J. Low Power Electron. Appl. **2013**, *3*, 250-266; doi:10.3390/jlpea3030250

OPEN ACCESS

Journal of
Low Power Electronics and Applications

ISSN 2079-9268

www.mdpi.com/journal/jlpea

Review

Exploration of Charge Recycling DC-DC Conversion Using a Switched Capacitor Regulator

Kaushik Mazumdar * and Mircea R. Stan

Charles. L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA; E-Mail: mircea@virginia.edu

***** Author to whom correspondence should be addressed; E-Mail: km3sj@virginia.edu; Tel.: +1-434-466-9375.

Received: 6 January 2013; in revised form: 5 July 2013 / Accepted: 9 July 2013 / Published: 29 July 2013

Abstract: The increasing popularity of DVFS (dynamic voltage frequency scaling) schemes for portable low power applications demands highly efficient on-chip DC-DC converters. The primary aim of this work is to enable increased efficiency of on-chip DC-DC conversion for near-threshold operation of multicore chips. The idea is to supply nominal (high) off-chip voltage to the cores which are then "voltage-stacked" to generate the near-threshold (low) voltages based on Kirchhoff's voltage law through charge recycling. However, the effectiveness of this implicit down-conversion is affected by the current imbalance among the cores. The paper presents a design methodology and optimization strategy for highly efficient charge recycling on-chip regulation using a push-pull switched capacitor (SC) circuit. A dual-boundary hysteretic feedback control circuit has been designed for stacked loads. A stacked-voltage domain with its self-regulation capability combined with a SC converter has shown average efficiency of 78%–93% for 2:1 down-conversion with I_{Load} (max) of 200 mA and workload imbalance varying from 0–100%.

Keywords: voltage stacking; charge recycling; DC-DC converter; switched capacitor; energy efficiency; near-threshold; hysteresis feedback; ripple; phase interleaving

---

1. Introduction

With the industry focusing more and more on SoC (system on chip) solutions for the low power portable applications, the role of efficient power management has become very crucial. An emerging trend for lowering energy is to scale the supply voltage V_{dd} to the near threshold region which brings not only quadratic dynamic energy savings, but also super-linearly reduced leakage currents. This shift in design trend, popularly known as near threshold computing (NTC) has been proposed to improve energy efficiency of the system, but this comes at the cost of severe degradation in performance [1]. Therefore multicore computing has been combined with NTC design to exploit parallelism and improve throughput while lowering energy/operation. However with the supply voltage being scaled down and the chip power density increasing with the number of cores, unsustainable increases in current density become a major design challenge. This leads to higher IR drop and I^2R power loss in the board/package resistance negating the effective energy gains.

To balance performance/energy, on-chip DVFS has been widely used to provide just-as-needed V_{dd}. With this approach, lower V_{dd} can be used for slower parts of the die, and higher V_{dd} for the high performance section of the chip. By bringing the DC-DC converter module closer to the processor and with boosted external voltage and local on-chip regulation, the current in the off-chip package and board parasitic can be reduced. However large variations in performance requirements and workload characteristics make on-chip DC-DC converter design very challenging. Also, while making claims about energy savings in the cores, often the energy overheads to generate such low V_{dd} power supply voltages are neglected [2]. While this can be a somewhat fair assumption for relatively narrow ranges, e.g., between full V_{dd} to 2/3 V_{dd} (as efficiency can be more than 90% in such cases), down conversion from V_{dd} to 1/2 V_{dd} or less can consume considerable overhead, which reduces the potential overall energy savings—under certain conditions, the overall power consumption might even increase.

Existing highly efficient regulator structures (like Buck-Boost architecture [3]) are not a good option for on-chip regulation as they need big inductors (also with small losses) which cannot be integrated fully on-chip. Low dropout (LDO) regulators, which are more commonly used for on-chip DC-DC conversion, enable fast voltage transition for multiple voltage domains. However LDOs suffer from low conversion efficiency, restricting their use for high step-down ratios. Switched capacitor (SC) converters are easier to integrate on-chip, but their output impedance varies with load and are not highly efficient with varying workloads [4]. Thus the benefits of wide range DVFS are negated by the lack of efficient conversion techniques. This bottleneck has been referred to as the *on-chip power regulation efficiency wall* (where the relatively poor efficiencies achievable with on-chip regulators limit the effectiveness of many low power schemes that depend on on-chip regulation) [5].

