

Low-power Design Methodology and Applications utilizing Dual Supply Voltages

Kimiyoshi Usami and Mutsunori Igarashi

Design Methodology Department
System LSI Design Division
Toshiba Corporation, Semiconductor Company
580-1, Horikawa-cho, Saiwai-ku,
Kawasaki 210-8520, JAPAN
Tel : +81-44-548-2346
Fax : +81-44-548-8318
e-mail: kimiyoshi.usami@toshiba.co.jp

Abstract - This paper describes a gate-level power minimization methodology using dual supply voltages. Gates and flip-flops off the critical paths are made to operate at the reduced supply voltage to save power. Core technologies are dual- V_{DD} circuit synthesis and P&R. We give a brief overview on existing low-power EDA technologies as background and discuss advantages and challenges of the dual-VDD approach. Through real design examples, we will show that the approach reduces power effectively while keeping the performance at negligible area overhead.

I. Introduction

Power minimization is one of the key technologies in modern VLSI design. Research and development in this field are motivated by growing markets of portable information devices, such as PDA's, cellular phones, digital camcorders and digital cameras. For VLSI's in those applications, high performance and low power are required simultaneously. High performance is required for data processing such as real-time encoding and decoding of picture/audio data, while low power for prolonging battery life.

Requirement for power minimization is not limited to portable applications. In consumer applications in which cost competition is extremely serious, low power is needed for reducing package cost. Power must be low if one uses cheap plastic packages instead of expensive ceramic ones.

Power minimization is also required for improving reliability of high-end microprocessors. Operation frequencies are raised for improving performance in those VLSI's, resulting in increasing power (current) dissipation. This tends to cause reliability problems such as electromigration in wires, hot electron effects in MOS transistors, IR-drop in power lines and ground bounce. Power reduction technologies keeping high performance are strongly required.

In this paper, we present design methodology that enables us to reduce power while keeping performance. First we will

give a brief overview on low-power EDA technologies as background. Next we will describe low-power techniques using multiple supply voltages in a core, especially focusing on a gate-level design methodology using dual- V_{DD} 's. We will discuss effectiveness and challenges of the methodology through experimental results at real chips.

II. Background

Average power dissipation in CMOS circuit consists of four components; dynamic power consumed in charging and discharging of load capacitances, short-circuit power arising from switching transient current, leakage power due to leakage current, and static power due to static current drawn continuously from the power supply [1]. Dominant fraction of average power is attributed to the dynamic power [2]. In this paper, we discuss low-power design methodology by focusing on the techniques of minimizing dynamic power.

The dynamic power is given by $P = CV^2 f \alpha$, where C is the load capacitance, V is the supply voltage, f is the clock frequency, and α is the switching activity. Hence, for minimizing the dynamic power, we need to minimize the parameters of C , V , f and α . Papers have been reported on the technologies that minimize these parameters while keeping the performance. Fig.1 summarizes power minimization technologies, in which parameters to be minimized and design levels are shown two-dimensionally. Currently the most popular techniques are the ones that minimize C and α . These techniques can be classified into two categories: (i) techniques which minimize $C \times \alpha$ on average and (ii) techniques which stop switching selectively. Techniques such as low-power logic synthesis, power-driven placement and transistor sizing are examples of (i), while techniques such as gated clock and operand isolation are those of (ii). Low-power technology mapping [3] is an effective technique in low-power logic synthesis. Cells are mapped so nodes with high switching activity be hidden inside the cells.

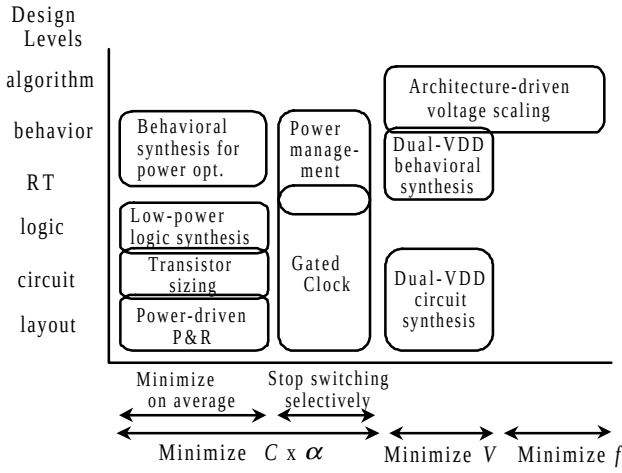


Fig.1. Power Minimization Technologies.

