thin Clients, industrial devices and convergence devices such as web enabled mobile phones. We wanted to find out how clock speed compared to the way real processors, installed in real systems, performed on real-world tasks.

While we recognize that there are a number of excellent computing architectures available today, we chose to focus this particular study on x86 processors used in embedded computing platforms.

This report will describe the benchmark requirements, present our benchmark suite, explain how scores are calculated, and provide examples using several popular processors.

System Benchmarking Requirements

If we were just studying embedded processors alone, we would turn to the industry standard EEMBC benchmarks (http://www.eembc.org). However, those benchmarks are focused almost exclusively on the processor itself and not the processor interacting with the other components and operating system. In other words, we needed a systems-level benchmark suite.

In developing our benchmark suite, our goal was to more closely replicate real-world usage in order to come up with a pragmatic measurement of processor performance. We knew we would need to incorporate performance parameters such as:

- raw CPU processing power
- L1 and L2 cache size and speed
- memory bandwidth
- multimedia performance
- file system performance
- available system headroom when running an application
- system storage (hard disk access) performance
- graphics processing
- communications, such as Instant Messaging

To do this, we created a new benchmark suite called EPRS (Embedded Processor Rating System). This benchmark suite uses recognized industry-standard benchmarks, well-respected by Performance Analysis Engineers for accuracy. We also included a number of popular benchmarks from the PC world because of their widespread usage and acceptance. In addition, two of the benchmarks (Surfbench™ and the IM Chat Test) were newly created because no industry standard of sufficient quality was available.

The Solution: EPRS Benchmark Suite

The EPRS benchmark suite is made up of the following tests:

- HINT for CPU and memory subsystem performance
- STREAM for memory subsystem testing
- SiSoft SANDRA™ for CPU, multimedia, memory and cache performance
- Synchromesh Computing Surfbench™ for Internet-oriented benchmarking
- HDBench for CPU, memory, graphics and hard drive performance
- Winbench '99 over RDP/ICA for thin client benchmarking (using Windows® CE and Windows XP)
- Synchromesh Computing IM Chat Test

HINT (Hierarchical INTegration), written by Ames Research Lab, Department of Defense, has gained a reputation for being the most scalable and accurate measure of CPU and memory subsystem performance.

Most benchmarks measure either the number of operations that can be performed in a given time period, or the time required to perform a given fixed calculation. HINT does neither; rather, it performs a particular calculation (estimating upper and lower bounds for the definite integral of a monotone-decreasing function) with ever increasing accuracy. The accuracy of the result at any given time is called the "Quality"; we may measure the improvement in quality at any given time as "Quality Improvements per Second," or QUIPS. As the computation progresses and the quality of the result improves, more memory and more operations are required to improve the answer: Higher is better. HINT curves are a function of raw CPU processing power, L1 and L2 cache size and speed, and main memory bandwidth.[1]

1. Nicholas Coult, Ph.D., Assistant Professor of Mathematics, Augsburg College

More information on HINT can be found at: http://hint.byu.edu/documentation/Gus/HINT/ComputerPerformance.html#Quips

STREAM, written by Dr. John McAlpin of Silicon Graphics, is another open source, industry-standard benchmark suite that does an excellent job of measuring sustainable memory bandwidth. Embedded processors are often connected to the Internet, and almost always must process large amounts of data typical of multimedia bit streams.

The STREAM benchmark is a simple synthetic benchmark program that measures sustainable memory bandwidth (in MB/s) and the corresponding computation rate for simple vector kernels. Computer CPUs are getting faster much more quickly than computer memory systems. As this progresses, more and more programs will be limited in performance by the memory bandwidth of the system, rather than by the computational performance of the CPU.
As an extreme example, several current high-end machines run simple arithmetic kernels for out-of-cache operands at 4-5% of their rated peak speeds --- that means that they are spending 95-96% of their time idle and waiting for cache misses to be satisfied.
The STREAM benchmark is specifically designed to work with datasets much larger than the available cache on any given system, so that the results are (presumably) more indicative of the performance of very large, vector style applications.[2]

More information on STREAM can be found at: http://www.austin.rr.com/mcalpin/papers/bandwidth/bandwidth.html

Sandra, a very useful PC tool written by SiSoft, stands for the System Analyzer, Diagnostic and Reporting Assistant. It also has a benchmarking component that has gained a certain amount of credibility by being quoted by other labs searching for a quick and direct way to measure CPU performance.

