고성능 시스템 설계에서의 클럭 신호 분배

정태경 · 이장호

Clock Distribution in High-Performance System Design

Taikyeong T. Jeong · Jang Ho Lee

요 약

수용 가능한 수준의 성능을 동시에 전달하고 분배하는 동안의 소비전력을 줄이는 문제는 고성능 시스템의 설계 분야에서는 더욱 더 결정적인 관심사로 받아지고 있다. 본 논문에서는 전력분배의 문제를 클럭 신호 발생과 분배의 관점에서 제시하고자 한다. 우리는 클럭 신호의 전력 효율성과 다른 응용제품 이외에도 무선통신의 회로에서도 찾아 검증하였다.

ABSTRACT

The problem of reducing power dissipation while simultaneously delivering acceptable levels of performance is becoming a critical concern in high performance system design. In this paper, we present this power dissipation problem from the clock generation and distribution side. We examine clock power efficiency and several applications as well as wireless communication circuits.

키워드

Clock distribution, transmission line, System-On-a-Chip (SOC), High-performance

Ⅰ. INTRODUCTION

While Very Large Scale Integration (VLSI) technology has evolved to satisfy the demand of faster speed and less power dissipation, high performance applications have further increased system power as well as operating speed.

In addition, low power clock distribution is one of the most critical areas in the design of high performance VLSI devices. Poor clock distribution can result in excessive clock skew between clusters on the chip, thus reducing the maximum operating frequency. Therefore, it is necessary to reduce the effect of clock skew on a chip. This requires reducing the interconnection delays and balancing the delays of all the clock paths [1].

In general, the total power consumption consists of dynamic (switching) power consumption, short-circuit power consumption, and leakage power consumption [2]. Therefore the total power consumption in CMOS circuit design can be expressed as follows.

$$P(x)_{total} = P(x)_{dyn} + P(x)_{static} + P(x)_{sh-cr} \quad (1)$$

* T. Jeong was with Department of Electrical and Computer Engineering, the University of Texas at Austin, Austin, TX 78712. He is now with the Myongji University (ph:: +82-31-330-6775, email: ttjeong@mju.ac.kr)

Manuscript received Dec, 12, 2005; This works partially supported by NASA Goddard Space Flight Center (Grant no. NNG05GJ38G).

접수일자 : 2005. 12. 16

Note that the average dynamic power dissipation of a CMOS gate is essentially independent of all transistor characteristics and transistor sizes as long as a full voltage swing is achieved. The principle of dynamic power consumption ($P(x)dyn$) is proportional to V_{DD}^2. The node transition factor (α), which is the effective number of power-consuming voltage transitions, is proportional to the voltage swing per clock cycle. Therefore, the expression for average dynamic power consumption becomes:

$$P(x)_{dyn} = \alpha C_{load} V_{DD}^2 f_{clk} \quad (2)$$

Power dissipation is composed mostly of dynamic power ($P(x)_{dyn\_power}$) that is derived from the equation:

$$P(x)_{dyn\_power} = \frac{1}{T}\int P(t)dt = \frac{V_{sup\ ply}}{T}\int i_{sup\ ply}dt \quad (3)$$

Consequently, power consumption in low power and high performance microprocessors and systems has been a concern in the realm of high performance devices. This is quickly become more important as newer chips are expending more and more power.

To reduce clock power, one or more of the terms needs to be reduced in the dynamic power dissipation equation. The frequency is generally not a candidate since increased functionality is dependent on VLSI systems running faster. Therefore, the other aspects of the power equation are considered using approaches such as clock gating, global clock power reduction, and adiabatic techniques. It is difficult to compare them since different techniques are used to quantify experimental results, but it is useful to examine the current techniques, more aggressive techniques, and adiabatic techniques.

The purpose of this paper is to provide a result that utilizes a wide variety of circuits and is as large of a standard functional unit as possible for validating low power clock distribution.

Ⅱ. LOW POWER CLOCK GATING

2.1 Minimization of Power

Some other design methodologies, including circuit-directed performance enhancing techniques which are incorporated in the main characteristics of energy efficiency design, ought to be aggressively exploited so as to eliminate or restrict the need for power demanding hardware circuits.

