

Low power architecture and design techniques for mobile handset LSI Medity™ M2.

Shuichi Kunie*, Takefumi Hiraga*, Tatsuya Tokue*
Sunao Torii**, Taku Ohsawa**

* Mobile LSI division, 2nd SoC Business Unit NEC Electronics Corporation
Kawasaki City, Kanagawa, 211-8668 JAPAN

** System IP Core Research Laboratory, NEC laboratories
Sagamihara City, Kanagawa, 229-1198 JAPAN

Abstract - This paper presents the low power architecture and design techniques for the mobile handset LSI Medity™ M2. M2 is a second-generation mobile handset LSI which integrates a Digital baseband and Application processor on a chip. M2 is capable of supporting 3.2 Mbps HSDPA, WCDMA communications, and rich, high-resolution multimedia applications, while power consumption is kept almost the same as in its predecessor chip M1. To reduce power consumption, M2 adopts hardware management clock control schemes, Multiple Vt transistors, an On-chip Power Switch, and Back-bias control. Preliminary measurement results show the design to work very well.

I. Introduction

Recent performance requirements for mobile phones continue to expand, now including higher resolution graphics, pictures, and rich JAVA applications. Simultaneous requirements for longer battery-life, however, run counter to the need for performance. This makes techniques for achieving lower power consumption an increasingly critical issue. Moreover, recent mobile phones also require modern packaging designs, such as very thin-widths and light-weights.

To meet these demands, we introduced Medity™ M1 in September, 2006. It integrated a 3G WCDMA digital baseband processor (DBB) and an application processor (AP) on a single chip. Since this integration offered a shared external memory (Mobile-SDRAM), it greatly reduced both the number of required chip components and overall cost. Medity M1 also applied a number of low power techniques to each IP, including root level clock gating, idle-time detection based clock frequency reduction, and power domain separation to cut off leakage power in idling macros. These mechanisms greatly contributed to the achievement of a best-in-class 700-hour standby time and a 3.5-hour calling time.

Next-generation mobile phones, however, will require a number of new functions.

- Higher resolution motion picture play and record (VGA-class), as well as support of higher resolution CMOS cameras. (5Mpixel-class)
- Surface digital TV play and record
- Support of multiple communication standards (GSM, CDMA)
- Support of high speed down-link packet access (HSDPA)

To meet these requirements, we have developed Medity M2 as a successor to "M1". M2 integrates third generation (3G to 3.5G) W-CDMA, HSDPA communications technologies, and application functions, using advanced low power technologies, optimized for mobile handsets.

The combination of these advanced technologies has resulted in a 50 percent reduction in power consumption. This improvement has been made possible by a host of advancements in circuit design and layout, including dynamic frequency scaling, automatic hierarchical clock control, LCD Direct Path technology, and on-chip power switch technology. In addition, the use of 65-nanometer process technologies, such as multiple Vt transistor technology, and UltimateLowPower™, which is a back-bias control technique based on Transmeta LongRun2™ have also contributed to lower power consumption. These advancements in power consumption technology not only help to extend battery life and usage time, they also contribute to minimizing the size of the mobile handset battery itself, leading to flexibility in designing slimmer handsets, and also to reducing environmental impact.

In Section II, we give an overview of the M2 chip architecture. In Section III, we discuss low power issues. In Section IV, we describe M2 low power techniques, frequency and clock control, the multiple Vt transistor design, and an on-chip power switch and back-bias technique. In Section V, we present real-chip measurement results. Finally, in Section VI, we summarize our study and discuss future work on EDA tools.

II. Medity M2 Overview

TABLE I shows Medity M2 functional specifications in comparison with those of M1. M2 includes many functional extensions from M1. And an architecture design to meet bus and memory bandwidth demands is a very critical issue here. M2 employs the hierarchical multiple bus architecture shown in Fig. 1. The main system bus is 64bit – AMBA AXI 166MHz, has a split transaction protocol, and connects four latency sensitive masters. ARM1176JZF-S™, a DSP, a 3D-Graphics accelerator, and a DBB-AP Bridge connect directly to a mobile-DDR SDRAM interface in order to expand bus bandwidth and reduce memory latency.

