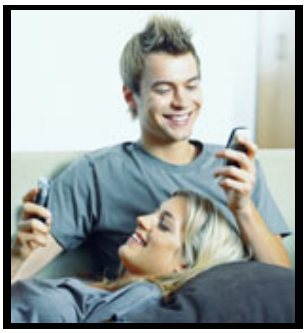
JOURNAL

INSIDE THIS ISSUE:

for the latest in benchmark scores
www.eembc.org

FROM THE PRESIDENT

Testing Your Mobile Phone's Java Performance



There are many reasons why it's important to test the Java performance in a mobile phone. But the number one reason is that, as a mobile phone user, you might be one of the estimated 220 million mobile gamers expected by 2009. In this case, your phone's Java performance will play a big role in the quality of your user experience. The good news is that there are many benchmarks available for testing a mobile phone's Java performance. The bad news is that there are many benchmarks available for testing a mobile phone's Java performance. As an industry-standard consortium, it's time we took action to ensure that the proper tests were being conducted on the world's mobile phones.

EEMBC has always closely guarded its benchmark source code to protect the benchmarks' credibility, and only EEMBC members are allowed to publish benchmark scores for their own products. With our GrinderBench™ Java benchmarks however, more than just processor performance is being measured; the score results show the performance of a whole Java implementation. This being the case, we've come to recognize that a different set of rules from our processor-centric benchmarks ought to apply to GrinderBench.

Based on a decision by the EEMBC Board, our consortium has agreed to make binary versions of the algorithms in the GrinderBench Java benchmark suite available for free-of-charge download from a new Web (and WAP) site, making it possible for anyone with a Java-enabled mobile phone to run the benchmarks and test their phone's performance in Java-enabled games, encryption, internet connectivity, and other applications.

The new site also presents scores for a wide variety of mobile phones that were tested under controlled conditions by EEMBC members. The preliminary data we've seen is already

(continued on page 3)

AN EEMBC MEMBER SPEAKS OUT

Benchmarking Multicore Platforms: Just how can I know?

By John Goodacre, ARM

Benchmarking has long been invoked as the best way for designers to determine how well any particular solution will work in a given application. The truth is that the abstraction of a solution into some manageable set of benchmarking tests provides the answer that one size really doesn't fit all.

Within EEMBC, there are multiple suites of tests, each with a number of test kernels that exercise a candidate platform in some contrived or generalized manner. Even so, we still see huge numbers of alternative suites and kernels being offered right across the computer industry. If I were to try and offer an overview of the benchmarking world, I'd probably start by splitting it into three distinct aspects.

Algorithmic excellence: benchmarks of this form concentrate solely on a single task that is typically focused on some form of signal processing.

(continued on page 4)

From the lab

NEWEST BENCHMARK SCORE REPORTS

freescale semiconductor

MC7448 1.7 GHz
Pre-production Silicon

Automotive/Industrial
Out-of-the-Box

Consumer
Out-of-the-Box

Digital Entertainment
Out-of-the-Box

Digital Entertainment
Optimized

Networking Version 1.1
Out-of-the-Box

Networking Version 1.1
Optimized

Networking Version 2.0
Out-of-the-Box

Office Automation
Out-of-the-Box

Telecom
Out-of-the-Box

Telecom
Optimized

IBM

IBM 970FX 2 GHz
Production Silicon

Automotive/Industrial
Out-of-the-Box

Consumer
Out-of-the-Box

Digital Entertainment
Out-of-the-Box

Networking
Out-of-the-Box

Networking Version 2.0
Out-of-the-Box

Office Automation
Out-of-the-Box

Telecom
Out-of-the-Box

Power Consumption: The Next Great Benchmarking Frontier

Alan R. Weiss, EEMBC Certification Laboratory (ECL, LLC)

For some time now, the semiconductor and processor industries, and even suppliers of embedded software, have considered power as a parameter with as much, or even more, importance than performance. EEMBC has not ignored this industry focus and has been working on a consistent, reliable, and certifiable method to measure the energy required to run EEMBC benchmarks in devices under test (DUTs). A power technical advisory group, led by Shay Gal-On of PMC-Sierra, has been discussing various techniques and concepts, with ECL acting as a technical advisor and laboratory to perform experiments. Dr. David Kaeli of Northeastern University has been another technical advisor, offering feedback and ideas.