Load capacitances inside the cells are smaller, resulting in minimizing $C \times \alpha$ on average. Power-driven placement is a layout technique in which placement is performed so wire length weighted by switching activity be minimized [4]. As transistor-sizing techniques for low power, gate resizing [5] and gate/wire sizing [6] have been reported. Techniques categorized in (ii) focus on disabling signal transitions when they are unnecessary. Gated clock is the most well-known technique in which switching of clock signals is stopped when register-loading is not needed. Operand isolation is an approach in which transitions of inputs are disabled at circuit blocks when they are idle [7].

Another parameter to be targeted at for power optimization is the supply voltage V . Reduction of supply voltage is very effective because power is reduced quadratically. However, the performance of MOS transistors is degraded with reducing V , leading to increasing the delay. Techniques are required which enable us to reduce supply voltage while keeping the performance of the entire chip. Architecture-driven voltage scaling [8] is an approach to save power by reducing V and f . A parallel architecture is built for keeping the throughput of the entire circuit. This approach is quite effective at DSP applications, although it comes with area penalty resulting from parallel datapaths.

Another approach is to make use of dual supply voltages in a core. Only the circuit part requiring high performance is made to operate at the original voltage V_{DDH} , while the rest of the circuit is made to operate at the reduced voltage V_{DDL} . This results in reducing power without degrading the entire circuit performance. This approach can be classified by the level of granularity into (i) module level, (ii) functional-unit level, and (iii) gate level. In a module-level approach reported in [9], V_{DDL} is used in potentially power-hungry modules, while V_{DDH} is used in other ones. A designer must perform RTL design and logic synthesis taking into account the delay at V_{DDL} . As approaches for (ii), papers have been reported on scheduling and optimal voltage selection for operations in behavioral synthesis [10][11]. A gate-level approach reported in [12] is to synthesize dual- V_{DD} structure at gate level from the netlist originally designed with single supply volt-

age. No extra burden is imposed on a designer. He/she can perform RTL design and logic synthesis in the conventional way. Implementation with dual supply voltages is automatically done using dual- V_{DD} circuit synthesis and P&R [13].

III. Gate-level Dual- V_{DD} Approach for Low Power

A. Basic Concept and Problem Definition

The basic idea of the gate-level dual- V_{DD} approach is to run the gates off the critical paths at the reduced supply voltage V_{DDL} , while those on the critical paths at V_{DDH} . This results in reducing the power without degrading the entire circuit performance. This idea can be realized by performing static timing analysis for a given circuit designed with single V_{DD} , extracting the gates off the critical paths, and assigning V_{DDL} to them. If netlist and timing constraints are given along with the V_{DDL} voltage, one could generate a dual- V_{DD} circuit using conventional algorithm such as gate resizing. Fig.2 shows a possible circuit example generated in this manner. However, this circuit has a problem when implementing in CMOS LSI's. Static current flows at a V_{DDH} gate if it is directly driven by a V_{DDL} gate. This is because the PMOS at the V_{DD} gate cannot be cut-off at the input level "High". A typical way of blocking the static current is to insert a level converter (level shifter) circuit into the interface of a V_{DDL} gate to a V_{DDH} gate. In Fig.2, level converters should be inserted into the portions marked as A to D. Level converter become an overhead of area and power. Thus the problem we need to solve is finally defined as follows: For a given circuit, choose gates and/or flip-flops to which we should apply V_{DDL} such that we minimize the number of level converters and the entire power while meeting the timing constraints.

