EEMBC and the Purposes of Embedded Processor Benchmarking
Markus Levy, President

ISPASS 2005

Certified Performance Analysis for Embedded Systems Designers

EEMBC: A Historical Perspective

-Began as an EDN Magazine project in April 1997
  -Replace Dhrystone
  -Have meaningful measure for explaining processor behavior
-Developed business model
-Invited worldwide processor vendors
-A consortium was born
EEMBC Membership

- Board Member
  - Membership Dues: $30,000 (1st year); $16,000 (subsequent years)
  - Access and Participation on ALL Subcommittees
  - Full Voting Rights
- Subcommittee Member
  - Membership Dues Are Subcommittee Specific
  - Access to Specific Benchmarks
  - Technical Involvement Within Subcommittee
  - Help Determine Next Generation Benchmarks
- Special Academic Membership

EEMBC Philosophy: Standardized Benchmarks and Certified Scores

- Member derived benchmarks
  - Determine the standard, the process, and the benchmarks
  - Open to industry feedback
  - Ensures all processor/compiler vendors are running the same tests
- Certification process ensures credibility
  - All benchmark scores officially validated before publication
  - The entire benchmark environment must be disclosed
Embedded Industry: Represented by Very Diverse Applications

- Networking
  - Storage, low- and high-end routers, switches
- Consumer
  - Games, set top boxes, car navigation, smartcards
- Wireless
  - Cellular, routers
- Office Automation
  - Printers, copiers, imaging
- Automotive
  - Engine control, Telematics

Traditional Division of Embedded Applications

<table>
<thead>
<tr>
<th>Low Power</th>
<th>High Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Cellular, routers
- Engine control, Telematics
Embedded Applications: Handled by Very Diverse Processors…

- Altera Corp
- AMCC
- AMD
- Analog Devices, Inc
- ARC International
- ARM
- Atmel Corporation
- CEVA
- Freescale Semiconductor
- Fujitsu Microelectronics
- IBM
- Imagination Technologies
- Improv Systems
- Infineon Technologies
- Intel
- IPFlex Inc.
- LSI Logic
- Marvell Semiconductor
- Matsushita Electric Industrial
- MIPS Technologies
- National Semiconductor
- NEC Electronics
- Oki Electric Industry Co., Ltd
- Patriot Scientific Corp.
- Philips Semiconductors
- PMC-Sierra
- Qualcomm
- Raza Microelectronics
- Renesas Technology Corp.
- Sandbridge Technologies
- Sony Computer Entertainment
- ST Microelectronics
- Stretch, Inc
- Sun Microsystems
- Tensilica
- Texas Instruments
- Toshiba
- Transmeta Corporation
- VIA Technologies

...And These Are Just From The List Of EEMBC Members

Evaluating Embedded Processors and Compilers

- Software compatibility and tool availability for CPU architecture
- Quality of tools
- Quality of service
- Level of integration
- Future availability and roadmap

Qualitative Comparisons
Quantitative Comparisons

- Feature set, peripherals
- Performance benchmarking
  - Native operations
  - Dhrystone mips
  - EEMBC
- Power consumption
- Price/Cost

Clients for Embedded Processor Benchmarking

- Framework to guide architectural choices for development stage of processors, compilers, etc.
- Researchers for experimenting and creating advanced technologies
- Platform OEMs
  - Determine performance bottlenecks
  - Understand how to improve end user performance
- Service Providers (e.g., Vodafone, DoCoMo)
  - Choosing the best platform to offer subscribers
- Content Providers (e.g., HI, Sega)
  - Determine state of the art, min/max performance
- End Users
  - Need help determining what to buy
  - Experience running real-world content (playability, response to user input)
Differentiating Between PC and Embedded Application Benchmarking

- What is enough performance?
  - For PCs, more is better
  - In embedded, enough to get the job done

The fastest isn’t always the ‘bestest’

Challenges of Creating Embedded Processor Benchmarks

- Reiterate: Embedded market is very diverse
- Best benchmark = ‘customer’s’ application
- The processor is often an SoC
- Focus of benchmarks
  - Processor/compiler
  - System level (memory, I/O)
- Synthetic versus real-world
Organizational Structure