Moreover, to connect legacy macros to this AMBA AXI bus, M2 employs three AMBA AHBs hierarchically. Each of these connects a grouping of macros to the AMBA-AXI bus via an AXI-AHB bridge. Two of the AMBA AHBs run at 166MHz, while the other runs at 83MHz.

TABLE I
Medyty M2 / M1 functional specifications

Category	Medyty M2	Medyty M1
Application	CPU/DSP	ARM1176, DSP
	System Bus	AXI
	LCD	WVGA
	Video codec H.264	D1 Enc, Dec (HW accelerator)
	3D Sound	128 chords
DBB	CPU	ARM926
	Communication	W-CDMA
		HSDPA
		GSM/GPRS
Clock	Application CPU/DSP	500MHz
	DBB CPU	250MHz
	DDRSDRAM	166MHz
		125MHz
Gate Count	Logic	Logic: 15Mgate
	SRAM	SRAM: 12Mbit
Process	Rule	65nm
	Transistor type	Low/Mid/High Vt
Power Supply	Core	1.2V
	I/O	3.0V, 1.8V
Package	Type	FPGA
	size	529pin, 14x14mm
Others		Power Switch UltimateLowPower

The appropriate AMBA AHB for any given macros to be connected to is determined in consideration of its bus bandwidth requirements and clock frequency specifications. The DDR SDRAM controller has a memory access request queue and can schedule very effectively, taking into consideration the possibility of burst transfer to the same page, bank conflicts, read-write changes, and QoS (Quality of Service) priorities.

III. Low power issues

There are basically three situations into which mobile phones might be put:

- Execution of heavy workload applications, such as 3-D graphics, TV phones, games, and video recording functions.
- Execution of light workload applications, such as telephone and E-mail functions.
- Standby mode, while waiting for the reception of telephone calls or E-mail messages.

In heavy workload situations, power consumption is extremely high, as applications require the booting up of specific engines and they run at higher frequencies.

In light workload situation, only a relatively low energy supply is required because operational frequencies are low and

almost no engines for specific applications are being used. The dynamic power for clock buffer switching and clock divider is dominant and leakage power has a bigger impact in this situation than in the heavy workload situation because the percentage of leakage power is higher with respect to total power consumption. That is, reducing the above dynamic power and the leakage power is very important for extending battery life.

In the standby mode, by way of contrast, most power consumption is due to leakage power, and reducing this leakage extends standby time significantly.

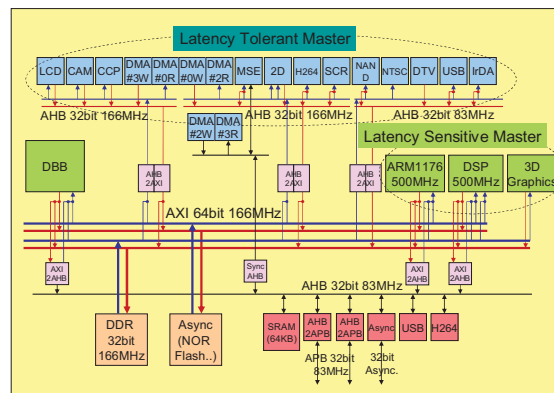


Fig. 1. M2 hierarchical multiple bus architecture

IV. Medity M2 Low power techniques

Medyty M2 provides the following effective measures to minimize both dynamic power and leakage power for improved LSI performance:

- Reducing dynamic power
 - Automatic frequency control and Hierarchical clock gating
- Reducing leakage power
 - Multiple Vt transistors employed in same block and on-chip power switching
 - UltimateLowPower™ (Back-bias control)

A. Automatic frequency control and Hierarchical clock gating

In M2, dynamic power is reduced by the use of both dynamic frequency control and clock gating techniques. While system software manages frequency adjustments and clock gating, it requires human-intensive work is required to develop such software. To avoid this difficulty, M2 employs the following two independent hardware-level clock regulation mechanisms:

- Automatic frequency control: As soon as an active monitor in M2 has detected that its operating condition is that of a predefined low workload condition, it notifies a Multi-frequency generator. When the Multi-frequency generator receives such notification, it considers the overall system status, and may start reducing all chip clock frequencies.

