Multi-Layer Adaptive Power Management Architecture for TSV 3DIC Applications

Ming-Hung Chang1,2, Wei-Chih Hsieh1, Pei-Chen Wu1, Ching-Te Chuang3, Kuan-Neng Chen1, Chen-Chao Wang2, Chun-Yen Ting2, Kuo-Hua Chen1, Chi-Tsung Chiu2, Ho-Ming Tong2, Wei Hwang1,2
Department of Electronics Engineering & Institute of Electronics1, National Chiao Tung University, Hsinchu, Taiwan
Advanced Semiconductor Engineering (ASE) Group2, Kaohsiung, Taiwan

Abstract
In this work, a multi-layer hierarchical distributed power delivery architecture for TSV 3DIC is proposed. By decoupling global and local power networks, the proposed power delivery architecture can be flexibly configured for different power requests. The decoupled power architectures can also greatly reduce the required decoupling capacitor sizes for voltage stabilization. Meanwhile, a multi-threshold CMOS switched capacitor DC-DC converter with up to 78% power efficiency is implemented in 65nm CMOS for hierarchical distributed power delivery architecture. An adaptive power management technique is presented to work in the local power network to increase the power efficiency. The proposed multi-layer hierarchical distributed power delivery architecture is also very useful for the heterogeneous integration in 3DIC chips.

Introduction
Chip performances have become interconnect-dominated in two-dimensional (2D) planar ICs as a result of device scaling driven by Moore's law in the past few decades. System-on-a-Chip (SoC) technology which is a highly integrated 2D solution for system scaling further exacerbates the long interconnect issue. The communication between logic and memory units residing at opposite ends of the chip suffers a long latency even though they are both on-chip. In addition to long interconnect issue, the SoC solution demands the designer to use one single process to fabricate digital, analog, and memory circuits at the same time. However, the chosen process cannot be the best choice for every design style. The increased design complexity to keep performance may offset the benefit of SoC integration.

Nowadays, three-dimensional IC (3DIC) technologies [1]-[3] are gaining rising interests as they become technologically viable and provide attractive alternatives to 2D planar ICs. Among various 3D interconnect methods, through-silicon via (TSV) technology has the potential to achieve the greatest interconnect density but also the greatest cost [3]. Unlike planar SoC integration, 3DIC technique allows heterogeneous integration of different fabrication processes. Different portions of the system can be fabricated with their most suitable process technologies and then combined within a single 3DIC chip. Memory stacked with logic circuits is a typical example of 3D system integration [4]-[6]. In addition to the heterogeneous integration, the long cross-chip interconnect latencies are also replaced by much shorter vertical interconnects between strata in the 3DIC chip. It was suggested that the average wire length will drop by a factor of $N^{1/2}$ whereas wire RC delay drops by a factor of $N$. In other words, system performance is enhanced.

Power delivery is another challenge in 3DIC technology. The power density per square area increases by a factor of $N^{1/2}$ even though the total power consumption is reduced [9]. So the power delivery requirements will increase with the number of stacked strata in 3DIC chips. Many researches had worked on the quality of the power delivery [9]-[11]. The analysis showed that the 3D stacking has a higher impact on IR drop noise [9]. A voltage high-to-low conversion near (or integrated into) the chip using a switching DC-DC converter is suggested [10] to reduce the input current from off-chip supply. As a result, the off-chip resources to maintain power integrity can therefore be reduced. A multi-story power delivery technique was also presented to employ the charge recycling concept in the 3DIC power network [11].

Previous works typically assumed a single supply voltage across the entire 3DIC chip stacks. However, the most attractive aspect of the 3DIC chip is the heterogeneous integration. The best scenario is that not only the fabrication processes but also the operating supply voltages can be chosen for different parts of the system to have the best trade-off between cost, power, and performance. A single supply voltage apparently cannot meet this demand. Therefore, a new concept of multi-layer hierarchical distributed power delivery architecture for TSV 3DIC is proposed in this paper. An adaptive power management technique is also presented for processing cores. The rest of this paper is organized as following. Section II introduces the proposed concept of multi-layer hierarchical distributed power delivery architecture. In Section III, a multi-threshold CMOS switched capacitor DC-DC converter is proposed. The adaptive power management technique is presented in Section IV. And the conclusions are summarized in Section V.