According to the total power dissipation equation, we can minimize power by minimizing the supply voltages and capacitance.

$$td = (\sum k\frac{Cload}{\beta W})\frac{V}{V - V_t^{\alpha}} \quad (4)$$

However, it should be noted that we could have a worst-case (V_{DD}=0, and C_{load}=0). In which it would be necessary to investigate more about the performance issue. The output node of a CMOS gate can normally undergo one power-consuming transition in each clock cycle period. This assumption is not always correct: the node transition rate can be slower than the clock rate, depending on the circuit topology, logic style, and input signal statistics [2]. The transition factor (α) is the effective number of power-consuming voltage transitions experienced per clock cycle.

Furthermore, the proposed design methodology will be used to model more practical and complicated circuits such as dynamic and low power clock circuits.

2.2 Circuit Validation

The main goal of this design methodology is to leverage voltage scaling to lower the power dissipation of circuits. Basically, the approach is to lower V_{DD} and resize circuits to reduce the power and nullify the delay problem.

For instance, the 21264 Alpha processor was reported to dissipate 40mW of the total power (110mW)in the global clock distribution and another 18mW of the total power in the latches and gates of the local clock network based on switching capacitances [3].

New designs are showing significant increases in local clock power that can be attributed to higher clock frequency, interconnections that do not scale as well as devices, and more dynamic gates and latches driven by the clock due to increasing integration and pipelining. Without aggressive clock gathering, the local clock distribution makes it the dominant power consideration in processor design.

We should optimize the chip and system performance using both the delay and the power parameters. In this case, reducing the delay is *high speed*, lowering V_{DD} creates *low power* circuits, and reducing the circuit size creates a *high density* circuit. Eventually we will be able to reduce power dissipation and achieve the same performance.

Ⅲ. GLOBAL CLOCK POWER REDUCTION

3.1 Clock Generation and Distribution

Current clock generation and distribution techniques attempt to reduce clock power by reducing the switching capacitance of the distribution network. Clock generation in microprocessors is accomplished through Phased-Locked Loop (PLL) design and distributed through H-trees and clock grids. The dominant techniques for reducing power in the clock tree are optimization of the distribution network and clock gating [11]. In an H-tree design, the path from the clock generation circuit to the individual latches is designed to have the same delay so that the clock is in phase at each point of use on the chip.

A grid reduces clock skew by reducing the resistance of the network, but increases the capacitance. The effective capacitance in clock gating is reduced by turning off portions of a microprocessor that are not being used. For this method to be effective, significant portions of the logic cells of the processor must have low switching factors and the overhead of gates and routing might cost more than the power saved. Clock gating showed that power consumption could be reduced dramatically in designs with low activity in design cells of the clock tree with power savings for a 10% activity factor, but units that have a 90% activity factor with clock gating dissipate as much power as a unit with 100% activity factor without clock gating due to the capacitive overhead of the wires and gates. Furthermore, gating adds delay, skew, and timing issues that are non-trivial.

An analysis of the Alpha 21264 floating-point unit shows a power reduction of 25% using clock gating [10]. In spite of these approaches, clock power continues to rise. Gating has proven to be effective in minimizing power, but adds considerable complexity to the design flow. Therefore, clock power continues of grow, leading to other methods of clock power reduction that have been proposed.

3.2 Simulation Results

Several alternative methodologies have been proposed to reduce power dissipation on the global clock distribution. The most conventional approach involves reducing the voltage swing of the clock signal. Converting from high to low voltage can be done using a simple inverter connected to the lower supply voltage, but converting back up requires a different amplifier or some other multi-transistor circuit that consumes more power than a simple inverter.

Nevertheless, some power savings should arise. As supply voltage (V_{DD}) continues to scale, power savings from the reduced swing will also decrease since the overhead voltage for reduced scaling will not be as large.

Table1. Power Comparison in Buffered vs. Reduced Swing Clock Distribution
표1. 전력소모의 버퍼 와 감소된 클럭의 분배표

Benchmark	Buffer Insertion-0.5μm		Reduced Swing-0.25μm		
	Base(mW)	GRIN(mW)	Base(mW)	Dual-V_{DD}(mW)	Low Swing-V_{DD}(mW)
R1	19	8.8	10.6	5.3	6.52
R2	24	18	19.9	11.1	13.4
R3	50	21	24.9	14	17.5
R4	63	37	52.2	28.7	36.2
R5	139	54	73.2	42.3	53.8

Table 1shows a comparison between an optimization buffer insertion for low skew/low power and a result of power savings of dual-V_{DD} and low-swing clock distribution [13]. The first column shows power scaling trends using the GRIN optimization algorithm in 0.5μm to 0.25μm technology operating at 200MHz and an assumed 2.5V for the 0.5μm

technology. The second column implemented dual voltages of 2.5V and 1.8V and then low-swing voltages that were a threshold above ground and below V_{DD} at 500MHz. Each column implemented a baseline distribution using buffer insertion at the root nodes for comparison. There is very little difference in power dissipation between the two approaches.

