Trends in Design and Implementation of Signal Processing Systems (DiSPS)

Moderator: Wonyong Sung, TC Chair of DiSPS

Presenter: Liang-Gee Chen & Mohammad M. Mansour

Graduate Institute of Electronics Engineering and Department of Electrical Engineering
National Taiwan University, Taipei, Taiwan
Acknowledgement of materials preparation

• Mohammad M. Mansour
• Andy Wu
• William Plishker
• Shuvra S. Bhattacharyya
• Shao-Yi Chien
Challenge from Plenary talk: Henry Tirri

Title: Making Sense of a Zettabyte World
Questions Raised

1. How do we capture, parse, and analyze, on the fly, zettabytes of data from radically different sources?

2. What will it take to build architectures that are truly robust across multiple facets and that have strong elastic properties?

3. How do we tackle energy efficient computing while balancing environmental concerns, business practicalities and the end user experience?

All challenges are related to DiSPS!!
Position of Design and Implementation in Signal Processing Systems

Applications: email, internet, ftp, p2p, video, ...

DiSPS

Computing Platform: SOC, Multi-core, MPU, FPGA, ASIP, ASIC
Position of Design and Implementation in Signal Processing Systems

Applications: email, internet, ftp, p2p, video, ...

Computing Platform: SOC, Multi-core, MPU, FPGA, ASIP, ASIC

DiSPS

Users
Resource Hungry Applications

The diagram shows the relationship between Mbps required and market progression for various applications. The x-axis represents market progression, while the y-axis represents Mbps required. Applications are categorized into Alternative, Substitutive, Preemptive, and Home of the Future. Examples include Email, Static Web pages, Integrated messaging, Peer-to-Peer Networking, Video Streaming, Video Telephony, Voice/Data/Video Telecommuting, Interactive TV, Broadcast TV, Interactive Shopping, Telepresence, TV download, Tele-education, Home of the Future.
Ultra High Definition TV

Multi-User/Multi-View 3DTV

Free-View-Point TV
Visual Recognition Technology

• New Vision to Life Revolution
  – Wearable vision, robot, surveillance, gaming, surveillance...

• Emerging technology trend:
  – High invariant ability to scale, view-angle, luminance change
  – High resolution video processing for accurate recognition
  – Low power design for energy saving
  – Intelligent human computer interface embedded inside for natural interaction

Seungjin Lee, 2010
Kinect, Microsoft
SixthSense, MIT media lab
ASIMO, Honda
Position of Design and Implementation in Signal Processing Systems

Applications: email, internet, ftp, p2p, video, …

DiSPS

Computing Platform: SOC, Multi-core, MPU, FPGA, ASIP, ASIC

Technology
Moore’s law still drives technology!!

(Data shown corresponds to the first version of each processor family)

Transistor Count

(Data shown corresponds to the first version of each processor family)

Source: http://www.intel.com/pressroom/kits/quickreffam.htm
Technology carries more bits!!
Broadcasting more wide area!!

Position of Design and Implementation in Signal Processing Systems

Applications: email, internet, ftp, p2p, video, ...

DiSPS

Computing Platform: SOC, Multi-core, MPU, FPGA, ASIP, ASIC
Various Types of Architectures

- Multi-core processor
- GPU
- Stream processor
- Reconfigurable processor
  - Fine-grained: FPGA
  - Coarse-grained
- Application-specific instruction-set processor (ASIP)
- ASIC
- ...
- Hybrid architecture
Trends of Design and Implementation in Signal Processing Systems

Applications: email, internet, ftp, p2p, video, ...

DiSPS should reflect the push & pull forces from both sides.

Computing Platform: SOC, Multi-core, MPU, FPGA, ASIP, ASIC
Demands from different perspectives

- Low level architecture/circuit techniques
  - Low voltage design
  - 3D IC
- Flexible Architecture
  - Easy of design
  - Real time application or instant request
  - More computing capability
- Hardware/Software co-optimization
  - More functions
  - Less energy
- System for Intelligent Signal Processing
  - Innovative architecture
I. Reduction of *Standby Memory Power* in Embedded Systems

Mohammad M. Mansour, American University of Beirut, Beirut, Lebanon
mmansour@aub.edu.lb
SRAM Standby Power: Technology Challenges

1. SRAM occupy large area on-chip and contribute to *power problem in standby* mode for embedded systems
   - Must scale-down supply voltage to reduce power

2. SRAMs are subject to *threshold voltage variations*
   - Random and temperature-dependent
   - Cannot be avoided using current technology, even getting worse with every new technology node
   - Prevents down-scaling of supply voltage at standby