B. Dual- V_{DD} Circuit Synthesis

Minimizing connections from V_{DDL} gates to V_{DDH} ones is a most effective way to minimize the number of level converters needed. The Clustered-Voltage-Scaling (CVS) structure proposed in [12] is the simplest one that does it. As

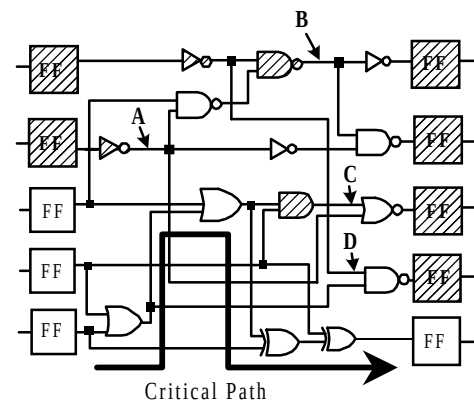


Fig.2. Circuit with portions operating at V_{DDL} (shown as shaded) and those operating at V_{DDH} .

shown in Fig. 3, clustered structure is built such that logic gates on the output side of the combinational network (i.e. logic gates closer to the D-inputs of flip-flops) become V_{DDL} gates. In the CVS structure the connection pattern of a V_{DDL} gate to a V_{DDH} gate does not exist. The only portions requiring level converters are the interface between the output of a V_{DDL} flip-flop and the input of a V_{DDH} gate. Therefore, the number of needed level converters is at most the same as the number of flip-flops. Because no level converters are needed between logic gates, the CVS structure minimizes the number of level converters needed.

An extended CVS (ECVS) structure has been reported in [14]. The limitation in the CVS is somewhat relaxed so the number of gates becoming V_{DDL} be increased. In the ECVS, insertion of level converters is allowed even between logic gates. This is effective at a gate with multiple inputs when one input is on the critical path, while others are not. In the CVS, not only this gate but also its transitive fan-ins must be V_{DDH} . In the ECVS, however, we give opportunities to the transitive fan-ins to become V_{DDL} by inserting level converters at the input of the gate. In practice, it is checked if the level-converter insertion is worth performing in reducing power. If it is, the insertion is performed.

Because the CVS or ECVS structure could not be synthesized with available EDA tools, we developed a dual- V_{DD} circuit synthesizer that generates them. The synthesizer takes a gate-level netlist designed with a single supply voltage V_{DDH} as input, identifies non-critical paths by invoking an internal static-timing-analysis engine, and replaces as many V_{DDH} cells with V_{DDL} ones under the given timing constraints. Level converters are automatically inserted into needed portions. A netlist with the CVS or ECVS structure is generated as output.

C. Dual- V_{DD} Place and Route

One of the most important issues in the dual- V_{DD} layout is the voltage separation between V_{DDH} cells and V_{DDL} cells. V_{DDH} and V_{DDL} cells have to be isolated because their N-well voltages are different. Voltage separation at row-by-row has been addressed in [14] [15], while the voltage separation within a row has been presented in [16]. We chose the row-by-row voltage separation shown in Fig. 4 because of (i)

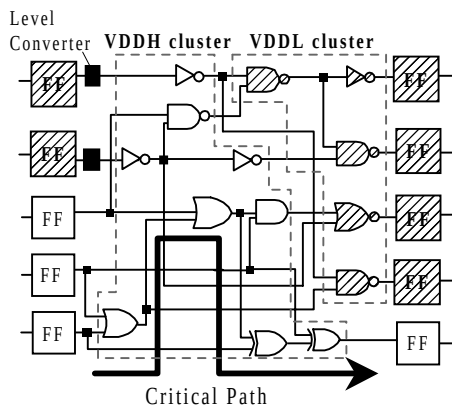


Fig. 3. Clustered-voltage-scaling (CVS) structure.