- Hierarchical clock gating: As soon as a function macro goes into an inactive state, the active monitor detects this, and the clock supplement to this macro is stopped in order to reduce clock buffer switching and clock divider power consumption. Clock supplements will be resumed with the reception of interrupts or activation requests from other, related macros. Further, an active monitor detects individual idle states in portions of macros, and when it does so, macro-local gating cells are used to stop clock supplements. Each macro has sub-macro level gating and leaf FF level clock gating.

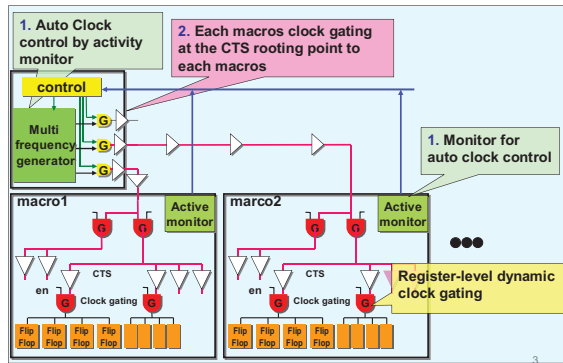


Fig. 2. Frequency management and clock structure

Fig.2 illustrates frequency management and clock gating, including a clock tree circuit and a regulation mechanism for each function macro. There is an active monitor in each macro and a Multi-frequency generator for the system as a whole.

The clock regulation mechanism has been designed very carefully, employing EDA tools and hand optimization to prevent the wasting of clock operations. The hardware-level mechanisms described above reduce power dynamically, without software clock management. M2 also provides conventional software for power reduction, and the combination of hardware-level and software-level measures results in significant reductions in dynamic power consumption, especially in the execution of light workload applications.

B. Multiple Vt transistors and on-chip power switch design

To reduce leakage power, M2 adopts a multiple Vt transistor design and power domain separation. By using three different types of transistors, High Vt, Middle Vt, and Low Vt, M2 achieves both high-speed operations and low leakage power.

In general, critical paths are synthesized with EDA tools, using Low Vt transistors in order to meet timing constraints. By way of contrast, paths with sufficient slack are generally synthesized using higher Vt transistors. It often happens; however, that low transistors are used in conditions in which high would be appropriate.

In M2, the use of multiple Vt transistors is optimized during each design phase. First, in the logical synthesis phase, only

High Vt and Middle Vt type transistors are used. In the layout phase that follows, EDA tools can be used to replace Middle Vt with Low Vt, except in domains in which power is always on. In such domains, in order to meet timing constraints, replacement to Low Vt can be done by hand without resulting, as would be the case with EDA tools, in the unnecessary use of Low Vt transistors.

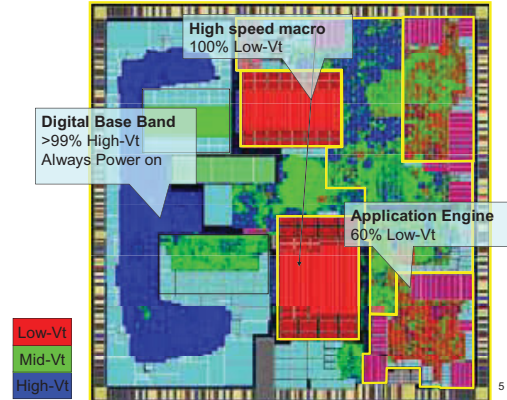


Fig. 3. Multiple Vt transistors in M2.

High speed macros, the application CPU and DSP, as well as certain application engines (e.g. a 3D graphics accelerator and a motion codec accelerator, which are activated only when executing heavy workload applications) are located in power domains separate from one another. M2 has 7 power domains, while M1 had 5 power domains. Fig. 3 shows multiple Vt transistors in M2. As noted above, reducing leakage power is critical in light workload situation and in the stand-by mode.

EDA tools are initially used in the layout phase in domains in which power is not always on, in order to conduct replacement with Low Vt for the purpose of meeting time constraints. In the next step, this replacement is continued for the purpose of reducing area requirements. For example, in a 3D graphics accelerator, the first step results in 14.3% of the transistors being Low Vt, and after area optimization, results in an additional 46.5% of the transistors being replaced of roughly 20% to Low Vt (for a total of 60.8%), and an area reduction is realized at a same time. This is all made possible by the use of an on-chip power switch.

