A 65nm Dual-mode Baseband and Multimedia Application Processor SoC with Advanced Power and Memory Management

Tatsuya Kamei, Tetsuhiro Yamada, Takao Koike, Masayuki Ito, Takahiro Irita, Kenichi Nitta, Toshihiro Hattori, Shinichi Yoshioka

Renesas Technology Corp.
Outline

1. Introduction
2. Chip Architecture
3. Power Management
   - Static Power (Hierarchical Power Domain)
   - Dynamic Power (Partial Clock Activation)
4. Memory Management
   - IP-MMU
5. Summary
Introduction(1) Trend of Mobile phone

- In a 2G handset, a baseband processor executed applications in addition to modem control. (RTOS)
- In a high-end 3G handset, applications require High-Level-OS and rich multimedia functions. 3(BB, AP, and MM) processors in a handset.
- System cost increase → 1 chip integration
Introduction(2) Challenges

- **Low Power**
  - Limited battery capacity
  - No fan

- **High Performance**
  - Digital TV
  - Mega-pixel camera
  - 3D graphics for games and UI

- **System Cost**
  - Memory capacity (SDRAM, NAND/NOR FLASH)
  - Packaging (SIP, POP)
Introduction (3) History of SH-Mobile G

- **1st generation SH-Mobile G1**
  - In mass production from 2006. 8
  - 141M Transistors, 90nm LP
  - Hierarchical Power Domain for leakage reduction

- **2nd generation SH-Mobile G2**
  - In mass production from 2007. 8
  - 249M Transistors, 90nm LP
  - Inter-Connect Buffer for high-performance multimedia

- **3rd generation SH-Mobile G3**
  - In mass production from 2008. 8
  - 307M Transistors, 65nm LP
  - Partial Clock Activation for dynamic power reduction
  - IP-MMU for memory footprint reduction
Outline

1. Introduction
2. Chip Architecture
3. Power Management
   - Static Power (Hierarchical Power Domain)
   - Dynamic Power (Partial Clock Activation)
4. Memory Management
   - IP-MMU
5. Summary
# SH-Mobile G3 Chip Specifications

| CPU | ARM926EJ-S, 166MHz [baseband]  
|     | ARM1176JZF-S, 500MHz [application]  
|     | SH-X2, 500MHz [multimedia]  
| 3G/2G modem | W-CDMA/HSDPA Cat. 8 (up to 7.2 Mbps), GSM/GPRS/EDGE  
| Multimedia hardware engines | Video Processing Unit (VPU) (MPEG4/H.264/VC-1 D1 size), Sound Processing Unit (SPU), 3D/2D Graphics, GPS, LCDC, Camera interface (up to 12M pixels), 512KBytes MediaRAM (MERAM)  
| I/O | 617 pins  |
Die micrograph

<table>
<thead>
<tr>
<th>Die size</th>
<th>9.3mmx9.3mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>65nm LP</td>
</tr>
<tr>
<td></td>
<td>8M(7Cu+1Al)</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>1.2V(internal), 1.8/2.5/3.3V (I/O)</td>
</tr>
<tr>
<td># of TRs, gate, memory</td>
<td>307M TRs (28.2M Gates, 30.7Mb RAM, 6.4Mb ROM)</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Chip Architecture
3. Power Management
   - Static Power (Hierarchical Power Domain)
   - Dynamic Power (Partial Clock Activation)
4. Memory Management
   - IP-MMU
5. Summary
Power Management

Both Static and Dynamic Power are very important for mobile application!