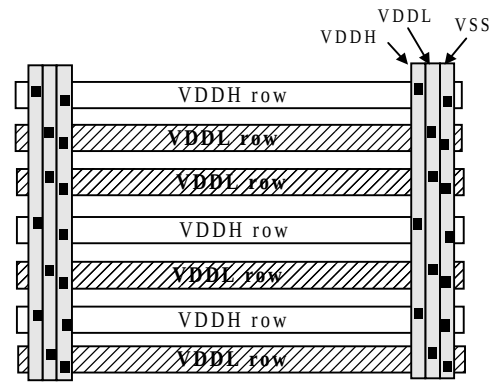


Fig. 4. Row-by-row voltage separation for dual- V_{DD} layout.

high-performance and (ii) applicability to gate arrays as well as standard-cell-based designs.

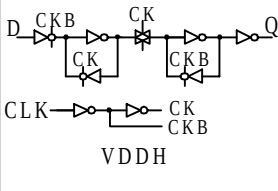
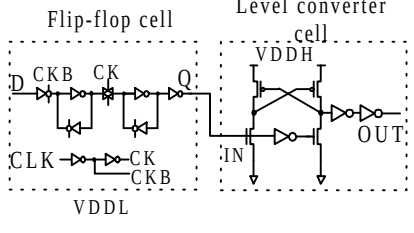
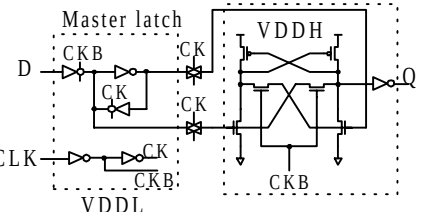
We developed automated techniques to generate the row by-row dual- V_{DD} layout and realized on our internal P&R platform. A couple of new technologies were required. One is the optimal voltage assignment to rows. We developed voltage assignment algorithms minimizing area overhead while taking into account wire congestion and timing constraints [15][17]. The other key technology is the cell placement under the voltage constraint. In the row-by-row architecture a V_{DD} cell must be placed in a V_{DDH} row, while a V_{DDL} cell in a V_{DDL} row. In other words, cells must be placed taking into account the voltage. Since this could not be achieved with available P&R tools, we enhanced the capability of our internal placement tool to achieve this.

D. Flip-Flop Circuit with Level-Conversion Function

The CVS and ECVS structure described above minimizes the number of level converters. The number is at most the number of flip-flops because a level converter is basically inserted at the output of a flip-flop. However, in the case of a circuit with a lot of flip-flops, inserted level converters cause a significant area penalty. We proposed an approach reducing the penalty by using a flip-flop circuit with a level-conversion function (FFLC) [17]. The circuit is depicted at the right-most column in Table 1. A master latch of the FFLC is the same as a conventional flip-flop, while a slave latch has a level-conversion function. Signals with V_{DDL} -swing at D and CLK are converted to V_{DDH} -swing at Q. Area penalty gets less by using FFLC compared to the case of inserting a level-converter cell at the output of a flip-flop (shown in column 2). Power is smaller than the conventional flip-flop operating at V_{DD} , resulting from the fact that the master latch of the FFLC operates at V_{DDL} . CLK-to-Q delay of the FFLC is larger only by 0.18ns than that of the conventional flip-flop. We chose to use the FFLC in the design because of its advantages.

The use of the FFLC allows us to shorten the turn-around-time (TAT) of the design. We developed a design flow to use the FFLC as a default flip-flop in logic synthesis. Register elements in RTL description are mapped into the FFLC' even in the single-voltage netlist. The use of the FFLC is

Table 1. Comparisons on flip-flops.

	Conventional flip-flop	Insertion of level-converter cell at output of flip-flop	Flip-flop with level-conversion function (FFLC)
Circuit structure			
Area *	1	1.64	1.21
Power *	1	0.74	0.63
Delay *	0.24 ns **	0.59 ns ***	0.42 ns **

* Results on area and power are normalized by those on the conventional flip-flop.
Power and delay were simulated at $V_{DDH} = 2.5V$, $V_{DDL} = 1.75V$, $|V_{th}| = 0.2V$.