M2 also employs an on-chip power switch. It is possible to reduce the leakage power by switching the power off via separated external power balls in a SoC (a technique used in M1). An on-chip power switch, however, can realize shorten switching time, to a roughly 30-40usec system on/off mode transition time. This makes it possible for M2 to reduce leakage power even during short idle time period.

In addition to improvements in system level power consumption, the M2 also helps reduce the total bill of materials (BOM) by decreasing the number of required external parts.

There are two major issues in the implementation of an

on-chip power switch: switch area cost and rush current during simultaneous turn-on.

The former issue can be resolved by replacement with Low Vt transistors, as previously noted, and the latter issue by hand optimization of rush current levels. Fig. 4 shows rush current image in a power-ons switching sequence.

An EDA tool calculates the transition of rush current from such given parameters as the number of switches, the number of simultaneous turn-ons of switches, and the interval timing among different switch groups. The biases of the results, we chose the best power switch parameters.

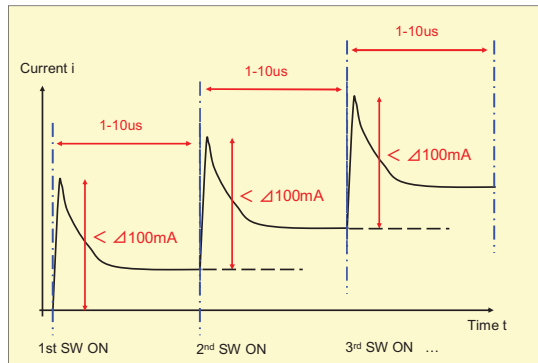


Fig. 4. Power-on switching sequence.

M2 also has a small processor used as a PMU (power management unit). On the basis of the current activity status of each macro, it controls PLL settings, provide clocks, de-asserts resets, allows interrupts, and manages the power switch for the CPU and for others logics. Since the PMU is controlled by system software, power management in M2 is very smooth and flexible and this helps to meet system requirements.

C. UltimateLowPower™ (Back-bias control)

In general, LongRun2™ uses a back-bias technique to reduce leakage power and is suitable for full custom, such as very high performance processors; however, NECEL UltimateLowPower™ employs a cell-based design for general-purpose processor, SoC and ASSP. It has three main features.

The first is its use of process and device techniques that the transistor cell libraries are designed and optimized whose leak current is well sensitive to the back-bias voltage.

The second is the use of a chip-architecture and chip sign-off system that enhances, in the design process, the general EDA tool flow, as well as STA rules and design conditions in ASIC flow. SoC designer can place a monitor macro with automatic back bias control like standard cell libraries and can use timing libraries with fixed voltages as sign-off point for STA against DVFS (Dynamic Voltage and Frequency Scaling)

The third is a circuit technique which automatically controls back-bias range using monitor macro and then employs DVFS, Transmeta LongRun® and LongRun2™ to prevent variation in transistor leakage.

Fig 5 shows optimized transistor characteristics for UltimateLowPower™. NECEL has developed a transistor which is more sensitive with respect to back-bias voltage, which helps it expand the leakage control areas. As a result, the controllability of the leakage current is significantly improved.

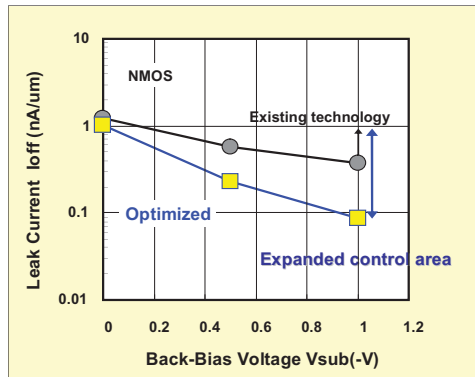


Fig. 5. Optimized transistor characteristics for UltimateLowPower™

Fig. 6 shows a monitor macro with automatic back bias control. By monitoring the timing of the ring oscillator integrated in the M2, it measures logic speed and compare it with target delay. After that it controls leakage power by automatic back-bias adjustment of Nch and Pch if the detection of the delay is small.