Power Delivery Architecture for TSV 3DIC
An example of die stacking in a 3DIC [9] is shown in Fig. 1. Practically the die with highest heat generation, i.e., highest power, is placed nearest to the heat sink. Other dies then organized between the top stratum and the package substrate considering connection and function relation. The power supply and the signals are transmitted proportionally. As a result, power drops by a factor of $N^{1/2}$ whereas wire RC delay drops by a factor of $N$. In other words, system performance is enhanced.

However, the benefits of going vertical 3D integration don't come at no cost. Heat removal is one of the major reliability issues in 3DIC [3], [7], [8]. A recent work had incorporated the microfluidic cooling technique in 2D and 3DIC chips to remove heat [7]. Routing analysis considering signal, power, and thermal distribution at the same time had also been presented [8].

Power delivery is another challenge in 3DIC technology. The power density per square area increases by a factor of $N^{1/2}$ even though the total power consumption is reduced [9]. So the power delivery requirements will increase with the number of stacked strata in 3DIC chips. Many researches had worked on the quality of the power delivery [9]-[11]. The analysis showed that the 3D stacking has a higher impact on IR drop noise [9]. A voltage high-to-low conversion near (or integrated into) the chip using a switching DC-DC converter is suggested [10] to reduce the input current from off-chip supply. As a result, the off-chip resources to maintain power integrity can therefore be reduced. A multi-story power delivery technique was also presented to employ the charge recycling concept in the 3DIC power network [11].

Previous works typically assumed a single supply voltage across the entire 3DIC chip stacks. However, the most attractive aspect of the 3DIC chip is the heterogeneous integration. The best scenario is that not only the fabrication processes but also the operating supply voltages can be chosen for different parts of the system to have the best trade-off between cost, power, and performance. A single supply voltage apparently cannot meet this demand. Therefore, a new concept of multi-layer hierarchical distributed power delivery architecture for TSV 3DIC is proposed in this paper. An adaptive power management technique is also presented for processing cores. The rest of this paper is organized as following. Section II introduces the proposed concept of multi-layer hierarchical distributed power delivery architecture. In Section III, a multi-threshold CMOS switched capacitor DC-DC converter is proposed. The adaptive power management technique is presented in Section IV. And the conclusions are summarized in Section V.

Power Delivery Architecture for TSV 3DIC
An example of die stacking in a 3DIC [9] is shown in Fig. 1. Practically the die with highest heat generation, i.e., highest power, is placed nearest to the heat sink. Other dies then organized between the top stratum and the package substrate considering connection and function relation. The power supply and the signals are transmitted proportionally. As a result, power drops by a factor of $N^{1/2}$ whereas wire RC delay drops by a factor of $N$. In other words, system performance is enhanced.

However, the benefits of going vertical 3D integration don't come at no cost. Heat removal is one of the major reliability issues in 3DIC [2], [3], [7], [8]. A recent work had incorporated the microfluidic cooling technique in 2D and 3DIC chips to remove heat [7]. Routing analysis considering signal, power, and thermal distribution at the same time had also been presented [8].

Power delivery is another challenge in 3DIC technology. The power density per square area increases by a factor of $N^{1/2}$ even though the total power consumption is reduced [9]. So the power delivery requirements will increase with the number of stacked strata in 3DIC chips. Many researches had worked on the quality of the power delivery [9]-[11]. The analysis showed that the 3D stacking has a higher impact on IR drop noise [9]. A voltage high-to-low conversion near (or integrated into) the chip using a switching DC-DC converter is suggested [10] to reduce the input current from off-chip supply. As a result, the off-chip resources to maintain power integrity can therefore be reduced. A multi-story power delivery technique was also presented to employ the charge recycling concept in the 3DIC power network [11].