Surfbench, written by Synchromesh Computing, measures the amount of headroom available in a system running a set of user scenarios, including playback of an MP3 file, two MPEG files, a Macromedia Flash movie (audio and visual combined), web page browsing of complex web pages, and multiple real-time clocks. Unlike any other benchmark in existence, Surfbench seeks to measure the amount of processing power left available to the user while running these scenarios. We believe that this is an excellent way to measure real-world performance for embedded processors. Synchromesh Computing is working with key microprocessor and system vendors in the thin client, Internet appliance, PDA, and mobile phone markets to make this a new industry standard.

HDBench is a useful tool for measuring general integer, floating point, hard disk, memory, and graphics performance. It is quite popular in Japan.

Back in the last century (actually 1999 or so), the standard for PC benchmarking included Ziff-Davis' **Winbench '99**. It may have outlived its usefulness as a standalone PC benchmark, but companies such as Microsoft® and Wyse use it to test and benchmark thin clients. It has a particularly strong graphics component (what was a weakness in the PC space – being susceptible to simply faster graphics cards – is actually useful when running this over RDP/ICA protocol). All of the processing occurs on the server, while the graphics is sent down the ethernet wire to the client (the target under test).

Because these processors support both Windows XP and Windows CE, we benchmarked both. CE is particularly important in high-end embedded environments.

Finally, Synchromesh Computing realized that Instant Messaging (IM) is a hot application. There are a number of services such AOL IM (AIM)™, Yahoo Instant Messaging™, and MSN Messaging™. However, to avoid the problems surrounding latencies associated with public services such as those, we decided to use an open source IM system called Jabber. In creating the **Synchromesh Computing IM Chat Test**, we created a system that launched two IM windows on the target, and fed those sessions using Rational Visual Test™ scripts.

EPRS Benchmarking Example

Computing the scores was the most interesting part of this study. Because we questioned clock speed as a measure of performance, we had to select a processor at a particular clock speed and set that as the baseline. We chose the 533 MHz VIA Centaur processor because its clock rate and power consumption are in the middle of what is available for the embedded market and its architecture is relatively simple (scaler versus superscaler). We chose to compare it to the AMD Geode™ line of processors because they have offerings both on the lower end (high integration) and higher end (superscaler cores). The processors used in this study are provided in the following list:

- AMD Geode™ GX 466@0.9W processor runs at 333 MHz
- AMD Geode™ GX 533@1.1W processor runs at 400 MHz
- VIA Centaur processor runs at 533 MHz
- VIA Nehemiah processor runs at 1 GHz
- AMD Geode™ NX 1500@6W processor runs at 1 GHz

System configuration information is provided in Table 1.

2. John D. McAlpin, creator, STREAM

Table 1. System Configurations[1]

Manufacturer	Processor	Speed	RAM	Hard Disk	Motherboard	OS
VIA	Nehemiah	1 GHz	DDR 256MB (shared)	18 GB	EPIA-M (133 MHz FSB)	Win XP Home Win CE
VIA	Centaur	533 MHz	SDRAM 256 MB (shared)[2]	Maxtor 60 GB	EPIA-M (133 MHz FSB)	Win XP Home Win CE
AMD	Geode GX	333 MHz	DDR 256 MB (shared)	Western Digital 80 GB	Hawk	Win XP Home Win CE
AMD	Geode GX	400 MHz	DDR 256 MB (shared)	Western Digital 80 GB	Hawk	Win XP Home Win CE
AMD	Geode NX	1 GHz	DDR 256 MB (shared)	Maxtor 120 GB	Modified ASUS A78V8X-MX	Win XP Home Win CE

1. *All machines were configured with a similar amount of memory, same performance in hard disks (same buffers and average seek times), same operating systems, and in general were made to be as identical as possible.*
2. VIA *does not support DDR in Centaur.*

After selecting the processors and configurations, we ran each system through the benchmark tests that make up the EPRS benchmark suite. We generated the overall scores for each of the tests. Some benchmarks produce a single figure of merit (for example, HINT generates QUIPS). Some, such as STREAM, produce a number of scores. For these tests, we used the Geometric mean within that suite so that the extremes were slightly discounted (benchmark suites such as EEMBC and SPEC, typically use the geometric mean, rather than the arithmetic mean).