President
Markus Levy

HW Power WG
Shay Gal-On, PMC

SW Power WG
Moshe Sheier, CEVA

Multiprocessing WG
John Goodacre (ARM)

Automotive WG
Manfred Choutka, Infineon

Consumer WG
Sergei Larin, Freescale

Java WG
Graham Wilkinson, ARM
Terrence Barr, Sun

Networking WG
Bill Bryant, Sun

Office Automation WG
Ron Olson, IBM

VoIP WG
Dan Wilson, ZSP/LSI

Telecomm WG

Accomplishments Since 1997

- First version of benchmarks include:
  - Automotive, Consumer, Networking, Office Automation, and Telecomm
  - Available since 1999
  - First Java CLDC benchmarks in 2002
  - Networking V2 in 2004
  - Digital Entertainment in 2005

More Than 1.5 Million Lines of Code Available
Consumer Subcommittee

- ConsumerMark™ = Version 1 consumer benchmarks
  - Benchmarks target digital cameras, basic imaging
    - JPEG Encode/Decode, Color Conversion, Filtering
  - Recently completed “Digital Entertainment” benchmarks called DENbench™
    - MPEG 2/4 Encode, MPEG 2/4 Decode, MP3, Encryption algorithms
    - Aggregate mark = DENmark™

Networking Subcommittee

- First generation networking benchmarks
  - PacketFlow, OSPF, Route Lookup
- Recently completed two suites of 2nd generation benchmarks
  - Internet protocol
    - NAT, Packet Reassembly, QoS, plus significantly enhanced first generation benchmarks
    - Aggregate score = IPmark™
- TCP/IP
  - Benchmark sends/receives its own packets to avoid I/O overhead
  - Aggregate score = TCPmark™
Office Automation Subcommittee

- First generation benchmarks test basic printer functions
  - Dithering, Image Rotation, Text Processing
  - Aggregate score = OAmark™
- Currently developing second generation benchmarks
  - Includes embedded version of Ghostscript

Automotive Subcommittee

- First generation benchmarks test a variety of workloads
  - Engine control, in-car entertainment, ABS
  - Aggregate score = EEMBC AutoMark™
- Currently developing 2nd generation benchmarks
  - Hardware-based for real-time analysis
  - Testing peripherals, interrupt structure, etc.
Telecomm Subcommittee

- First generation benchmarks test DSP properties
  - Autocorrelation, Bit Allocation, Convolutional Encoder, FFT, Viterbi
  - Aggregate score = EEMBC TeleMark™
- Currently defining next generation benchmarks
  - Initial focus on VoIP

Embedded Java Benchmark Suite

- Designed to analyze the entire Java platform, not just the Java execution engine
- Based on real application code:
  - Internet-usage benchmark stresses CLDC threading
  - Gaming benchmark stresses computations
  - Photo benchmark stresses photo decoding using PNG format
  - M-commerce benchmark stresses cryptography algorithm decoding
Cold Start Versus Warm Start

- First time versus steady state
- Warm start relies on incremental compilation or JIT compilation
- Cold start benefits hardware accelerators

Methodology of Creating Embedded Processor Benchmarks

- Benchmarks derived from multiple sources
  - Challenging to develop code from scratch, but we do
  - Benchmarks can be developed starting with industry standard reference code
  - Benchmarks donated from members/industry
- EEMBC defines and develops data sets
- ECL integrates the Test Harness to ensure a common benchmark API
  - Makes it ‘easy’ to get most platforms running
  - Support for hardware and simulated platforms
EEMBC Test Harness

- Target Hardware
  - System Memory
    - Digital Entertainment Benchmark Code
    - ‘Media Stream’
    - Test Harness Control Code
  - Memory Controller
  - Processor
    - 1 cache
    - 0 cache
  - Timers

EEMBC Scoring Methods

- “Out-of-the-box” scores:
  - Standard EEMBC source code
  - Any publicly available compiler
  - Any compiler optimizations
  - Must report compiler and switches
- “Full-fury” scores:
  - Can rewrite EEMBC source code
  - Can fine-tune in assembly language
  - Can use special function libraries
  - Can use special hardware in CPU
Out-of-the-Box Implications