Previous works typically assumed a single supply voltage across the entire 3DIC chip stacks. However, the most attractive aspect of the 3DIC chip is the heterogeneous integration. The best scenario is that not only the fabrication processes but also the operating supply voltages can be chosen for different parts of the system to have the best trade-off between cost, power, and performance. A single supply voltage apparently cannot meet this demand. Therefore, a new concept of multi-layer hierarchical distributed power delivery architecture for TSV 3DIC is proposed in this paper. An adaptive power management technique is also presented for processing cores. The rest of this paper is organized as following. Section II introduces the proposed concept of multi-layer hierarchical distributed power delivery architecture. In Section III, a multi-threshold CMOS switched capacitor DC-DC converter is proposed. The adaptive power management technique is presented in Section IV. And the conclusions are summarized in Section V.
controlled collapse chip connection (C4) bumps connect the lowest stratum of the chip to the package substrate. Such an arrangement is necessary to comply with the thermal requirement of the chip.

It was suggested that the TSV architecture has small impact on Ldi/dt noise [9] whose major source is the off-chip passive components. But the IR drop noise will become worse as the number of the stacked strata increases because of additional resistive and capacitive path of TSV connections. Since the most power hungry die is located the furthest from the power source, it will suffer the most IR drop noise as well as Ldi/dt noise. Increasing the total number of TSV improves the IR drop noise at the cost of more blockage for device and metal layer placement. More decoupling capacitors are also required to suppress the Ldi/dt noise as load current increases. These will incur a lot of area penalties.

On the other hand, heterogeneous integration of 3DIC is another challenge for power delivery. Dies or blocks for different purposes could have their best power-performance trade-off at different supply voltages. It certainly cannot be satisfied with a single supply source from off-chip. It is also costly to have multiple (off-chip) supply sources and power networks in the 3DIC chip.

Therefore, a multi-layer hierarchical distributed power delivery architecture is proposed considering the issues above. A simple illustration is shown in Fig. 2. The global power and the local power networks are decoupled by linear voltage regulators. And each power domain is powered by a dedicated voltage regulator with the requested voltage. Note that the "layer" represents the conceptual location in the power hierarchy. For example, the global power network is the first layer and the voltage regulators are used because they have lower area consumption.

In Fig. 2, the global power network is connected to the off-chip power source and transmitted by TSVs to different strata. Note that the number and the location of TSV are flexible despite that only one in the center is shown in the figure. The IR drop noise exists as mentioned. Meanwhile, the inserted voltage regulator will induce a voltage drop. Therefore, the voltage source of the global power network should be slightly raised in the proposed architecture to endure these voltage drops. The cost of such a decoupled power architecture is the increased number of voltage regulators. However, the required decoupling capacitor to maintain a stable supply voltage can be greatly reduced.

The conventional power architecture connects TSVs to off-chip supply sources and directly supplies the load circuits by these TSVs. The fluctuation on TSVs are directly seen by the circuit. Extended analyses are performed based on Lin's and Yang's work [12], [13] on TSV 3DIC power integrity. The height and the diameter of the TSV are set at 50μm and 25μm, respectively. Each of power and ground has a 8x8 TSV matrix. Every stratum is assumed to have a ring oscillator to consumed 5A current. The worst effective supply voltage is measured at the farthest stratum from the off-chip power source.

**Figure 1.** A conceptual 3DIC stacking with TSV connection [9].

**Figure 2.** The multi-layer hierarchical distributed power delivery architecture.

**Figure 3.** Simulation results of effective supply voltage versus decap size.
Figure 4. Another configuration of the multi-layer hierarchical distributed power delivery architecture for heterogeneous integration.

source. It reports that to have a minimum 0.8V effective supply voltage from a 0.9V off-chip supply, a decoupling capacitor of 9nF is required in every stratum.

Fig. 3 shows the analysis when using the proposed decoupled power architecture. The right axis is the size of decoupling capacitor (decap) at TSV node. The left axis shows the total capacitance of the regulator, including a fixed 1nF internal capacitor and a variable output decoupling capacitor. The z-axis represents the acquired minimum effective supply voltage under different TSV and regulator decap combinations. There are two reference planes in the figure. The horizontal one indicates the required 0.8V minimum effective supply voltage whereas the vertical one shows the location of a total of 9nF decaps. Therefore, it can be easily interpreted whether the proposed decoupled power architecture is better.