** CLK-to-Q delay *** CLK-to-OUT delay

very helpful for minimizing the change between the single-voltage netlist and the dual- V_{DD} netlist. This is because there is no need to insert level converters at the outputs of the FFLC's, resulting in minimizing the change in gate count and path delay. This allows us to utilize the placement result of the single- V_{DD} layout as initial floorplanning information at dual- V_{DD} layout. This leads to shortening the TAT.

E. Supply Voltage Reduction in Clock Tree

We applied the dual- V_{DD} approach also to a clock tree [13]. A root driver is made to operate at V_{DDH} , while clock buffers in the clock tree are made to operate at V_{DD} . Since the output voltage swing of the leaf buffers is V_{DDL} , flip-flops driven by the leaf buffers must be able to receive the clock signal with the V_{DDL} swing. By using the FFLC circuits as flip-flops, this constraint is automatically satisfied. The dual- V_{DD} clock tree is generated at the layout phase using a Clock-Tree-Synthesis (CTS) technique.

IV. Application Examples of Dual- V_{DD} Approach

A. Media Processor

We applied the dual- V_{DD} circuit synthesis and dual- V_{DD} P&R to a media processor [13].

In applying this approach, the first thing we had to do was to find the optimal V_{DDL} voltage. We repeated dual- V_{DD} circuit synthesis and monitoring the power while changing V_{DDL} voltage under V_{DDH} of 2.7V. The result is shown in Fig.5. The power versus V_{DDL} curve is U-shaped, showing that the power gets the minimum at V_{DDL} of 1.9V. Intuitively we can see that the power can not be reduced at higher V_{DDL} . However, the power is not reduced either when V_{DDL} is too low. The reason is as follows: In the dual- V_{DD} approach, power reduction is determined by a couple of factors: the power reduction at a single cell by scaling the supply voltage from V_{DDH} to V_{DDL} , and the number of cells replaced with V_{DDL} ones. At lower V_{DDL} , the power at a V_{DDL} cell gets

smaller, while the number of the cells replaced with V_{DDL} ones gets fewer. This is because the performance of V_{DDL} cells gets poorer at lower V_{DDL} . In other words, to replace cells with V_{DDL} ones gets harder while meeting the timing constraints. Hence the power is not reduced when V_{DDL} is too low.

Determining the optimal V_{DDL} voltage is an important step in the dual- V_{DD} design. However, if we need to repeat the procedure described above to find the optimal V_{DDL} at each design, it would take a lot of time. Instead, if there is an approach to give us the optimal V_{DD} in a top-down fashion, it will be more efficient. The research on this issue has been reported in [18]. The authors built a theoretical model relating the V_{DDL} / V_{DD} value and the path-delay distribution of circuits to the power reduction. By using the model, they analyzed the V_{DDL} / V_{DD} value maximizing the power reduction. They showed that the optimal V_{DDL} / V_{DDH} value exists between 0.6 and 0.7 irrespective of the path-delay distribution of a circuit. In the application to the media processor, the optimal V_{DDL} was empirically obtained as 1.9V under V_{DDH} of 2.7V, leading to $V_{DDL} / V_{DDH} = 0.70$. This shows a good agreement with the claim from the theory.

We performed the dual- V_{DD} circuit synthesis to random-logic modules including 2.5K through 14K cells. The per-

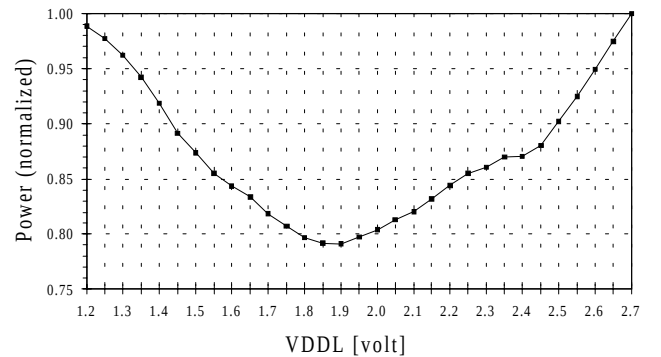


Fig.5. Power-VDDL curve resulting from the dual-VDD circuit synthesis.