- First application-oriented compiler benchmark
- Tests architecture’s ‘C friendliness’
- Can serve as a C compiler benchmark suite

Applying EEMBC for Industry

- Published and non-published scores
- Vendors use this information for competing at various levels
### Total Score

<table>
<thead>
<tr>
<th>Processor</th>
<th>Score1</th>
<th>Score2</th>
<th>Score3</th>
<th>Score4</th>
<th>Score5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SuperH SH5-103</td>
<td>1.23</td>
<td>1.29</td>
<td>0.96</td>
<td>1.09</td>
<td>0.98</td>
</tr>
<tr>
<td>SuperH SH4-202</td>
<td>1.11</td>
<td>0.98</td>
<td>1.13</td>
<td>0.99</td>
<td>1.08</td>
</tr>
<tr>
<td>MIPS 20Kc</td>
<td>1.45</td>
<td>1.07</td>
<td>1.06</td>
<td>1.17</td>
<td>0.81</td>
</tr>
<tr>
<td>Infineon TriCore/TCIM</td>
<td>0.72</td>
<td>0.94</td>
<td>0.97</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>IBM ppc405d4v6</td>
<td>0.89</td>
<td>1.01</td>
<td>0.89</td>
<td>1.05</td>
<td>1.04</td>
</tr>
<tr>
<td>ARM1026EJ-S</td>
<td>0.98</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>ARM1020E</td>
<td>0.77</td>
<td>0.36</td>
<td>0.98</td>
<td>0.83</td>
<td>0.87</td>
</tr>
</tbody>
</table>

(Scores normalized by the OTB average of each benchmark suite)

### Used to Demonstrate System-Behavior

- Example based on 64-bit MIPS processor running Networking Packet Flow benchmark
- SRAM 20% higher performance than DRAM
  - But delta drops to 10% at 150 MHz
  - All from 4 extra clocks on leadoff cycle
Write-Through vs. Write-Back

- 100 MHz implementation is using cache write-through
- 150 MHz implementation is using cache write-back
- Numbers are in iterations/sec

Case Study: Processor Comparison

- Chart highlights performance-related features
- Demonstrates that benchmarks test more than processor core
Components of Consolidated Scores for Digital Entertainment

64 Tests

- **DENmark™** (Geomean of entire suite * 10)
- **MPEG 2/4 Encode/Decode**
- **MP3 Player**
- **Crypto Benchmarks**
- **Static Image Benchmarks**

- **MP3 Decodemark™**
  - Apply geomean * 1000

- **Crypto Benchmarks**
  - AES
  - DES
  - RSA
  - Huffman Decode

- **Imagemark™**
  - RGB->YIQ
  - RGB->HPG
  - RGB->CMYK
  - JPEG Compression
  - JPEG Decompression

The Comprehensive DENmark™

- **DENmark combines all scores in suite**
- **Makes it easy to compare processors**
- **Loved by ‘marketing’**
- **Minimizes engineering value**
MPEG Decode and Encode

MPEG-2 Encode
(5 data sets)
MPEG-4 Encode
(5 data sets)
MP3 Player
(5 data sets)
MPEG-2 Decode
(5 data sets)
MPEG-4 Decode
(5 data sets)

Elements of Cryptography

AES
DES
RSA
HuffmanDecode

CryptomarkTM
(Apply geomean * 10)
Processing Static Images

- RGB->YIQ (7 data sets)
- RGB->HPQ (7 data sets)
- RGB->CMYK (7 data sets)
- JPEG Compression (7 data sets)
- JPEG Decompression (7 data sets)

Imagemark™
(Apply geomean * 10)

Overall Performance Comparison

- Bigger caches and higher clock speeds = fastest raw performance

Scores Normalized to AMD Processor

- DENmark™
- Imagemark™
- Cryptomark™
- MPEG
- Encode mark™
- MP EG
- Decode mark™

IBM 750GX - 1GHz
Freescale MPC7447A - 1.4GHz
AMD Geode NX1500@4W - 1GHz
ADSP-BF533 - 594MHz
Architectural Efficiency Comparison