The analysis in Fig. 3 first shows that the required 0.8V minimum effective supply voltage is easily met by inserting voltage regulators. It also shows that the decap size at the regulator output has a much larger effect on suppressing the noise whereas the size of the decap at TSV node is almost irrelevant. The 0.8V effective supply is achieved when the regulator decap is larger than 1.5nF (a total capacitance of 2.5nF) even when the TSV decap is zero. There is a 72.2% capacitance reduction. If only half of the supply fluctuation is allowed, i.e. 0.85 effective supply, the proposed decoupled power architecture can still has an over 30% reduction of decoupling capacitor.

It is also observed in the analysis that after inserting voltage regulators, the current flows through TSVs are smoother. The produced Ldi/dt noise is much smaller (less than 50mV depending on the size of decap) than original architecture since most of the inductance comes from off-chip components. Meanwhile, since the TSVs in the proposed decoupled power architecture does not supply circuits directly, the stability issue is relaxed. Therefore, the off-chip passive resources used to stabilize the global power supply can be greatly reduced as well. Note that there may be overdrives in the regulators for lower strata where the IR drop noise is relatively small. The high voltage tolerant circuit designs such as the ones in [10] or [14] can be used to prevent gate-oxide reliability issue.

The proposed power delivery architecture is very flexible as mentioned. A different configuration suitable for heterogeneous integration is shown in Fig. 4. A heterogeneous system usually requires different supply voltages for different function blocks, ranging from high (3.3V or higher) to ultra-low (sub-threshold operation) voltages. The multiple voltages requirement can be achieved by adopting the proposed multi-layer power delivery architecture. As shown in Fig. 4, the first layer power TSVs are connected to power source, supplying a high voltage. The voltage of the power source obeys the same rule that the degraded voltage level in the presence of voltage drops is still higher than requested. The clean high voltage is then provided to the high voltage domain through a voltage regulator.

Because of the inherent power efficiency limit, the linear regulator is not suitable for large voltage converting ratio. A switching DC-DC converter is a better option. The switching DC-DC buck converter can be positioned at the second layer of the power hierarchy as in Fig. 4 to produce a lower voltage for further usage. Reference [10] and [15] both suggested and demonstrated the fully integrated switching buck converters.

In summary, the proposed multi-layer hierarchical distributed power delivery architecture provides a flexible power delivery configuration depending on the applications. The load circuits and the power TSVs are decoupled by regulators (or switching converters). The analysis has shown that the decoupled power architecture reduced the requirement of the decoupling capacitors as well as off-chip passive resources. The inserted regulators are the major overhead. However, the reduction of decoupling capacitors compensates the overhead since the capacitor typically occupies large silicon areas. Moreover, the proposed power delivery architecture integrates power domain concept. It can support all kinds of power management methodologies in order to accomplish a power efficient design.

Multi-Threshold CMOS Switched Capacitor DC-DC Converter Design

Nowadays, ultra-low voltage operation becomes a major way to realize full energy saving for battery-operated or self-powered mobile applications. For such applications with a maximum 100μA of load current requirement, a switched capacitor (SC) DC-DC converter is essential as shown in the third layer of the power hierarchy in Fig. 4. In this work, a multi-threshold CMOS SC DC-DC converter with 0.2V to 0.4V output voltage capability is proposed.

The block diagram of our SC DC-DC converter is shown in Fig. 5. It uses a pulse frequency modulation (PFM) mode of control to regulate the output voltage. A delay-line-based digital comparator clocked by the signal CLK is generated.
corresponding signals for non-overlapping PFM generation. When the output voltage, $V_{OUT}$, is above $V_{REF}$, the switches are all set to $\Phi_1$ mode. It means $V_{BAT}$ starts to charge the capacitor. When $V_{OUT}$ falls below $V_{REF}$, the comparator triggers a $\Phi_2$ pulse, which charges up the output load capacitor, $C_{LOAD}$. The non-overlapping clock generator block prevents any overlap between the $\Phi_1$ and $\Phi_2$ ON phases. A PFM mode control is crucial to achieving high efficiency for the extremely low power system being built.