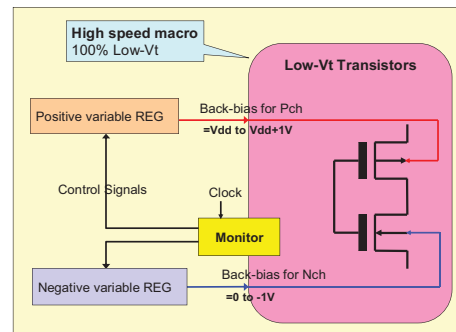


Fig. 6. Monitor macro with controlling back-bias automatically

In the deeper sub-micron process, the differences in leakage power due to process variation become very huge.

FAST type transistor is high speed and big leakage power transistor in first corner on the chip. TYP type transistor is normal speed and middle leakage power transistor in center corner on the chip. SLOW type transistor is low speed and small leakage power transistor in slow corner. In 65nm process generation, variation in leakage power range is 10 times or more, and effect of this chip yield cannot be ignored.

Fig. 7 shows the effect image of back-bias in UltimateLowPower™. The characteristics of the transistors can be shifted from FAST type to TYP type with controlling back-bias voltage unless the effect of the TYP and SLOW type characteristics.

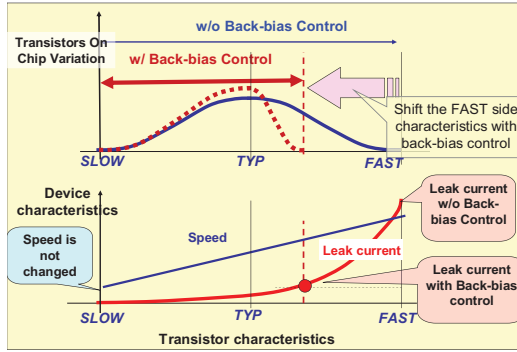


Fig. 7. The effect image of back-bias

M2 applies UltimateLowPower™ to high speed CPU and DSP applications, which rely on low-Vt transistors. UltimateLowPower™ maintains optimal threshold voltage and helps reduce leakage power. CPU and DSP speed can be optimized while leakage power is reduced, without leakage variation, by using a monitor macro.

V. Power Measurement Results

Fig. 8 shows relative power consumption for various power modes in Medity M2. In this graph, M2 is executing a music player displaying static pictures to an LCD on a Linux operating system. This is a light workload, which means that Medity M2 requires maximum effectiveness in power consumption despite the fact that the CPU is running at a 500MHz maximum frequency.

Using all its low power features (frequency control (FRQ control), clock gating (CLK control), and power switch control (P-SW control)), M2 achieves an 80% reduction in power over that required without use of the low power features. This indicates that these low power features work well and that M2 will meet power consumption requirements.

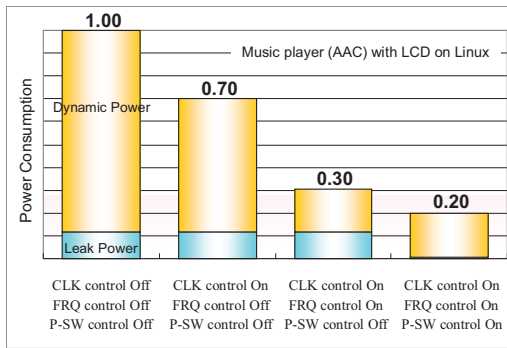


Fig. 8. The effect of each low power techniques.

Table II shows power consumption for various applications. As may be seen, M2 achieves a level of less than 70mA during the execution of such heavy workload applications as H.264 D1 30fps decoding.

Fig. 9 shows the results of the reduction of leakage power

with UltimateLowPower. DFAST is the target delay value around a FAST sample's ring-oscillator. DTYP is the target delay value around a TYP sample's ring-oscillator.

TABLE II
Power consumption of various applications

Application	mA@1.2V (w/o I/O)
VGA displaying	5.2mA
Audio dec. (Enhanced AAC+, 48Kbps)	23.8mA
Video dec. (H.264, QVGA, 30fps)	22.3mA
Video dec. (H.264, D1, 30fps)	66.5mA

The DFAST of target delay for FAST samples is without any back-bias, which means that the leakage current in FAST samples (especially in F/F and S/F) will be large. Leakage current with back-bias (target delay is DTYP), however, is the same as in T/T samples, and leakage current in TYP samples (T/T) is not affected. That is, the speed of TYP samples is not affected. This shows that back-bias control with a monitor macro in UltimateLowPower™ effectively decreases leakage power due to process variations.