3.3 Distributed Circuit Model

Another alternative clock power distribution method that has been gaining popularity in the design community is to resonate a clock signal on a transmission line in various configurations. Figure 1 shows a lossless series terminated transmission line with a high impedance load that can be used to approximate a global clock wire and voltage waveforms at various points along the line [8].

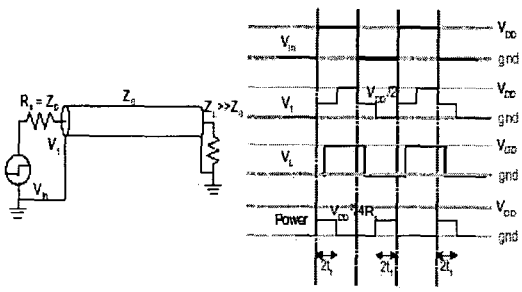

Fig 1.Distributed Circuit Model for Transmission Line
그림1. 트랜스미션라인의 분배된 회로 모델

The theory of operation of these transmission lines can be explained using a distributed circuit model for the transmission lines. It can be defined as a distributed parameter network within the circuit parameters distributed throughout the line.

Of particular practical interest in microwave electronics is the *lossless* transmission line - a transmission line where *R* and *G* are negligible.

In general, if the time of flight across the interconnection line (as determined by the speed of light) is much shorter than the signal rise/fall times, then the wire can be modeled as a capacitive load, or as a lumped or distributed *RC* network. If the interconnection lines are sufficiently long and the rise times of the signal waveforms are comparable to the time of flight across the line, then the *inductance* also becomes important, and the interconnection lines must be modeled as transmission lines. In addition, as the width of the metal lines shrink, the transmission line effects and signal coupling between neighboring lines become even more pronounced [3].

In a lossless transmission line, phasor analysis will be used to obtain the steady-state solution. Added the phasors $V(x)$, $I(x)$ and the term in parentheses must be zero, namely

$$\frac{dV(x)}{dx} = -j\omega LI(x) \tag{5}$$

$$\frac{dI(x)}{dx} = -j\omega CV(x) \tag{6}$$

Equation 5 is the differential equation satisfied by the phasors $V(x)$ and $I(x)$ along a lossless transmission line. Equation 7, which means the propagation constant (in radians per meter), $\beta = \omega\sqrt{LC}$ and Z_o can be defined as the characteristic impedance of the transmission line as $Z_o = \omega L/\beta = \frac{\omega L}{\omega\sqrt{LC}} = \sqrt{\frac{L}{C}}$. In other words, the function of $A\cos(\omega t - \beta x)$ is a positive quantity. This shows that the point of constant phase is moving toward the right (i.e., toward the load in a transmission line). Consequently, the wave functions are defined as

$$v(x,t) = A\cos(\omega t - \beta x) + B\cos(\omega t - \beta x) \tag{7}$$

$$i(x,t) = \frac{A}{Z_o}\cos(\omega t - \beta x) - \frac{B}{Z_o}B\cos(\omega t - \beta x) \tag{8}$$

The function represents a traveling wave moving at a velocity (V_p) toward the load. This wave is called an outgoing wave when viewed from the source, or an *incident wave* when viewed from the load. The analysis of $B\cos(\omega t+\beta x)$ will show that this function represents a traveling wave moving at velocity (V_p) toward left (i.e., toward the source in a transmission line)[20]. This wave is called an incoming wave

when viewed from the source, or a *reflected wave* when viewed from the load. This is an ***incident wave*** and ***reflected wave*** preferred in transmission-line work, respectively.