The external voltage input to the this stratum is set to be 0.5V. The gain setting at no-load provides a voltage ratio output of the input voltage. The switching losses in the converter are dominated by the energy expended in turning the charge transfer switches on and off. Thus, the switch widths are designed such that the charge transfer capacitors just settle at the end of a charge transfer cycle. In order to scale switching losses with load power, the charge transfer switches have adjustable widths which enabled by signals $V_{OP1}$/$V_{OP2}$. These two signals are generated by the proposed delay-line comparator. It can help to decrease the switching power, leading to an enhancement in efficiency at lower load current level.

The proposed delay-line-based digital comparator is implemented without any analog op-amp. It makes our comparator can be operated in near-/sub-threshold regions. With different voltage sources, higher supply voltage makes the delay line having faster speed than those with lower supply voltage. The proposed digital comparator composes of two stacking-inverter delay chain and two D-flip-flop. One of the two stacking-inverter delay line is connected to $V_{REF}$; while the other is connected to $V_{OUT}$. With CLK inserted at the same time, the upper delay line will be faster if $V_{REF}$ is greater

### Table I. Switched Capacitor DC-DC converter summary and comparison.

<table>
<thead>
<tr>
<th></th>
<th>2007 PESC [16]</th>
<th>2009 JSSC [17]</th>
<th>This Work (MOS Cap.)</th>
<th>This Work (MOM Cap.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>0.18μm</td>
<td>65nm</td>
<td>UMC 65nm</td>
<td>TSMC 65nm</td>
</tr>
<tr>
<td>Input Voltage</td>
<td>1.2V</td>
<td>0.5V</td>
<td>0.5V</td>
<td>0.5V</td>
</tr>
<tr>
<td>Output Voltage</td>
<td>0.3V to 1.1V</td>
<td>0.3V</td>
<td>0.2V to 0.4V</td>
<td>0.2V to 0.4V</td>
</tr>
<tr>
<td>Voltage Variation</td>
<td>8%</td>
<td>10%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>Max. Load Current</td>
<td>400μA</td>
<td>400μA</td>
<td>100μA</td>
<td>110μA</td>
</tr>
<tr>
<td>Power Efficiency</td>
<td>75%</td>
<td>75%</td>
<td>70%</td>
<td>78%</td>
</tr>
<tr>
<td>Switched Cap.</td>
<td>600pF</td>
<td>600pF</td>
<td>30pF</td>
<td>30pF</td>
</tr>
<tr>
<td>Response Time</td>
<td>540ps</td>
<td>288ps</td>
<td>500ns</td>
<td>450ns</td>
</tr>
<tr>
<td>Area (μm²)</td>
<td>750,000</td>
<td>120,000</td>
<td>7,200</td>
<td>21,280</td>
</tr>
</tbody>
</table>

Figure 5. Block diagram of the proposed multi-threshold CMOS switched capacitor DC-DC converter.
than $V_{OUT}$. Therefore, $V_{OP1}$ will be set to be "1". On the other hand, $V_{OP1}$ will be set to be "0" if $V_{OUT}$ is greater than $V_{REF}$.

For $V_{OP2}$ generation, the second D-flip-flop is inserted in the middle of the delay line. While $V_{REF}$ is much greater than $V_{OUT}$, the delay of longer inverter chain which is connected to $V_{REF}$ is still shorter than that connected to $V_{OUT}$, even though it passes more inverters. Therefore, the difference between $V_{REF}$ and $V_{OUT}$ is large when both $V_{OP1}$ and $V_{OP2}$ are set to be "1". In this situation, the switched capacitor matrix requires more driving ability to make the voltage transfer faster.

The efficiency of the proposed multi-threshold CMOS SC DC-DC converter is able to achieve no less than 60% over a load current range from 10μA to 100μA. The resolution of the delay-line-based digital comparator is 5mV. Because each technology node provides different kinds of capacitor, we implement this work in both TSMC (using MOM capacitor) and UMC (using MOS capacitor) 65nm standard bulk CMOS. The specifications of this work and comparison is shown in Table I.

Temperature Aware Adaptive Power Management Structure

In Section II, the power delivery architecture is discussed. The global power TSVs are separated from local power domains by voltage converters. The decoupled power delivery structure not only reduces the required passive resources for power stabilization but also facilitates the appliance of power management.