This group's latest milestone is a white paper, "Measuring Power with National Instruments LabView™." The title gives some indication of our approach. Rather than use a power meter for measurements, we believe EEMBC and its stakeholders will be best served by the use of National Instruments LabView software running on a host platform (such as Windows XP or Linux) containing a digital-to-analog converter (DAC) board. This board is connected to a breakout board, which is turn is connected to the DUT, or, more accurately, to power access points to which a power resistor has been soldered.

One major advantage of using this PC-based measurement setup is that the test results are automatically recorded. This system logs all of our samples to a file and provides a complete audit trail of our work. These datapoints can be plotted on a graph, resulting in a reasonable picture of power consumption while the benchmark is being run. Warm up periods are taken into account to assure accuracy and stability of measurements. We will measure using two different sampling frequencies to factor out possible aliasing concerns.

ECL has designed a method to modify the EEMBC Portable Test Harness to trigger the power measurement system programmatically. By automating the system, we are providing useful tools to EEMBC members and ECL-certified power consumption measurements for their customers.

The following questions and answers address the work on power measurements that EEMBC and ECL have done so far.

Q: Can the measurement equipment do the job?

A: The National Instruments 250-kHz DAC board and software tools proved suitable for the task at hand, although it may be prudent to move to a 1-MHz board (1 million samples per second). ECL was quickly able to build the National Instruments VI program and we can instrument the code to trigger the logging system.

Q: Is the equipment cost effective?

A: This same equipment doubles as the data acquisition and signal generation equipment used for the Automotive/Industrial Real Time Version 2 development, benchmarking, and certification work. The cost of a DAC board plus the National Instruments software is around $4000, far less than the hardware-only system, and it can be used for future EEMBC work.

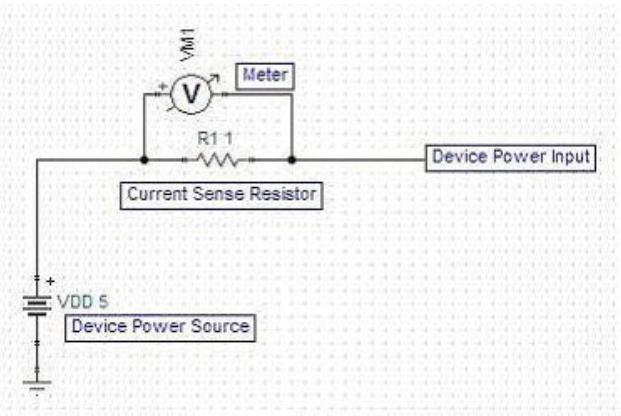
Q: Can we measure power on a variety of platforms?

A: Yes. We think that any system where the power rails have power access points (points we can solder on a power resistor) will work. We also must be able to access a signal pin on the board for control purposes.

Q: Can we measure more than one power rail?

A: Yes, so long as there are access points.

(continued on page 3)



FROM THE PRESIDENT

(continued from page 1)

quite revealing. Some phones have outstanding Java performance and will make game playing a joy. Other phones have such poor Java performance, it's stretching the facts to claim that they can run Java games.

In addition to the 'fun' and curiosity aspect for end-users, how will GrinderBench be used? This depends on the audience, which includes service providers, hardware OEMs (the mobile phone manufacturer), Java Virtual Machine (JVM) developers, and silicon providers. While "Vendor X" may build the phone, user complaints always come through the service provider. Hence, all service providers should make use of EEMBC's new service and try out every phone that they are evaluating and subsequently offering to the public. Vendors will be able to make use of this service to test their phones and subsequently improve them. The same is true for JVM developers and silicon providers.

In doing performance testing, these stakeholders may want to go beyond just running the benchmarks. They might also want to examine the actual source code, do some performance profiling, and determine how to improve their products. We invite companies such as these to become members of the EEMBC Java Subcommittee or to license the GrinderBench benchmark software.