Each of the benchmark tests were also normalized in comparison to a constant. For example, consider the Surfbench benchmark score where a 333 MHz Geode GX processor is compared against the 533 MHz VIA Centaur. We start with the raw scores as follows:

533 MHz VIA Centaur: 83.57% Headroom Available

333 MHz Geode GX processor: 77.43% Headroom Available

To normalize for clock speed, we multiply the Geode GX processor score times the clock speed of the reference (in this case, the VIA Centaur), and then divide by the score obtained by the VIA itself.

$$77.43 * 533 / 83.57 = 493.84$$

What this says is that on this particular test, the 333 MHz Geode GX processor is running at a performance level of 493.84 MHz in comparison to the reference processor from VIA running at 533 MHz. This process was repeated for each of the tests included in the benchmark suite.

Finally, the EPRS score is computed by taking the geometric mean of all of the accumulated scores for each benchmark test suite. In doing this for the 333 MHz Geode GX processor, we get an EPRS score of 505.89 MHz. While not as fast as the reference chip, the 333 MHz Geode GX processor performed much better than its clock speed would indicate. The EPRS benchmark suite scores for the 333 MHz Geode GX processor are presented in Table 2.

Table 2. 333 MHz Geode™ GX Benchmarks

Benchmark Tests	How Measured	Scores
Synchromesh Computing Surfbench	Headroom Score	493.84
HINT	QUIPS Score	446.59
STREAM	Geomean	849.53
Sandra Multimedia	Geomean	399.63
Sandra Filesystem	Geomean	318.96
Sandra Cache and Memory	Geomean	454.34
Sandra Memory Bandwidth	Geomean	1274.82
HDBench	Overall Score	380.95
Winbench '99 Business Graphics RDP/ICA on Windows XP	Score	297.65
Winbench '99 Business Graphics RDP/ICA on Windows CE 4.2	Score	656.12
Synchromesh Computing IM Chat Test	Overall Score	479.70
	EPRS =	**505.89**

Analysis shows that the support of DDR (double data rate) memory combined with better memory subsystem in general (and superior integration of hardware and software components) helps propel the 333 MHz Geode GX processor to achieve an excellent rating.

We can see the effect of the better memory subsystem, for example, by highlighting HINT, which as Table 2 shows, indicates a QUIPS score of 446.59.

Figure 1 compares the effect of HINT on the 333 MHz Geode GX processor, the 400 MHz Geode GX processor and the VIA Centaur.

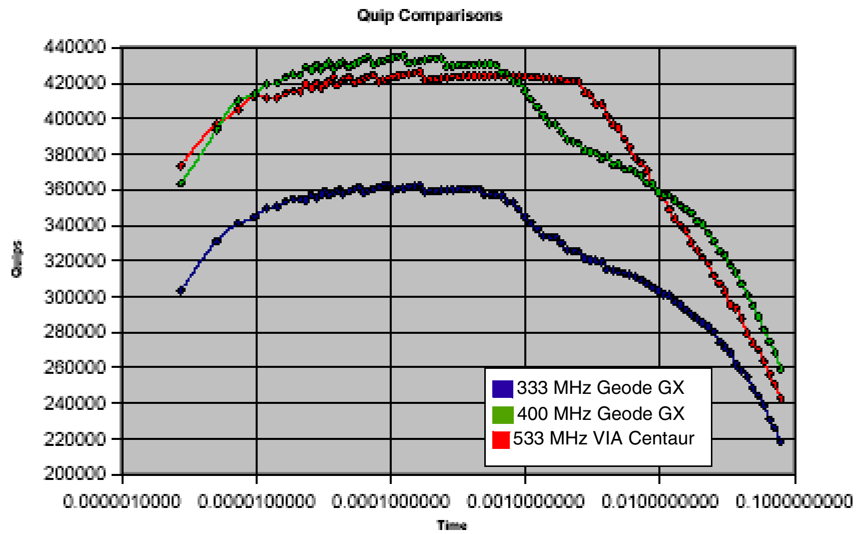

Figure 1. Effect of HINT Benchmark on 333 MHz & 400 MHz Geode™ GX Processors and 533 MHz VIA Centaur

The 400 MHz Geode GX processor keeps up with the 533 MHz VIA Centaur. At 0.00001 seconds, the Centaur's bigger L1 caches kick in, but by 0.0010 seconds, the superior DDRAM performance takes over. The tracking of the 400 MHz Geode GX processor is quite close to the 533 MHz VIA Centaur.