1.1. Voltage Stacking Idea

Voltage stacking simply refers to the power delivery arrangement of two or more circuit blocks such that the ground of one block becomes the power connection for the next, thus the blocks being connected as a *series* stack for power delivery, with all of them sharing the same current (thus the *charge being recycled* in the stack), while their V_{dd} values are added. With this internal recycling, for almost the same power consumption, the current drawn through the supply will be reduced to $1/n$ of conventional parallel

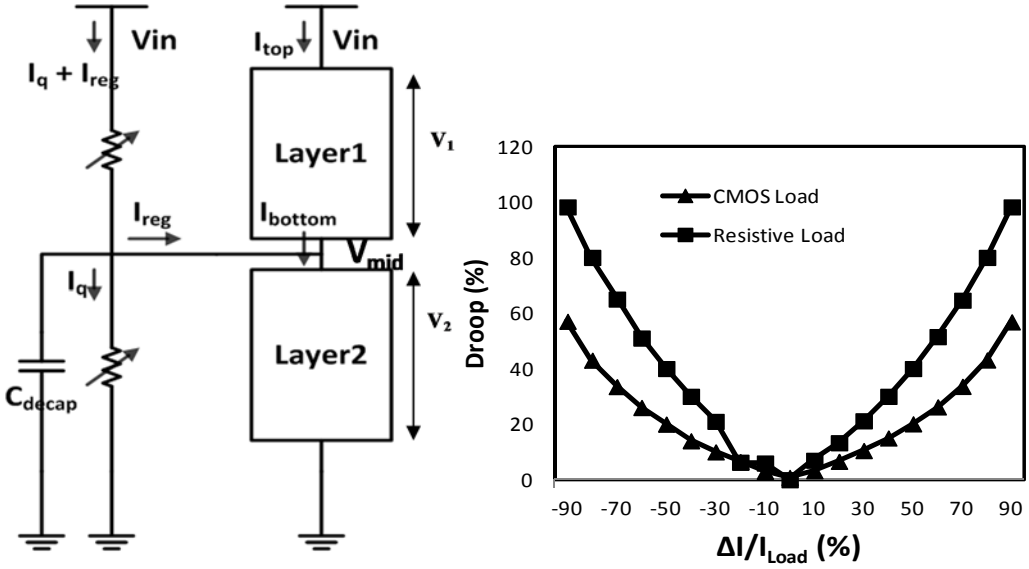
loads (n is the number of cores stacked) [5–7]. This will also reduce the off-chip I^2R power loss by a factor of n^2 and IR drop by a factor of n. The efficiency of this technique depends on the current mismatch among the stacked domains. In the best-case scenario when all the domains are perfectly balanced, IR noise is the lowest and regulation efficiency the highest, while the more the imbalance, the less the efficiency. Thus there is an implicit self-regulating loop that controls the internal voltage distribution and unless the current consumption across the layers is balanced, it can contribute to internal voltage noise [6].

In this work we explore voltage stacking as an alternative technique for on-chip DC-DC conversion for NTC. We discuss in depth the mechanism behind this highly efficient power delivery technique as well as the challenges. Optimization strategies were developed to improve the regulation performance and a novel dual-boundary hysteresis based feedback circuit was designed for increasing the efficiency. With simulations we show how performance can actually benefit from the absence of feedback circuitry for low power loads. We also compare our proposed push-pull SC converter with traditional SC regulators to demonstrate a performance improvement. The paper focuses on the simplest possible case of only two cores using voltage stacking, but the same methods and optimizations can be extended to a larger number of cores with voltage stacking as long as maximum voltage levels for the process are not exceeded by the circuit. A notable contribution of the work is the characterization of the loads. Finally, simulations are shown for various workload conditions to demonstrate the usefulness of voltage stacked power delivery.

2. DC-DC Conversion using Voltage Stacking

The idea of charge recycling through voltage stacking to generate on-chip low voltage at higher efficiency has also been considered before (Figure 1a) [7].

Figure 1. (**a**) Implicit DC-DC down conversion through voltage stacking [7]; (**b**) Implicit voltage conversion for Resistive/CMOS load.



(**a**) (**b**)

The power efficiency of this technique depends on the mismatch between the stacked domains, which in turn depends on the activity of the circuits, the evaluation node capacitance and the voltage swing in the domains. The efficiency is given by:

$$\text{Efficiency} = \frac{\text{Power\_logic}}{\text{Power\_system}} = \frac{V_{IN}\, I_{top} + V_{INT}\, |I_{reg}|}{V_{IN}\,(\, I_{top} + I_{reg}| + I_q\,)} \quad (1)$$

where $|I_{reg}|$ is the difference between the top and bottom stack and I_q is the quiescent current of the regulator [7]. Different works in the literature have handled this imbalance in different ways. In [7], charge balance was maintained between the domains through active regulation of the intermediate node using a push-pull linear regulator. However if the top core has larger current requirements than the bottom core, then the closed-loop regulator will force the excess current to ground, thus wasting power. To compensate for this loss, granules were shifted between the domains using switching logic—this came with a power and area overhead needed for the switching logic. In [8], a shunt regulator was used, and, to balance the different domains, software scheduling was done to distribute the workload at runtime. In our own work, we recycled the imbalance current among the stacked domains using an explicit regulator to improve the efficiency of this technique and maintain the output node within a certain tolerance limit [9]. By using a SC regulator, the achievable efficiency can be more than LDO. This idea is an extension of our work on GALS-based stacked cores which allow the intermediate node to implicitly track the workload of the different cores [5].