With several different Java benchmarks now on the loose, GrinderBench stands as a thoroughly engineered, tested, and certifiable standard for measuring the performance of Java implementations. We've changed our rules so everyone with a Java-enabled mobile phone can experience GrinderBench while ensuring that only valid and reliable scores on these phones will be published. Compared with the alternatives, GrinderBench may be somewhat less exciting to run; it won't bring up any fancy images on your cell phone display. The rather plain interface allows the benchmarks to focus on the true performance aspects of the Java implementation. This is also the reason why GrinderBench doesn't provide a progress meter while the benchmark is running, as even that interferes with the actual benchmark execution. But under the hood, there's a lot of processing going on, making this a very powerful benchmark and laying a solid foundation for future versions of these industry-standard benchmarks.

The new GrinderBench Web site at www.grinderbench.com will debut in late August 2005.

Markus Levy

Power Consumption: The Next Great Benchmarking Frontier

(Continued from Page 2)

Q: Can we make measurements over a wide range of frequencies?

A: Yes. For experimental purposes, we measured a 180-MHz processor, a 266-MHz processor, and a 600-MHz processor. We can take at least 250,000 samples per second, and if we run the benchmarks over at least a two- to three-second range, that will generate a lot of data samples. If necessary, a faster DAC board can be inexpensively acquired.

Q: Can we measure performance and power simultaneously?

A: Yes, although separate power and performance runs would be better to remove any potential for intrusive behavior.

Q: Can we find a way to get consistent results?

A: Consistent to within 5% is achievable and practical. The power resistor itself contributes about 1% variance, and the DAC board another 1%.

Q: What is the difference across measuring tests, and are there differences between benchmarks in terms of power?

A: We saw consistent differences between benchmarks and different power characteristics (waveforms, graphs). We think that benchmarking across different benchmarks is meaningful based on our experiments.

Q: When will this process be submitted for approval to the EEMBC Board of Directors?

A: A few more experiments must be performed, and the entire process from start to finish must be demonstrated on a suitable number of platforms, but we believe that there are no technical obstacles to completing the process development.

NEWS BRIEFS

Broadcom Corporation and **SunPlus Technology** are the newest members of the EEMBC Board of Directors. Broadcom supplies wired and wireless broadband communications semiconductors and had revenues of $2.4 billion in 2004. Customers for its system-on-chip and software products include manufacturers of computing and networking equipment, digital entertainment and broadband access products, and mobile devices. Taiwan-based SunPlus provides custom and standard-product IC solutions for handheld electronics, interactive toys, digital cameras, MP3 and DVD players, and many other consumer electronic systems. The company's products include LCD controllers/drivers, microcontrollers, multimedia ICs, and memory chips.

Interest in EEMBC benchmarks on the part of academic researchers is on the rise in Europe. Since the last issue of EEMBC Journal, faculty at the **University of Hertfordshire** (UK), **Chalmers University of Technology** (Goteborg, Sweden), and the Institut für Datentechnik und Kommunikationsnetze (IDA) of the **Technical University at Braunschweig** (Germany) have become EEMBC U academic members.

Most datasheets for the **DENBench 1.0 Digital Entertainment benchmark suite** are now available from the Digital Entertainment section of the EEMBC Web site.

Benchmarking Multicore Platforms: Just how can I know? *(Continued from Page 1)*

Generalized competence: this is where I'd put the EEMBC suites. These benchmarks take a cross section of realistic workloads and provide standardized implementations which are applied to solutions that provide a generalized processing capability.

Consumer satisfaction: these forms of benchmarks are designed to provide user-based scenarios and will typically only execute on the final, complete solution.

Even in offering these distinctions, I hope it's clear that the boundary between each can be somewhat blurred. Technology is constantly evolving, and what may once have been very much a specific signal processing algorithm can now be performed by a general-purpose processor. Likewise, a generalized task, by virtue of rigidity through standardization, becomes a candidate for fixed-function signal processing. As far as benchmarks are concerned, end users really only care about how well their purchased solution will operate. It is for this reason that I separate the consumer satisfaction benchmarks into their own category. These tests target end user applications on a defined type of platform — and this is where another key use of benchmarking comes into conflict.

Using benchmarking to measure the consequences of design is fine, and can be accomplished by all three forms of benchmarks. The challenge comes when designers use the benchmarks to make design choices and architecturally explore the options they have within a given design space. In this case, the distinction between these three forms of benchmark must be clear and concise.