3.4 Global Clock Distribution

The initial voltage step is 50% of the input voltage and it remains so until the reflected waveform from the load bumps it up. Power is drawn in this ideal case during the intervals there is a voltage drop across the series termination resistance. If the length of the line is designed such that after 2 τf, the reflected pulse arrives just as the net pulse is beginning, the reflected pulse effectively charges the line for the next pulse and power dissipation effectively drops to zero. There will be current drawn in the first pulse, but after that, charge will be conserved. In reality, there are losses in the load and transmission line, but also a great potential for power saving [13].

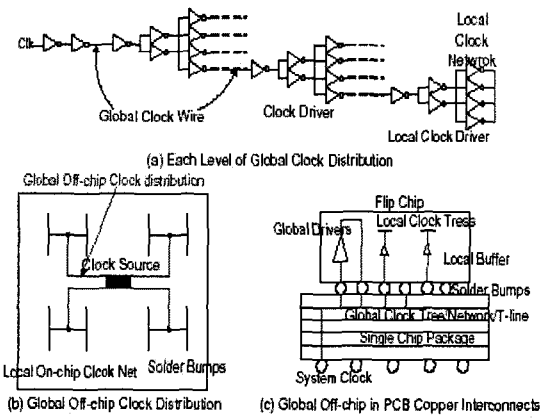

Fig 2. Transmission Line Configurations
그림 2. 트랜스미션라인의 구성과 설정도

Each level of the global clock distribution is designed as a transmission line as in Figure 2 (a). Furthermore, the transmission line is moved off-chip to take advantage of PCB copper interconnects as shown in Figure 2 (b), (c) [12]. The clock drivers demonstrated up to a 66% power reduction over distributed buffers, but no data is given for the test chip which will have much lower savings due to the extra drivers in Figure 2 (a).

All of these techniques work best in global distribution since a wire is driven with a single gate at the load end. It does not work as well for local clocking due to the fact that the initial wave may not exceed the turn-on voltage of the gate, and gates close to the driver will receive the clock pulse before those at the end so skew can be a problem. With the introduction of thick high-level metal layers in copper the off-chip solutions may not be as necessary as before with aluminum (*Al*) interconnect. The sum of waves at any point in the line is given by

$$Vx,t = 2A\sin(\omega t)\cos\frac{x-l}{v} \tag{9}$$

The term $\sin(\omega t)$ defines the time varying portion of the signal and it has equal phase at every section of the line with an amplitude of $2\,Acos\,(x\text{-}l)\,/\,v$ term. Clock signals can be taken directly from the distribution line at appropriate locations with nearly exact phase. Lossy transmission lines and distributed receivers introduce phase errors.

IV. ADIABATIC TECHNIQUES

Another scheme for energy loss systems is the design of the adiabatic circuits. It is an attempt to minimize the voltage drop across dissipative loads to reduce I^2 x R losses in the system and to recycle as much stored charge as possible. The following equation between switching time and transiting energy illustrate the goal of adiabatic circuit design

$$E_{dissipation} = P \times T = I^2 R \times T = (\frac{CV}{T})^2 = \frac{RC}{T} CV^2 \tag{10}$$

To recycle charge and reduce charging time, the clock is slowed down and logic is driven by the clock instead of the system power supply. By reducing the charging time, the energy dissipated can be reduced [5]. Consequently, the clock signals are generated using off-chip *LC* units or switched capacitor circuits that recycle energy, with the capacitance of the distribution network making up some of the unit or

switched-capacitor storage capacitance. True adiabatic systems are rare due to non-idealized CMOS devices and the overhead of recycling charge from small capacitive loads [5].

The AC-1 microprocessor is a 16-bit architecture designed in 3.3V, 0.6μm technology, and uses an adiabatic circuit shown in Figure I

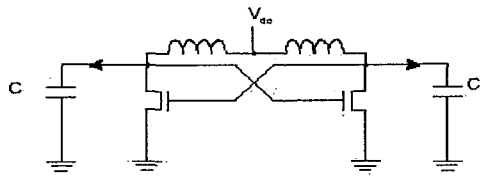

Fig 3. A Schematic of Resonant Driving Circuit
그림3. 공명된 회로의 구성도

Its stability fluctuates with the load capacitance, which will vary depending on the data patterns, and varies by several percent. The processor's clock could be run from a conventional clock or resonant clock. The clock power dissipation was 90% of the total power which dropped to between 60% and 70% in resonant mode when run with a conventional clock.

Another aspect of a low power microprocessor, based on the AC-1, was implemented to compare a power dissipation drop by 50% in the resonant driven microprocessor over the static one [7]. While not a very good comparison, there is some evidence of power reduction; however, their processor only ran up to 54MHz.