To better appreciate this, we created a chart (Figure 2) for these three processors comparing all of the benchmark scores. Once again, we see that the Geode GX processors hold their own and occasionally surpass the “faster” VIA Centaur.

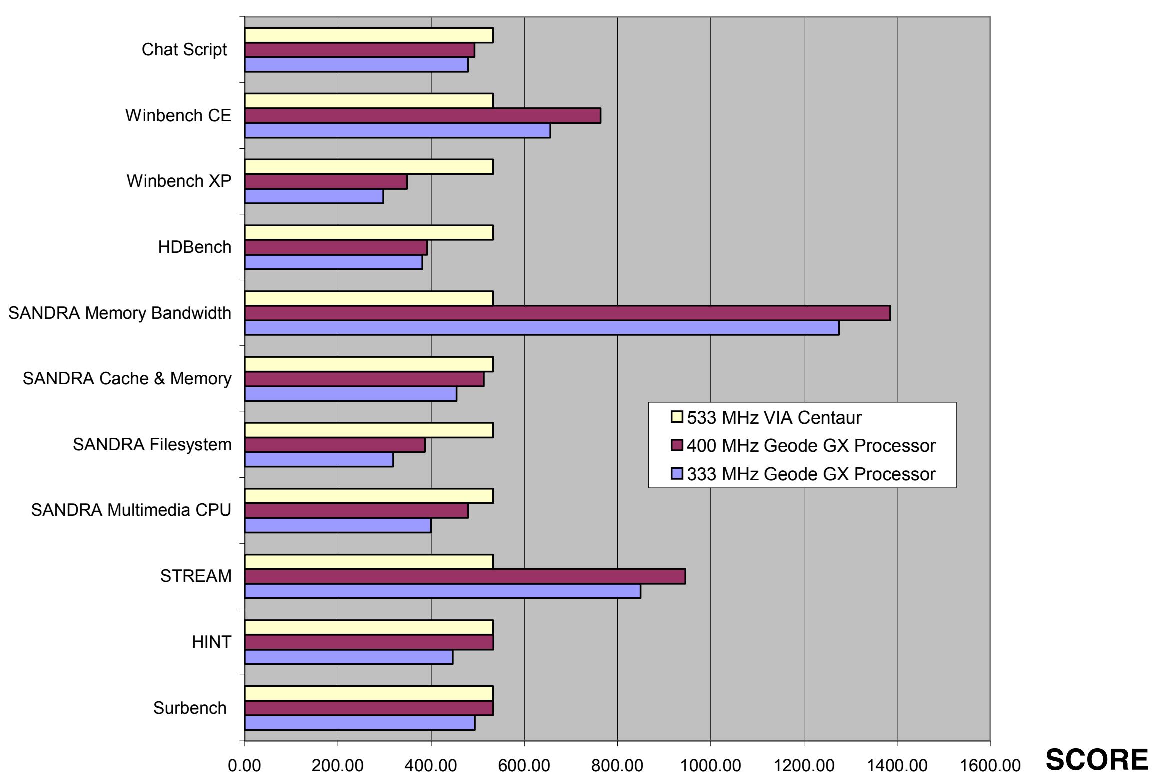

Figure 2. 333 MHz & 400 MHz Geode™ GX Processors Compared to the 533 MHz VIA Centaur

Taking this a bit further, we decided to discover the EPRS for the 1 GHz AMD Geode NX processor, and for comparison we used the 1 GHz VIA Nehemiah. Applying the benchmark suite produced the results presented in Table 3.

Table 3. 1 GHz Geode™ NX Benchmarks

Benchmark Tests	How Measured	Scores
Synchromesh Computing Surfbench	Headroom Score	1148.90
HINT	QUIPS	2135.66
STREAM	Geomean	2464.06
Sandra Multimedia	Geomean	1000.00
Sandra Filesystem	Geomean	1219.59
Sandra Cache and Memory	Geomean	2021.05
Sandra Memory Bandwidth	Geomean	3258.05
HDBench	Overall Score	1072.19
Winbench '99 Business Graphics RDP/ICA on Windows XP	Score	1204.23
Synchromesh Computing IM Chat Test	Overall Score	1036.70
	EPRS =	**1554.88**

The EPRS score indicates that the 1 GHz Geode NX processor, although clocked at 1.0 GHz, performs at approximately 1.6 times the 1 GHz VIA Nehemiah. Figure 3 presents a comparison of the benchmark test scores for the 1 GHz Geode NX processor and the 1 GHz VIA Nehemiah.