centage of the cells replaced with V_{DDL} ones was 76%. To investigate the reason of the rich replacement, we analyzed the distribution of path-delays in the design. Results are shown in Fig.6. The X-axis shows the path-delay normalized by the cycle-time. Hence the normalized delay of 1.0 indicates the path-delay is the same as the cycle-time, meaning it is a critical path. Moving toward the left along the X-axis, the path-delay decreases and the slack increases. The Y-axis shows the number of paths normalized by the total number. We investigated 5.8M paths using static path analysis. In the original design, although as many as 15K critical paths exist, they occupy only 0.3% in the total paths. More than 60% of the paths have delays of half of the cycle-time. By applying the dual- V_{DD} circuit synthesis, the center of the distribution was moved toward the right. Thus the dual- V_{DD} circuit synthesis effectively spends the excessive slack remaining in the non-critical paths to build the dual- V_{DD} structure.

We analyzed the power using the PowerMill simulation. Power was reduced by 47% in the random-logic modules while keeping the performance of 75MHz.

B. Graphic Controller LSI

We applied the dual- V_{DD} circuit synthesis to three function modules in a graphic controller LSI [19]. Power reduction was compared with that of the conventional gate-resizing technique. We investigated how the power reduction is affected by the wire capacitance. We performed dual- V_{DD} circuit synthesis and gate-resizing while changing the wire capacitance. Results are shown in Fig.7. As the wire capacitance gets dominant, power reduction in the gate-resizing significantly decreases, while that in the dual- V_{DD} circuit synthesis does not change so much.

Possible reasons are as follows: when the wire capacitance is not dominant over the gate capacitance, each gate could be down-sized without increasing the total delay. In other words, down-sizing a certain gate reduces the output load capacitance of the parent gate, allowing to down-size the parent gate itself without increasing the delay. This down-sizing chain results in reducing the total power. However, when the wire capacitance is dominant, the down-sizing chain becomes less possible because the down-sizing

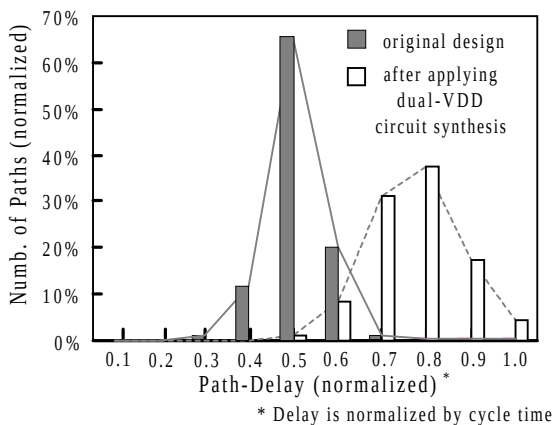


Fig.6. Distribution of path-delays

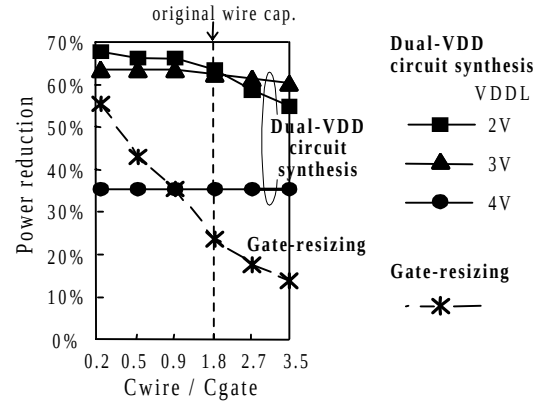


Fig.7. Affect of wire capacitance on power reduction at dual-VDD circuit synthesis and gate resizing

of a gate affects less to the load capacitance. Another reason is that reducing gate capacitance does not contribute so much to the total power when wire capacitance is dominant.