- **Static Power Consumption (Leakage)**
  
  **Objective:** Standby time
  
  **Technique:** Hierarchical Power Domain (HPD)

- **Dynamic Power Consumption**
  
  **Objective:** Operational time
  
  **Technique:** Partial Clock Activation (PCA)
## Static Power Management

<table>
<thead>
<tr>
<th>Tr. Vth</th>
<th>Leakage</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HVth</td>
<td>Good (Low)</td>
<td>Poor (Low)</td>
</tr>
<tr>
<td>MVth</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>LVth</td>
<td>Poor (High)</td>
<td>Good (High)</td>
</tr>
</tbody>
</table>

- **Triple Vth technology**
- **Power Domain Separation (Partial Power-Off)**

<table>
<thead>
<tr>
<th># of domains</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few</td>
<td>Easy to control</td>
<td>Limited Leakage reduction</td>
</tr>
<tr>
<td>Many</td>
<td>Optimal Leakage reduction</td>
<td>Difficult to control</td>
</tr>
</tbody>
</table>

**Hierarchical Power Domain (HPD)**
Hierarchical Power Domain

Power area activity:
- System
- Real-time

Almost Power Off
- ARM1176 (logic)
- SYS Peripheral
- SH-X2 (logic)
- 3D Graphics
- LCDC, MediaRAM
- Sound related IPs
- LVDS serial interface
- GPS BB

Almost Power On
- ARM1176 (RAM)
- SH-X2 (RAM)
- RT Peripheral
- INTC, DDR. cntrl1
- INTC, RTSHwy
- DSP
- ARM926
- BB- bus & DMAC

Always On
- Common Power Domain
  - C-SHwy, Local Bus, DDR cntrl2
  - PLL
  - Clock Generators
  - Back-up FFs
  - Repeater cells

20 domains in total
Hierarchical Power Domain

Power area activity:

- System
- Real-time

20 domains in total

Almost Power Off

- ARM1176 (logic)
- SYS Peripheral
- SH-X2 (logic)
- 3D Graphics
- LCDC, MediaRAM
- Sound related IPs
- LVDS serial interface
- DSP
- ARM926

Almost Power On

- ARM1176 (RAM)
- SH-X2 (RAM)
- RT Peripheral
- INTC, RTSHwy
- DSP
- RAM
- BB-bus
- DMAC

Always On

Right domain must be ON

Common Power Domain

- C-SHwy,
- Local Bus,
- DDR cntrl2
- PLL
- Clock Generators
- Back-up FFs
- Repeater cells

Almost Power On

- Always On

Almost Power Off

- Almost Power Off

20 domains in total
Outline

1. Introduction
2. Chip Architecture
3. Power Management
   - Static Power (Hierarchical Power Domain)
   - Dynamic Power (Partial Clock Activation)
4. Memory Management
   - IP-MMU
5. Summary
Dynamic Power Management

- **At low power scenario** (music replay etc.), estimated power would go over the budget.
- All un-used domains already in power-off
- Clock gating already applied deeply
- In this case, **Common Power Domain** consumed large percentage of whole chip power (60%).

Note: The diagram shows the power consumption breakdown across different domains, with a particular emphasis on the Common Power Domain.
Partial Clock Activation: Motivation

[Problem cause]
- Big modules have already been in power-off.
- But accumulated “small” power is not negligible.

PLL and clock tree (PLL→modules)
Partial Clock Activation: Diagram

- Clock tree separation
- Intermittent activation

**Diagram**

- **PLL cntrl**
  - Clock gating control

- **Common Power Domain**
  - STOP
  - BYPASS

- **CPU_stby**
  - clkA
  - clkB

- **CPU**
  - clk1
  - clk2
  - clkC

- **1st clock driver A**
- **1st clock driver B**
- **1st clock driver C**

- **always clock-active region**
- **intermittent clock-active region**

**Clock tree separation**

**Intermittent activation**
Partial Clock Activation: Control

- Clock for SPU and serial IO is slowed down (PLL2 is bypassed.)
- PLL1 and clock tree is stopped in synchronization with CPU stand-by. They are reactivated in intermittent manner for housekeeping operation. (data copy from SD-card to SPU)
Partial Clock Activation: Result

Scene: Long time music replay

mA(1.2V)

Without PCA

With PCA

PCA: Partial Clock Activation

x0.6

x0.3

33.6

19.6

common

Others
Clock
Bus
DSP
CPU
Leak
Outline

1. Introduction
2. Chip Architecture
3. Power Management
   - Static Power (Hierarchical Power Domain)
   - Dynamic Power (Partial Clock Activation)
4. Memory Management
   - IP-MMU
5. Summary
Memory Management: Motivation

In general, efficient memory management is required for mobile phone due to limited memory capacity.