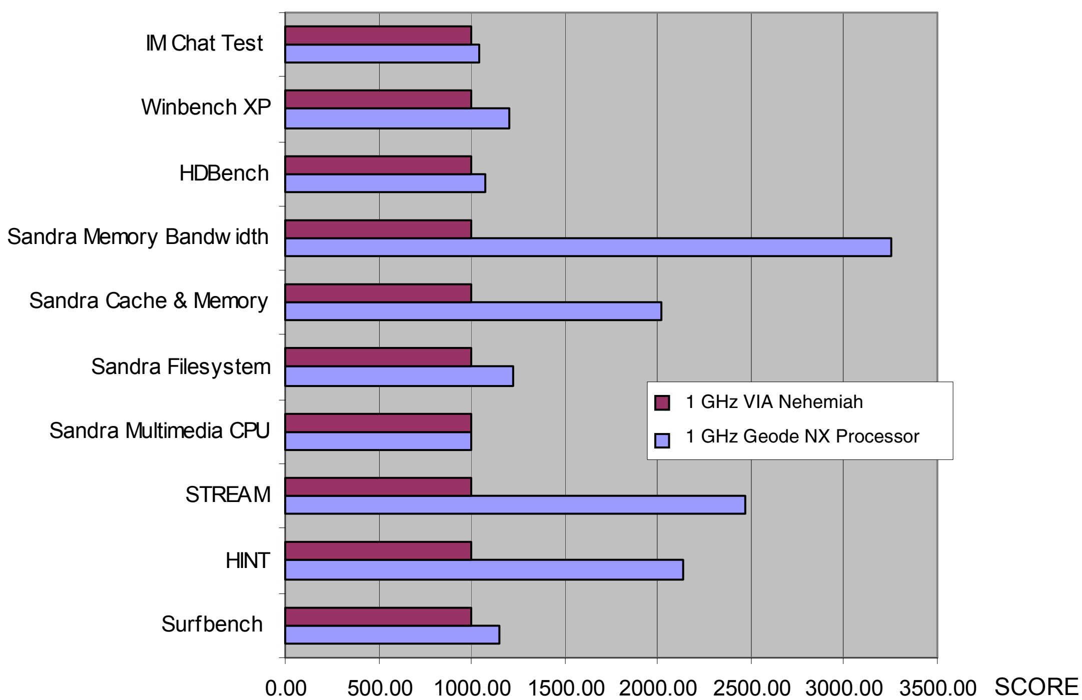

Figure 3. 1 GHz Geode™ NX Processor Compared to the 1 GHz VIA Nehemiah

Conclusion

Clock speed is not the only factor that goes into performance, and with this study we have shown that it is logical to rate a processor based on a suite of recognized benchmarks. Memory subsystem, graphics, bus speed, and instructions per clock cycle all play a role in overall performance.

Indeed, we believe that processor companies in this application space need to move beyond clock speed, and use EPRS benchmarking to position their products, to price their products, and provide meaningful product information to their customers.

Synchromesh Computing welcomes all x86 vendors to submit their systems for benchmarking and certification.

About Synchromesh Computing

Synchromesh Computing was formed to bring the rigor of industry-standard benchmarking found in such organizations as SPEC and EEMBC to the desktop PC, Internet appliance, thin client, PDA, and mobile phone worlds. A sister company of the EEMBC Certification Laboratory (ECL), it has certified hundreds of benchmark scores for EEMBC, the Embedded Microprocessor Benchmark Consortium. No benchmark scores can be published without ECL's certification.

Located in Austin, Texas with customers world-wide, both ECL and Synchromesh Computing have the full faith and confidence of the entire microprocessor and computer industry and are known for fairness in benchmarking and performance.

Synchromesh Computing
6507 Jester Blvd., Suite 511
Austin, TX 78750 - USA
Telephone: (512) 219-0302
Fax: (512) 219-0402
www.synchromeshcomputing.com