V. LC CLOCK POWER SAVING CIRCUIT

A wireless distribution system contains transmitter at the center of the chip broadcasts the synchronizing clock signal to receivers located at each block of the devices. The cost in power of the transmitters and receivers is shown to be less than that of a clock grid, but a little more than an H-tree [8]. Another approach uses a two-phase clock driver where the complementary outputs are each loaded with large capacitors and connected by transmission gates so that one line is charged by the discharging of the complementary line providing an 18% improvement in power-delay product of the clock distribution over conventional drivers in 3.3V, 0.35μm technologies. Care must be taken with this technique to prevent the clock lines from short circuiting, power that is lost in the transmission gates.

To reduce clock power, the problem of local clock distribution must be solved. In order to do this, techniques of recycling clock power will be borrowed from adiabatic logic. Several forms of clock generation, including *LC* units, are used in the Voltage Controlled Oscillator (VCO) on IBM's the 1GHz processor [4]. Passive *LC* oscillators have better phase noise characteristics than VCO. In particular, a passive *LC* unit can be used to filter input noise and does not generate switching and power supply noise like an active VCO. This is not such a large problem as PLL jitter has scaled well with technology. Interconnect delay has not scaled as well and buffering used to distribute global clocks is subject to noisy power supply lines. The large local capacitive load on the clock tree requires a large number of buffers, adding to the noise issue. Reducing the size of the load would help reduce the need for large buffers. The potential advantages of a local *LC* clock driver derive from its charge recycling. It would reduce the power required from the global clock network, thus reducing the number of drivers, jitter, and skew in the global clock. It is also conserving power in the local distribution.

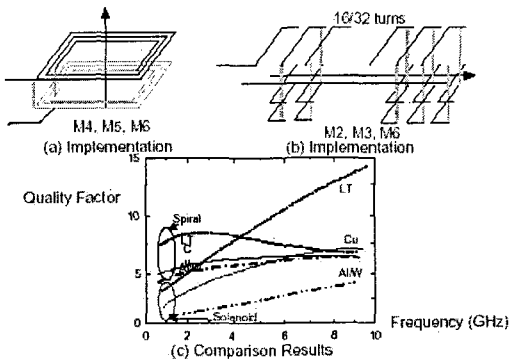

Fig 4. Comparisons of Spiral and Solenoid Inductor Configurations.
그림4. Spiral과 Solenoid Inductor의 비교결과

On chip inductors that locally-controlled *LC* units using on-chip solenoid inductors could provide significant power saving. Several spiral and solenoid inductors, as shown in Figure j (a) and Figure j (b) were implemented in 0.25μm aluminum (*Al*) technology with tungsten (*W*)vias. The spiral inductors used the top 2 layers of the 6 layer *Al* process and ranged in size from 100 x 100 μm^2 to 400 x 200 μm^2.

The copper data is a result of extrapolating 0.25μm technology to 0.18μm and assuming values for all copper wires and vias. Figure j (c) shows the measured quality factors of the aluminum inductors and extrapolated quality factors for the copper inductors. Although the spiral inductor's Q is higher at low frequency than solenoid inductors, it has a larger area, higher substrate losses at high frequency, and is more impacted by dense flip-flop packaging environments. The solenoid inductors do not suffer from substrate losses and are less influenced by external environments and are more suited to digital circuits.

Therefore, the power dissipated in a *LC* unit is equal to $P(x)_{avg} = I_{in}^2 R/2$, where the 1/2 term is due to the input current being a sinusoid. Assuming a sinusoid output voltage with a maximum value of V_{DD}, this can be re-written as $P(x)_{avg} = V_{DD}^2/2R$. Dividing this by power dissipation in a modern clock circuit is given by Equation 12, with the simplification that in the clock network $\alpha = 1$, gives the power ratio ($P(x)_{ratio}$)

$$P(x)_{ratio} = \frac{P(x)_{resonant}}{P(x)_{buffer}} \frac{1}{2RCf} \tag{11}$$

where R is the resistance in the oscillator, f is the operating frequency, and C is the distribution capacitance. If we assume that the capacitance of the two clock networks is the same, which it should be for local distribution networks, and that both circuits are operating at $\omega_0 = 2\pi f$, then using the fact that $Q = 2\pi f_0 RC$, we find that the power ratio of these two approaches is $P_{ratio} = \pi/Q$. Consequently, any Q value larger than π will result in power savings. This is an idealized approach since resistance in the inductor and capacitor will reduce the Q of the circuit and modify the power dissipation, but for Q larger than 5, these will be ignored [9]. The standard clock buffer is also an idealized situation since the dynamic power equation does not directly account for power dissipated to maintain edge rates and skew so this will offset the non-ideal conditions of the resonant circuit.