3. SRAM design must guarantee stability measures
   - Read, Write measures at run-time
   - Retention measures at standby

• **Idea**: Reduce SRAM *supply voltage* to *data retention voltage* (DRV) during standby while guaranteeing stability measures
Dynamic DRV Techniques

• Traditional techniques for determining DRV
  – Analytical techniques
  – Statistical techniques
Both fail to capture random variations and do not yield optimum DRV

• New technique: Provide circuit support on-chip to measure and calibrate DRV
  – Captures threshold voltage and temperature variations
  – Easily integrated into BIST units
  – Yields optimum DRV

• Operation:
  – Test SRAM integrity at $V_{test}$ (starting from $V_{DD}$).
    • If test passes, lower $V_{test}$ one step and retest SRAM integrity.
  – If test fails, raise $V_{test}$ two steps and check integrity
  – If test passes, save $V_{test}$ as the DRV
On-Chip DRV Circuit

- TEST CONTROL UNIT
- Delay Counter
- SRAM BIST
- ADC
- SRAM ARRAY
- Comparator
- DC-DC Converter
- VDD
- VBat
- DRV

Connections:
- TEST CONTROL UNIT to Delay Counter
- Delay Counter to SRAM BIST
- SRAM BIST to SRAM ARRAY
- SRAM ARRAY to VDD
- VDD to VBat
- VBat to Comparator
- Comparator to DC-DC Converter
- DC-DC Converter to ADC
2. New Semiconductor Technology: 3D package or Heterogeneous Integration

By: Andy Wu
2D vs. 3D IC

2D SoC
- Long Global wire

Shorter wire

3D IC

http://www.jonathassociates.com/AJA/droppedImage.png
Major Advantages of 3D IC Technology

- **Improve Performance and Break Bandwidth Limitation by TSV**
  - UCSB report: the micro-processor stacking with L2 cache and DRAM architecture are 2X faster than conventional 2D architecture [Loi 06].
  - Intel shows that 3D processor can improve performance by 15%. [Black 04]
  - With the extra dimension in 3D NoC, the traffic delay can be greatly reduced. [Kim 07]

- **Reduce Power Consumption**
  - UNC research results: 3D IC can shorten the interconnection by 15% and reduce the power consumption by 10% [Al–Sarawi 98].

- **Enhance value and reduce form factor by Heterogeneous Integration**
  - With vertical stacking, we can increase transistor density per area footprint. It is beneficial for memory and FPGA ICs, which focus the capacity.
  - Heterogeneous integration can be achieved with 3D IC. For example, Logic + Memory, Sensor + Circuit.
3D Integrated Circuits

• More than Moore!

• What is brought to the table?
  - One more degree of freedom
  - Simplified Integration
  - Wire length
  - Via Density

• Opportunities from architects’ perspective
  - Latency (e.g., shrunk global wires)
  - Power consumption
  - Bandwidth
  - Flexibility
  - Heterogeneity

Source: Prof. Hsien-Hsin S. Lee, GIT, USA, talk on “3D Integration”
CPU/GPU/Embedded RAM Integration Architecture in 3D ICs

- **Integrated CPU/GPU architecture** is the future trend for many-core architecture
  - ARM Mali Platform, AMD Fusion and Intel Sandy Bridge
  - Advantage:
    - Fast communication between CPU and GPU
  - Issue
    - System performance will be limited by memory bandwidth

- **3D stacked memory** provides massive memory bandwidth required for integrated CPU/GPU architecture
3. High Level Architecture Exploration and System optimization

By: Shao-Yi Chien
Average NRE Cost and Architecture Selection

- Multi-core Processor
- ASIP
- GPU
- Stream Processor
- FPGA
- Coarse-grained Reconfigurable Processor
- ASIC

- Longer Time-to-Market (for System Design)
- More Advanced Technology
- Increasing Average NRE Cost (Shipped Amount)

Optimal Solution?
SoC Example (MPU + DSP + ESW)
Embedded Software Platform

### Applications
- P2P Stream Server
- Multimedia Portal
- Web Container
- Web Service
- Java AP, Java 3D
- VOIP, IM
- Web Browser
- Mail
- GPS Nav.
- DRM
- Image Viewer
- Security SW
- Media Player
- Office, PDF
- Graphics
- AV CODECs
- Baseband
- Transcoding
- Biometrics
- Crypto Engine

### Middleware
- JVM, OSGi
- DLNA
- GUI, 3D, JSR-184/239
- Network protocol, SIP, RTP
- DirectFB, OpenGL ES
- DSP Middleware
- GSM, 3G, WiMAX
- APIs

### System Software
- Embedded Linux + RTOS
- Multi-Core IPC
- Power Mgmt
- Multi-Core Micro Kernel
- Multi-Core Compiler Toolkits
- Multi-Core IDE and Debugger

### Multi-core SoC Platform
- Digital Camera
- Wireless
- Multimedia Accelerator
- Multi-Core, MPU + VLIW DSP
- Multi-Core interconnection
- DMA / Memory Control
- ESL: Virtual Platform Design
- Ethernet, 1394
- USB, HDMI
We will need ...