Meanwhile, power reduction in the dual- V_{DD} approach does not depend upon whether the wire capacitance is dominant or not. This is because the dual- V_{DD} approach reduces the power dissipated in charging and discharging of wire capacitance as well as that for the gate capacitance.

Both of dual- V_{DD} circuit synthesis and gate-resizing are the techniques that reduce power by utilizing excessive slack in the circuit. However, they show very different sensitivity to the wire capacitance in power reduction. The dual- V_{DD} approach shows effectiveness in power savings even in the case that the wire capacitance is dominant.

C. MPEG4 Codec Core

To an MPEG4 codec core, we applied both the dual- V_{DD} approach and a VTCMOS technique to reduce power. VTCMOS is a technique to control the effective threshold voltage by applying substrate bias to MOS transistors [20]. At the active mode, transistors are made to operate at low V_{DD} with low V_{th} . At the standby mode, the effective threshold voltage is made to be larger by applying substrate bias to block the leakage current. Hence the VTCMOS technique allows us to reduce supply voltage at the active mode at the entire chip while keeping the performance.

Meanwhile, since VTCMOS does not care about excessive slack in non-critical paths, there is room for further power reduction at the gates in non-critical paths. By applying the dual- V_{DD} approach, we further reduce the supply voltage in non-critical paths to reduce power.

We analyzed the power of an MPEG4 codec core designed in 0.3 μ m CMOS technology. Results from the power simulation using PowerMill are shown in Fig.8. Operation frequency is 30MHz. VTCMOS using $V_{DD}=2.5V$ and $V_{th}=0.2V$ reduced power by 42% compared to the conventional design using $V_{DD}=3.3V$ and $V_{th}=0.55V$. In addition, applying the dual- V_{DD} approach with $V_{DDL}=1.75V$ reduced power further by 25% at the entire chip. We applied the dual- V_{DD} approach to the portions except memory macros.

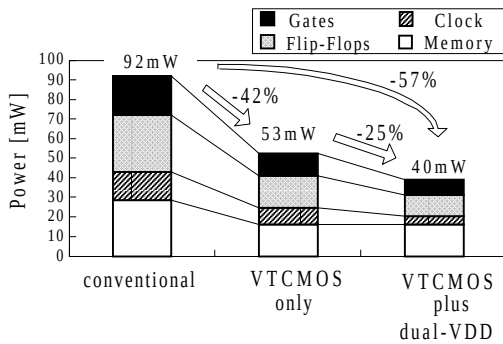


Fig.8. Power reduction resulting from VTCMOS and dual-VDD approach at MPEG4 codec core

Power was reduced at logic gates by 30%, at flip-flops by 37%, and at the clock tree by 49%, respectively. At the portions to which we applied the dual- V_{DD} approach, power was reduced by 35% even after applying VTCMOS. Thus the power in the conventional design was reduced by 57% by combining both techniques together.

To verify at real silicon, we fabricated test chips and measured average power dissipation. The chip to which we applied both the dual- V_{DD} approach and VTCMOS dissipated 45mW, leading to power reduction by 58% from the conventional design. The number of power reduction shows a good agreement with the above-mentioned results from the power simulation. The area overhead was 5% at the portion to which we applied the dual- V_{DD} approach, being almost negligible at the entire chip.

V. Conclusions and Future Work

We have described power minimization techniques reducing load capacitance, supply voltage, clock frequency, and switching activity. Currently the most popular techniques are those to reduce capacitance and switching activity. Techniques to reduce voltage, however, are very attractive because power dissipation is reduced quadratically. The dual- V_{DD} approach using two different supply voltages in a core enables us to save power while keeping the entire performance. We presented dual- V_{DD} circuit synthesis, P&R techniques and dual- V_{DD} clock tree generation. Through application examples, we showed the effectiveness of the dual- V_{DD} approach.