For the targeted signal processing benchmarks, a time consuming, very expensive, unique, and complex redesign of the benchmark implementation for each design variation must be undertaken. In the case of consumer satisfaction benchmarks, the sheer size of the benchmarks makes it impossible to execute them in any of today's design exploration tools, without the designer needing to either invest excessive amounts of simulation time, or to simplify the benchmarks to a point they can be executed. This brings us back to generalized competence benchmarks and EEMBC.

The EEMBC benchmarks for many years have been offering designers both the ability to compare final solutions and to help improve their products. Delivered as portable C-code, a simple recompile allows these tests to run on some new or improved general-purpose processing unit. But the world is changing. The general-purpose compute engines (CPUs) targeted by EEMBC are starting to call on techniques long used in embedded systems, for example by using multiple general purpose compute engines, i.e. a multiprocessor implementation, instead of a single uniprocessor. The distinction between multiple processors at the system level and multiple processors at the CPU level is an important one when you start to look at what must be done within benchmarking. Generalized benchmarks that historically focused on a CPU's performance now must also understand the multiprocessor (MP) and multi-thread (MT) forms of processor implementations. But the world of system-level multiprocessing is quite diverse. It is fair to ask whether this is something best left to the consumer satisfaction level of benchmarks.

When looking into processors that provide software with a view of multiple CPUs, it's clear that the traditional programmer's view of writing for a single flow of instructions cannot be used across processors capable of executing multiple flows of instructions concurrently. Any benchmark code must change, but to offer comparisons with existing uniprocessor designs, the benchmark must be portable between the two forms of processor.

Symmetric multiprocessing (SMP) is a software technique where code is written so that one or more threads (flows of instructions) can be represented in a portable form. The task of multi-threaded programming is limited. Its role is really only to define the syntax and the API of how to represent code for a multiprocessor. It does not define the semantics of how a workload is to be partitioned to accomplish the given work.

When looking at SMP, there are three main forms for semantic decomposition of workloads. Together these can be used to provide the required meaning in the development of MP benchmarks.

Decomposition of a single task: in this scenario, a single algorithm is parallelized to execute concurrently across the multiple processors. Similarities exist here in the targeting of a signal processing algorithm to a specific (MP) architecture. The key difference is that the benchmark is focused at general purpose processors in which the designer will always maintain a C-code target. You could consider this as simply a code optimization of an existing EEMBC kernel.

Execution of multiple algorithms: this form of benchmark takes multiples of the algorithms that represent a user scenario and runs them together in much the same way the end user would. It thus starts to encroach on consumer satisfaction tests. The key understanding gained from this type of test is how well a given solution can manage multiple concurrent activities.

Multiple streams of data: in embedded computing, performance is not just a matter of how well a solution can solve a single task or multiple tasks, but of how well it can process a given input data — its throughput. Here the multi-processor is using its multiple threads to split up and share the data across the processors. To measure how well this is done, a different kind of benchmark is needed.

In short, there is no simple answer to the question "Just how can I know?" Thankfully, though, the available benchmarks are broad and cover a wide design space. But we must keep in mind that as technology changes, so will the ways in which devices address a problem and offer a solution. In measuring how well they do this with benchmarks, the important thing is to ensure that these instruments of measurement remain appropriate, and evolve in parallel with the greater changes that are taking place in processor architectures, implementations, and applications.

EEMBC Calendar

ARM Developers' Conference
October 5 – Santa Clara Convention Center
Markus Levy will present *"Update on Industry-Standard Performance Analysis"*

October 6 – Santa Clara Convention Center
Markus Levy and Graham Wilkinson will present *"EEMBC Java Development and Testing with GrinderBench"*

GSPx 2005 Pervasive Signal Processing Conference
October 24-27 – Santa Clara Convention Center
Markus Levy and Danny Wilson of LSI Logic will co-present *"A Layered Approach to Evaluating Processor and System Performance for VoIP Applications"*

Major recent and forthcoming contributed articles by Markus Levy
"Evaluating Digital Entertainment System Performance" (**IEEE Computer**, July 2005)

"New Benchmarks Aim to Verify Microcontroller Real-Time Capabilities" (**Embedded System Engineering**, May 2005)

"Performance-Tests für Bildverarbeitungs-Systeme" (forthcoming Q4 in **Elektronik Praxis**)