**Goal:** To reduce system cost by minimizing memory footprint

**However,**

1. HW-IP (Accelerator) requires *address-contiguous* large memory regions.
2. Memory fragmentation proceeds with application execution. Large memory chunk may not remain after some applications have run (even after exit)!
3. To avoid this fragmentation, large memory regions have to be *statically allocated at boot phase*.
4. As a result, system footprint on main memory becomes very large although *actual memory size in use at the same time is not so large!*
IP-MMU: Concept

(a) Without IP-MMU

(b) With IP-MMU

Address Translation by IP-MMU
IP-MMU: Benefit

Benefit

- Large address-contiguous memory regions can be composed of physically fragmented memory chunks. Memory regions can be freed after applications terminated and they can be allocated on demand. Reduction of system footprint.

- HW-IP can see Virtual Address Space like CPU. CPU and HW-IPs can directly share the "pointer" to the memory. Easy development of applications utilizing HW-IPs.
IP-MMU : Address Space View

- First half (2GB): TLB (Translation Look-aside Buffer)
  - Dynamically allocated by OS (4KB granularity)
  - Shared with a software process on OS
- Second half (2GB): PMB (Physical-address Mapping Buffer)
  - Statically allocated at boot time (always alive)
    (e.g. LCDC frame buffer)
  - Access window to arbitrary physical address
IP-MMU: Implementation (Basic)

Address space judgment

µTLB (8 ent.)

PMB (16 entries, full-assoc)
64MB page

main TLB (64 entries, 2-way)

Hardware Page Table Walker

Bus request

Bus response

Bus request

Bus response
IP-MMU: Problem

Inter-Connect Buffer (ICB) is the bus bridge that buffers data with 512 KBytes MediaRAM for inter-IP communication.

[Problem1]
Inserting IP-MMU at bus I/F (1) may degrade performance.

[Problem2]
Inserting IP-MMU to each IP side (2) is costly because 16 IP-MMUs are needed for ICB.
IP-MMU : Solution

High data throughput
Multimedia IPs

IP #0 R
IP #1 W
IP #2 R
IP #2 W

IP #3

IP #15

ICB access control

IP indep. μTLB
#0 - #2
2 ent. x4

IP comm. μTLB
#3 - #15
20 ent.

PMB

main TLB
TLB RAM

HW PTW

Buff. cntrl

Buff. RAM
512KB

Buff. on #0-#15

Bus interface

On chip bus

30 μTLB entries in total

IPs that share μTLB

common
IP-MMU: Evaluation results

- Statically allocated memory size
  53MB
  - Camera, display (35MB)
  - Graphics (18MB)
  - 10MB free (-81%)

Note: All IP are used with IP-MMU.

- Performance evaluation
  Degradation is negligible.
  Video camcorder application performance:
  30 fps @D1(720x520) for MPEG4/AVC format on G3
Summary

- The SH-Mobile G3 integrates a dual-mode (3G/2G) baseband processor, an application processor, and a multimedia processor with 65nm LP for mobile phone.
- Hierarchical Power Domain architecture separates the chip into 20 power domains and reduces leakage in many use cases.
- Partial Clock Activation reduces power even in low power scenario. It achieved more than 40% reduction of chip power in long time music replay scene.
- IP-MMU enables HW-IPs to use physically fragmented memory. By using this, more than 80% of statically allocated memory can be freed in the actual system without performance degradation.