- Performance/MHz yields different results
- But, performance is not a linear relationship

Energy Efficiency Comparison

- Performance/Watt yields striking differences
- Critical analysis for embedded applications
Applying EEMBC For Research

- Ecole Polytechnique Federale De Lausanne
- Kent University
- Korea University
- MIT
- Northeastern University
- Northwestern University
- Tokyo Institute of Technology
- University of California, Berkeley
- University of Bristol
- University of Delaware
- University of Illinois
- University of North Texas

Research Usage Model

- Benchmarks applied to a variety of functions
  - Processor/architecture development/experimentation
  - Compiler testing
  - Building and testing simulation models
- EEMBC makes most of the EEMBC benchmarks available for academic research
- Publication of scores
  - Relative scores can be shown for commercial products (otherwise requires certification)
  - Can show absolute scores for research projects
- Source code remains confidential (similar to SPEC)
EEMBC’s Next Generation Benchmarks Are Hot

Active Working Groups: Power

- Standardizing on power/energy measurement
  - Measures the energy consumed while running benchmarks
  - Implementations for hardware and simulator-based platforms
Hardware Versus Simulator Power Measurements

**Hardware: Easy to run, hard to measure**
- Benchmarks run at processor speed
- Where to attach measuring device?
- What system components to include?
- When to measure?

**Software: Hard to run, easy to measure**
- Benchmarks run with gate level netlist
- Capture any data anywhere in program

Challenges of Hardware-Based Power Measurements

- What components to include in the measurement?
  - Performance/energy dependent on core/system
  - Memory hierarchy
- How will measurements be performed?
  - Simple meter or oscilloscope
- Can we use existing benchmark suite?
  - Required for consistency
- Do the current benchmarks capture data value sensitivities?
- Sampling rate
  - Frequency of sampling?
  - Sample with multiple frequencies to avoid looking at same benchmark point
Active Working Groups: Multiprocessing

- Multiprocessing: three parallelism characterizations:
  - Task decomposition: Takes a single algorithm and parallelize it to share its workload over multiple processors
  - Multiple Algorithms: Examines how the bigger system, including the OS, handles the workloads from multiple concurrent algorithms.
  - Multiple Streams: Examines the bigger system, but concentrates more on the data throughput, and how a system can handle multiple 'channels' of data.

Academia Using EEMBC To Help Support Research Projects

- VIRAM1: A Media-Oriented Vector Processor with Embedded DRAM
  - Joseph Gebis, Sam Williams, and David Patterson, Computer Science Division, University of California, Berkeley; and Christos Kozyrakis, Electrical Engineering Department, Stanford University.

- Vector vs. Superscalar and VLIW Architectures for Embedded Multimedia Benchmarks
  - Christoforos Kozyrakis (Stanford University) and David Patterson (University of California at Berkeley)

- A Standalone GCC-based Machine-Dependent Speed Optimizer for Embedded Applications
  - Sylvain Aguirre (University of Applied Science, Yverdon, Switzerland), Vaneet Aggarwal (Indian Institute of Technology, Kanpur, India), Daniel Mlynek (Swiss Federal Institute of Technology, Lausanne, Switzerland)
Time for a Pop Quiz

- **Question:** How can we work with Academia to establish the EEMBC standard?
  - **Answer:** Involve more universities and non-profit research institutions

---

Time for a Pop Quiz

- **Question:** Does EEMBC support co-development projects.
  - **Answer:** Co-development projects are proving to be quite useful. One example relates to EEMBC’s hardware power measurement standard being co-researched with Northeastern University’s Computer Architecture Research Lab
Time for a Pop Quiz

• Question: Is EEMBC accepting code donations to include into its benchmark standard?

• Answer: Absolutely. Currently working on an awards program.

Time for a Pop Quiz

• Question: How does EEMBC deal with the increasing challenge of developing benchmarks to test platforms of growing complexity?

• Answer: EEMBC is interested in establishing sponsored code development programs.
EEMBC Wants Your Support!

• Questions?
• Time to participate
• Contact
  • Markus Levy; Markus@eembc.org

www.eembc.org