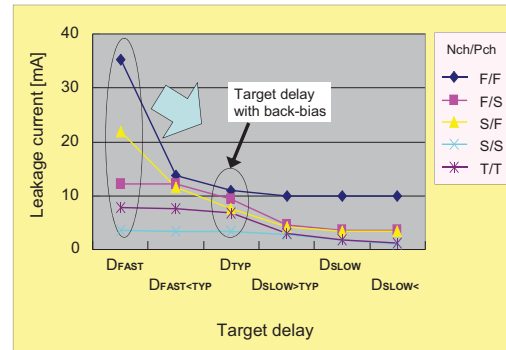


Fig. 9. Relationship with leakage current and target delay

VI. Summary and Conclusions

Meditry M2 achieves best low power results on 65nm process even if the logic and memory is exceed 2 times comparing M1 power consumption and contributes to mobile area with such a high performance. Fig.10 shows M2 die and package photos. M2 has its application portion and digital base band portion on a single chip, whose size is only 8.52x8.52mm². The package is 529 pin PoP(Package on Package) and can use a general memory chip.

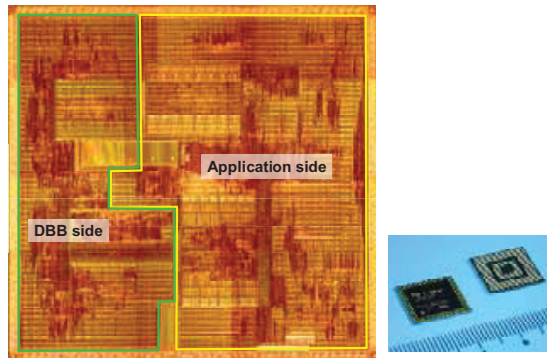


Fig. 10. M2 die and package photos

Truly ultimate low power consumption is achieved by power optimization in all layers, from the front-end RTL design level to the back-end transistor primitive level. M2 applies a combination of low power techniques to all layer levels, and as a result, power consumption is roughly the same as in the previous M1, despite the doubling of its performance and the addition of new functions.

We have, however, encountered several fundamental problems in the design of M2 with respect to the use of EDA tools, particularly regarding the DVFS, the power switch, and the replacement of multiple Vt transistors. While many EDA tools support multiple Vt Transistor design in each design phase, their use tends to result in excessive transition to Low Vt transistors, due timing closure.

The unfortunate effect of this tendency is to increase both leakage power and the number of hold buffers needed to maintain minimum hold-timing constrains. In this area, hand optimization sometimes offers better results than do EDA tools. Also, automatic on-chip power switch optimization with EDA tools is neither optimized nor automated and requires numerous iterations of rush current calculations.

In addition, the layout data size of recent deep sub-micro processes has continued to increase. This increases development time for EDA tools. Some designing requires more than 100 hours for only a single operation. So the development iteration “RTL, synthesis, layout, STA” spends huge time and it decreases productivity of such a large SoC design significantly. To boost the design efficiency, increasing time to time performance of EDA tools is most critical issue.

Finally, tool-based low-power designing needs to be matured even further, and it is very important to design low power SoCs using CPF (Common Power Format) / UPF (Unified Power Format) standards. The tools themselves, however, as well as design flow process, are still in the early stage of development. While differences in interpretation among individual vendors have made it difficult to obtain satisfactory design results, we may hope that future development will contribute to improving the quality of EDA tools.

References

- [1] Nomura, M. et al., “Delay and Power Monitoring Schemes for Minimizing Power Consumption by Means of Supply and Threshold Voltage Control in Active and Standby Modes,” *IEEE J. of Solid-State Circuits*, vol. 41, no.4, pp.805-814, Apr. 2006.
- [2] Torii, S. et al., “A 600MIPS120mW 70uA Leakage Triple-CPU Mobile Application Processor Chip,” *ISSCC Digest of Technical Papers*, pp.136-137, Feb.2005.
- [3] Royannez, P. et al., “90nm Low Leakage SoC Design Techniques for Wireless Applications,” *ISSCC Digest of Technical Papers*, pp.138-139, Feb.2005.

LongRun2 is trademark of Transmeta Corporation. NEC Electronics is either a registered trademark or trademark of NEC Electronics Corporation. All other product or service names mentioned herein are the trademarks of their respective owners.