• Hardware/Software Co-optimization to provide
• More architecture exploration
• More Parallel computing
• Less power consumption

• ...

• Open Problem ??
4. Novel Design Methodology with Heterogeneous Parallelism

William Plishker and Shuvra S. Bhattacharyya
Department of Electrical and Computer Engineering, and Institute for Advanced Computer Studies
University of Maryland, College Park
{plishker, ssb}@umd.edu

Document version: May 9, 2011
Model-Based Design for Embedded Systems

• High level application subsystems are specified in terms of components that interact through formal models of computation
  – C can be used to specify intra-component behavior
  – Object-oriented techniques can be used to maintain libraries of components

• Popular models for embedded systems
  – Dataflow and KPNs (Kahn process networks) → widely used for signal processing system design
  – Synchronous languages
  – Continuous time
  – Discrete event
  – FSM and related control formalisms

• Two important motivations are to expose concurrency and exploit parallelism
Parallelism in DSP Applications

- Homogeneously parallel applications (e.g., basic filtering, matrix operations, IPv4 packet routing) are well-serviced by homogeneous architectures (e.g., traditional FPGAs and GPUs).

- Modern and emerging signal processing applications typically have significant components that are not homogeneously parallel.

- However, these applications must also be optimized for key metrics, such as latency, throughput and power consumption.

- Such applications have parallelism, but it exists
  - At different levels of granularity
  - With different forms of dataflow structure (not just conventional static dataflow)
Example Applications and their Performance-critical Subsystems

- **High energy physics**: Triggering system for the Large Hadron Collider (LHC): filtering, calculations, and ID
- **Software defined radio**: Protocols within GNU Radio -- a mix of digital signal processing blocks
- **Network processing**: higher level (TCP, http) routing decisions --- partial packet stream reassembly
- **Medical imaging**: automatic, image registration --- info. theory kernels in an image processing pipeline.

→ common characteristic: evolution from regular kernels to more complex, performance-critical subsystems
Key Trend: Heterogeneous Parallelism

- Design methods and tools must evolve to support design, analysis, optimization, and integration across different granularities and different dataflow structures (e.g., static, multirate, parameterized, dynamic, multi-dimensional, ...) for parallelism
Example: Intensity-Based Medical Image Registration

- Doctors want multiple types of data to be accurately overlaid
- Align the features of one image with the features of another
  - Rigid or non-rigid, mono-modality or multi-modality
Image Registration Solutions

• Four Basic Steps to Registering Images
  – Feature Detection
  – Feature Matching
  – Transform Model Estimation
  – Image Resampling and Transformation

• Solutions will vary in their treatment of each of these steps

• For intensity-based image registration, the overall process is iterative
Iterative Image Registration

- Start with a candidate transform
  - 2 to 9 parameters for rigid registration
  - Many parameters for non-rigid
- Apply the transform to a floating image
  - Interpolate individual voxel transformations where necessary with B-splines, thin plate splines, etc.
- Interpolate to match reference image space
  - Nearest neighbor, partial volume, trilinear, etc.
- Calculate the similarity between this image and resulting image
  - Mean Squared Difference (MSD), Mutual Information (MI), etc.
- Use the similarity result to generate a new candidate transform

Diagram:
- Generate Transform
- Apply Transform
- Interpolate
- Calculate Similarity
- Floating Image
- Reference Image
Heterogeneous Parallelism in Image Registration

To Probe Further

Part 1: Applications

Part 2: Architectures

Part 3: Programming and Simulation Tools

Part 4: Design Methods
5. Low-power “DSP Software”
   – Energy-Tagged Instruction Set Architectures

Mohammad M. Mansour, American University of Beirut, Beirut, Lebanon
mmansour@aub.edu.lb
II- Energy-Tagged Instruction Set Architectures

- Notion of Low-power DSP software
II- Energy-Tagged Instruction Set Architectures

• Micro-architecture exposes knobs for compiler to turn on/off blocks depending on instruction type to save energy
• All instructions in ISA are tagged with energy numbers

\[ \text{inst } \langle \text{op1} \rangle \ < \langle \text{op2} \rangle \ < \langle \text{op3} \rangle \ < \langle \text{energy} \rangle \]

• An energy/power optimizing compiler
  – Produces a mix of instructions with lowest energy/power on target μ-architecture
  – Deactivates inactive datapath blocks to save power

• Challenges:
  – Accurate tagging of instructions with energy numbers
  – Support for multiple threads on parallel DSP platforms
6. Design and Implementation for New DSP Algorithms

- Compressive sensing
- Bio-medical signal processing
- Bio-inspired processing
- ...