For future work, research efforts will be needed on noise analysis and avoidance scheme for the design with dual- V_{DD} . In particular, coupling noise should be carefully analyzed where an interconnect driven by V_{DDL} is located between two interconnects driven by V_{DDH} . We performed a design by paying special attention to inter-module wires and a clock net with V_{DDL} swing [13]. Research on systematic noise analysis and avoidance methodology will be required.

References

- [1] L. Benini and G. De Micheli, *Dynamic Power Management*, Kluwer, 1998.
- [2] A. Chandrakasan and R. Brodersen, *Low power digital CMOS design*, Kluwer, 1995.
- [3] C-Y. Tsui, M. Pedram, and A.M. Despain, "Technology decomposition and mapping targeting low power dissipation", *Proc. 30th Design Automation Conference*, 68-73, 1993.
- [4] H. Vaishnav and M. Pedram "Pcube: A performance driven placement algorithm for low power designs", *Proc. Europe. Design Automation Conference*, Sept. 1993.
- [5] R.I. Bahar, G.D. Hachtel, E. Macii, and F. Somenzi, "A symbolic method to reduce power consumption of circuits containing false paths", *Proc. Int. Conference on Computer Aided Design*, 368-371, 1994.
- [6] J. Cong and C-K Koh, "Simultaneous driver and wire sizing for performance and power optimization", *IEEE Trans. VLSI Systems*, vol.2, pp.4, 408-423, Dec. 1994.
- [7] E. Musoll and J. Cortadella, "High-level synthesis techniques for reducing the activity of functional units", *Proc. Int. Symposium on Low Power Design*, 99-104, Apr. 1995.
- [8] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design", *IEEE J. Solid-State Circuits*, vol.27, no.4, pp.473-484, Apr. 1992.
- [9] M. Mizuno, et al, "Adaptive search-window motion estimation for MPEG2 encoder LSIs", *Technical Report of IEICE (in Japanese)*, DSP97-109, pp.33-39, 1997.
- [10] J.M. Chang and M. Pedram, "Energy minimization using multiple supply voltages", *Proc. Int. Symposium on Low Power Electronics and Design*, 157-162, 1999.
- [11] M. Johnson and K. Roy, "Optimal selection of supply voltages and level conversions during data path scheduling under resource constraints", *Proc. Int. Conference on Computer Design*, 72-77, 1999.
- [12] K. Usami and M. Horowitz, "Clustered voltage scaling technique for low-power design", *Proc. Int. Symposium on Low Power Design*, 3-8, April 1995.
- [13] K. Usami, et al, "Automated low-power technique exploiting multiple supply voltages applied to a media processor", *IEEE J. Solid-State Circuits*, vol.33, no.3, pp.463-472, 1998.
- [14] K. Usami, et al, "Automated low-power technique exploiting multiple supply voltages applied to a media processor", *Proc. Custom Integrated Circuits Conference*, 131-134, May 1997.
- [15] M. Igarashi, et al, "A low-power design method using multiple supply voltages", *Proc. Int. Symposium on Low Power Electronics and Design*, 36-41, 1997.
- [16] C. Yeh, Y. Kang, S. Shieh, and J. Wang, "Layout techniques supporting the use of dual supply voltages for cell-based designs", *Proc. 3rd Design Automation Conference*, 62-67, 1999.
- [17] K. Usami, et al, "Design methodology of ultra low-power MPEG4 codec core exploiting voltage scaling techniques", *Proc. 35th Design Automation Conference*, 29.1, 1998.
- [18] M. Hamada, et al, "A top-down low power design technique using clustered voltage scaling with variable supply-voltage scheme", *Proc. Custom Integrated Circuits Conference*, 495-498, 1998.
- [19] K. Usami, T. Ishikawa, M. Kanazawa, and H. Kotani, "Low-power design technique for ASICs by partially reducing supply voltage", *Proc. IEEE ASIC Conference*, 301-304, Sept. 1996.
- [20] T. Kuroda, et al, "A high-speed low-power 0.3 μ m CMOS gate-array with variable threshold voltage (VT) scheme", *Proc. Custom Integrated Circuits Conference*, 53-56, 1996.