Each power domain is supplied by a dedicated voltage converter or regulator as in Fig. 2 or Fig. 4. The first requisition for voltage converters is that they should be suitable for on-chip integration such as the ones in [10], [15], [18]. The regulators are also required to be tunable in order to support dynamic voltage scaling. In such way, the power domains can be controlled by dynamic voltage and frequency scaling (DVFS) technique based on the power and performance requirements.

Although the DVFS technique can be directly applied for power management, the prime concern in 3DIC field has been shifted from power to heat. The power density per square area in a 3D stacked chip increases by a factor of $N^{1/2}$ [9] where as N is the number of the stacked strata in the chip. The increased power dissipation turns into heat. Typical heat removal devices can only be attached to the surfaces of 3DIC chip as the example shown in Fig. 1. As a result, the heat generated from other dies (die 2 and die 1 in Fig. 1) cannot be removed effectively because the thermal transfer coefficient of substrate materials is not high enough. The chip temperature will hence be increased to impact circuit performance.
A temperature aware power management concept is therefore proposed to limit the internal heat generation as shown in Fig. 6. The architecture contains different levels of power management techniques. First a DVFS controller is attached to the target power domain (process unit) for performance control based on system requirement. A variable output voltage regulator is a must to support internal voltage scaling. An adaptive power control block is inserted between the regulator and the process unit to have better power efficiency by utilizing unused slack [19]. Meanwhile, a temperature sensor [20] is added to monitor the circuit temperature. An arbiter is also inserted to give temperature sensor the priority over the DVFS controller.

Typically the DVFS controller defines the supply voltage and frequency according to performance or power request. However, the generated heat may impact the circuit reliability because of less efficient heat removal especially in inner strata of a 3DIC chip. The higher priority given to temperature sensor is to force the circuit to operate at lower voltage to lower the power consumption and heat generation. Meanwhile, the overall system controller should be designed to handle the performance degradation as a result of temperature control.

Since the proposed power delivery architecture separates the system into power domains, the power domains can be controlled individually. Fig. 7 shows an example block diagram of local power management structure. The figure also shows the flexibility of the proposed power delivery structure that more than one power domain can be arranged under the voltage regulator. The power domains can be further classified as always-on or not-always-on ones. Always-on blocks are supplied directly by the regulator whereas not-always-on ones are further controlled by an adaptive power control system [19] as illustrated in Fig. 7.

A test chip is implemented on UMC 65nm standard bulk CMOS technology to demonstrate the local power management structure. The floorplan is shown in Fig. 8. The whole chip is supplied by a variable output digital voltage regulator from [18]. There are three 64-bit multiplier blocks which are adaptive power controlled [19]. An always-on built-in-self-test (BIST) block including an 64-bit multiplier, pattern generation and results comparison logics is located in the center of the test chip. MOS capacitors are used for the internal decoupling capacitors as shown in Fig. 8.

The layout view of the implemented test chip is shown in Fig. 9(a) whereas the fabricated die photo is in Fig. 9(b). The simulation has verified the function of the proposed hierarchical adaptive power management structure. Three local sub-power domains are individually controlled that each one can be turned on or off based on the function demand. For each sub-power domain, an average of about 15% power reduction have been observed which is a result of adaptive power control.

**Conclusions**

A multi-layer hierarchical distributed power delivery architecture for TSV 3DIC is proposed in this work. By decoupling global and local power networks, the proposed power delivery architecture can be flexibly configured for different power requests. By adopting the proposed architecture, the required size of decoupling capacitor for voltage stabilization can be greatly reduced especially on the global power network. A delay-line-based digital comparator is used to implement the proposed multi-threshold CMOS switched capacitor (SC) DC-DC converter. Our SC DC-DC converter is capable of 0.2V to 0.4V output voltage with up to 78% power efficiency. An adaptive power management technique is presented to work in the local power network to increase the power efficiency. Meanwhile, the proposed multi-layer hierarchical distributed power delivery architecture is also very useful for the heterogeneous integration in 3DIC chips.

**References**