Paradigm of Compressive Sensing (CS)

- Sample = Compress

\[ y = \Phi f \]

“cheap” encoder

“expensive” decoder

L1 minimization
Advantages from Compressive Sensing

• Compressive sensing
  – Takes advantage of the fact that many signals are sparse under some basis
  – Each measurement carries the same amount of information and use less measurement to recovery original signal

• Compressive sensing and sparse representation have been proved to be useful in many new applications
  – Image processing: denoise, super-resolution, inpainting, segmentation
  – Image analysis: classification, recognition, modeling
  – Other: Audio, Bio...
BIOMEDICAL SIGNAL PROCESSING SYSTEMS
Biomedical Applications

Advanced Signal Processing Algorithms

Smart Biomedical Sensor

Bulky Systems in Hospital

Miniaturized Signal Acquisition and Wireless Telemetry ICs
On-sensor Biomedical Processors

- Signal processing capability on the portable, wearable, or implantable biomedical sensors
  - Data reduction for low-power wireless transmission
  - Real-time processing for timely warning
  - On-line feedback of electrical stimuli or control signals
  - Analysis for multi-channel signals with complicated algorithms
Wearable/Implantable Smart Body Sensor

Power & Size Reduction By Chips

300 cm³
140 mW

< 1 cm³
< 100 µW

Extend Applications to:
• Sports
• Entertainment
• Assisted Living
Realtime Applications

Sudden Cardiac Arrest

Next-generation User Interface

Epilepsy

Spinal Cord Injury


http://www.emotiv.com/
Progressing in ECG Sensor IC

Compact SoC
- ASIC for compression and encryption (ASSCC 2008 [3])
- Low power sub-Vt uC (VLSI 2009 [4])

Wireless Tx and uC (ISSCC 2008 [2])
- Low power sensor interface (VLSI 2008 [1])

Smart Sensor
- Holter’s Device
- Readout and storage

SCA Detection
- Heart rate estimation
- Compression and encryption
- Wireless raw data receiving

Functionality
Heterogeneous Processors on a Platform

• Application specific processor (ASP) gain in power efficiency for main computation

• General purpose processor (GPP) achieve flexibility for relatively small computation
Biomedical SPSs

• Biomedical sensor equipped with signal processing capability is the trend
  – Low power, real-time decision, and on-line feedback for wearable or implantable applications

• System optimization for Signal Processing is the key to succeed
  – Biomedical platform, heterogeneous processors, digitally-controlled interface, pipelined processing, and parallel-folding structure
BIO-INSPIRED SIGNAL PROCESSING
New Paradigm: Bio-Inspired Computing

Computing

Understanding
Brain Inspire Video Intelligent

Intelligent Recognition

Bring the Intelligent to the Circuits

Occlusion Synthesis
Great potential to scale

- Neocortical computing hardware scales well with the growth of resources since more and more neocortices and thus intelligence can be emulated in one SOC
## Future Apps: Bio-Inspired System

### Comparison

<table>
<thead>
<tr>
<th>Component</th>
<th>Computer</th>
<th>Pigeon’s Brain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td>Speed</td>
<td>Fast (GHz)</td>
<td>Slow (120Hz)</td>
</tr>
<tr>
<td>Component</td>
<td>Digital Logic</td>
<td>Neuron</td>
</tr>
<tr>
<td>Operation</td>
<td>Sequential</td>
<td>Massive Parallel</td>
</tr>
<tr>
<td>Function</td>
<td>Computation</td>
<td>Learning Adapting</td>
</tr>
<tr>
<td>Component Error</td>
<td>Fatal</td>
<td>Tolerable</td>
</tr>
</tbody>
</table>
Trends of Design and Implementation in Signal Processing Systems

Necessity is the mother of invention!!
Trends of DiSPS could be:

- Low level architecture/circuit techniques
  - Low voltage design
  - 3D IC
- Flexible Architecture
  - Easy of design
  - Real time application or instant-on requirement
  - More computing capability
- Hardware/Software co-optimization
  - More functions
  - Less energy
- Novel and Intelligent Signal Processing
  - Innovative design and implementation system