Moreover, the capacitive load will consist of the distributed gate capacitances of the functional unit implemented. This problem is due to the resistance in the metal, vias, and contacts and the inductance. The wires and gates must appear mostly capacitive to be effective and provide a high Q value.

VI. CONCLUSIONS

In general terms, high-performance design with a basic CMOS input/output and receiver design is a main part of chip and system design. An actual CMOS receiver design consists of impedance matching and slew rate control, mixed supply voltages including level shifters, electrostatic design and other constraints.

The *clock generation and distribution* is a very effective design for low power and high performance circuits and system. It is applied to an oscillation-based test circuit to implement a low-cost and comprehensive low power methodology for fast logic circuits, which allows a tolerance margin in high-performance system design. The research is also focusing on new low power clock design methodology that helps engineers predict the best way to design computer chips in power sensitive ways.

참고문헌

[1] J. M. Rabaey, "Low-Power Silicon Architecture for Wireless Communications," *ISSCC'94: IEEE International Solid-State Circuits Conference*, Washington, 1994

[2] S. M. Kang,, Y. Leblebici, "CMOS Digital Integrated Circuits-Analysis and Design," 2nd Edition *McGraw-Hill International Edition*. 1999 USA.

[3] E. W. Dobberpuhl, "A 200-MHz 64-b Dual-Issue CMOS microprocessor," *IEEE Journal of Solid-State*

Circuits, vol. 27, no. 11, pp. 1115-1567, Nov 1992

[4] P. Hofstee, N. Aoki, et. al, "A 1.0 GHz Single-Issue 64-Bit PowerPC Processor," *IEEE International Solid-State Circuits Conference*, pp. 92-93, 2000 *Proc. ISSCC*, pp. 30-31, 186-187, Feb 1998., pp. 230-231, 1998., vol. 33, pp. 1600-1608, 1998

[5] W. C. Athas, and L. Svensson, "Reversible Logic Issues in Adiabatic CMOS," *Proc. Workshop on Physics and Computation*, pp. 111-118, Nov 1994

[6] K. J. Nowka, T. Galambos, "Circuits design techniques for a Gigahertz Integer Microprocessor," *International Conference on Computer Design*, pp. 11-16, Oct 1998.

[7] W. C. Athas, et. al., "A Low-Power Microprocessor Based on Resonant Energy," *IEEE J. Solid-State Circuits*, vol. 32, no. 11, pp. 1693-1701, Nov 1997

[8] B. A. Floyd, and K. K. O, "The projected power consumption of a wireless clock distribution system and comparison to conventional distribution systems," *IEEE Int. Conf. Interconnect Technology*, pp. 248-250 May 1999

[9] T. Jeong, A. Ambler, "Power Efficiency System for Fligth Application (PESFA) Mission: Low Power Dissipation in Digital Circuit Design for Flgith Application/Space Communications," *IEEE Trans. Aerospace and Electrical Systems,* vol 42, 2006

[10] A. Vittal, "Low-power buffered clock tree design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol, 16, no. 9, Sep 1997

저자소개

Taikyeong T. Jeong received the Ph.D. degree from the Department of Electrical and Computer Engineering, the University of Texas at Austin in 2004. He performed research in the area of high performance circuit design and power efficiency system design. He joined the Myongji University, where he is now an assistant professor, working on VLSI design for next generation space robotics devices and high performance circuit and system. His research interests include VLSI design, computer architecture, SoC for communication and network applications, and high performance system design.

Jang Ho Lee received the Ph.D. degree in computer science from the University of Michigan, Ann Arbor in 2000. He worked on collaborative systems at IBM T. J. Watson Research Center, NY as a postdoctoral researcher in 2000. He is currently an assistant professor at the Department of Computer Engineering at Hongik University, Korea. His research interests include computer-supported cooperative work, grid computing, and distributed systems.