2.1. Voltage Stacking: Dependency on Nature of the Load

The idea of implicit down-conversion through voltage stacking is very intuitive if we think of the loads as identical resistors stacked upon each other. However, real load may have behavior very different from an ideal resistor (Figure 1b). Even resistive load of different magnitudes will act as an imbalance, leading to internal voltage noise. Thus this notion of implicit down-conversion will vary not only on the workload difference, but also on the nature of the load. To demonstrate this load dependency, we modeled two different kinds of load, CMOS load (multiple blocks of ring oscillators) and resistor load, drawing the same amount of current.

From first order analysis, $V_{CMOS}\ \alpha\sqrt{I_{CMOS}}$ while $V_{Resistive}\ \alpha\ I_{Resisitve.}$ Therefore, load current imbalance will cause a larger variation in resistive load than CMOS load. This quadratic versus linear dependency between voltage and current is also evident from Figure 1b where mid-voltage droop due to resistive load variation is larger than due to CMOS load variation.

2.2. Positive/Negative Regulation

In order to recycle the current imbalance between two voltage-stacked cores, current needs to be either *sourced* or *sunk* from the regulator, depending on which core consumes more current. When the bottom core consumes more, current needs to be sourced (*positive regulation*) which is similar to the conventional case (Figure 2a). However, when the top core consumes more, current needs to be *sunk* and we call this *negative regulation* (Figure 2b). To achieve this unique feature, we use a modified version of a conventional switched capacitor circuit [4]. The design is implemented with eight switches and two capacitors (Figure 3a). Unlike the case of a conventional push-pull design, with such an SC

solution the *sunk* current is fed back to the top core, thus reducing power waste. The two fly-capacitors change roles periodically, providing the *source/sink* of charge.

Figure 2. (**a**) Positive Regulation (Similar to conventional regulator, sourcing I_{Load}); (**b**) Negative Regulation (Regulator absorbs current, sinking I_{Load}).

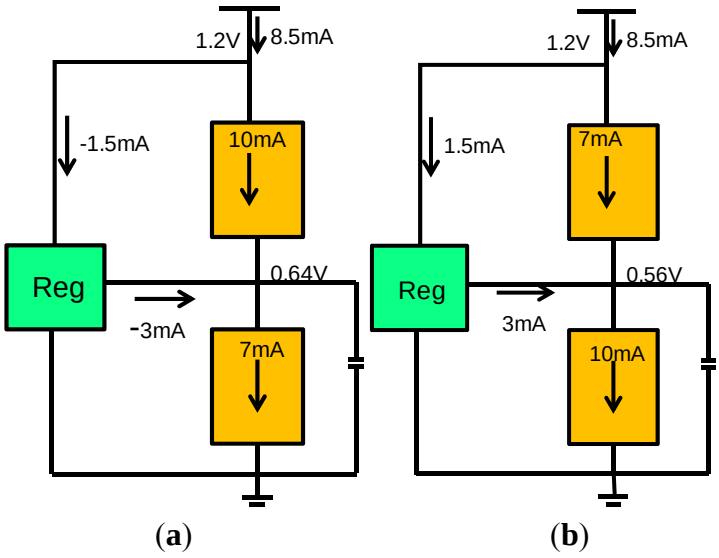
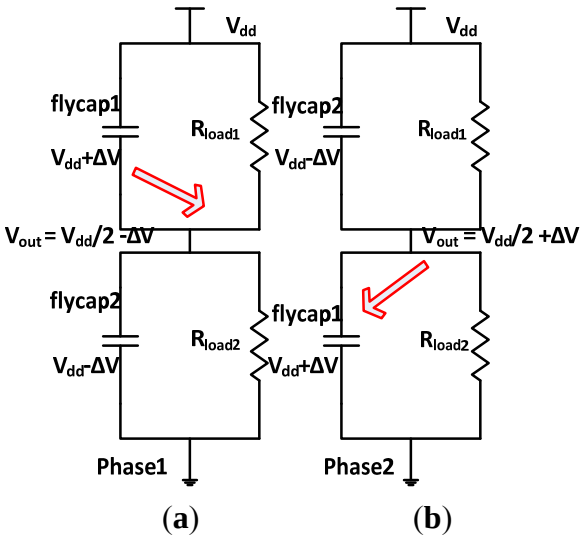


Figure 3. (**a**) Illustrates the push-pull switched capacitor designed to assist voltage stacking; (**b**) Fly-capacitors are swapped over the phases to regulate the *imbalance* between the stacks. Arrows indicates the direction from which charges are flowing to and from V_{out}.



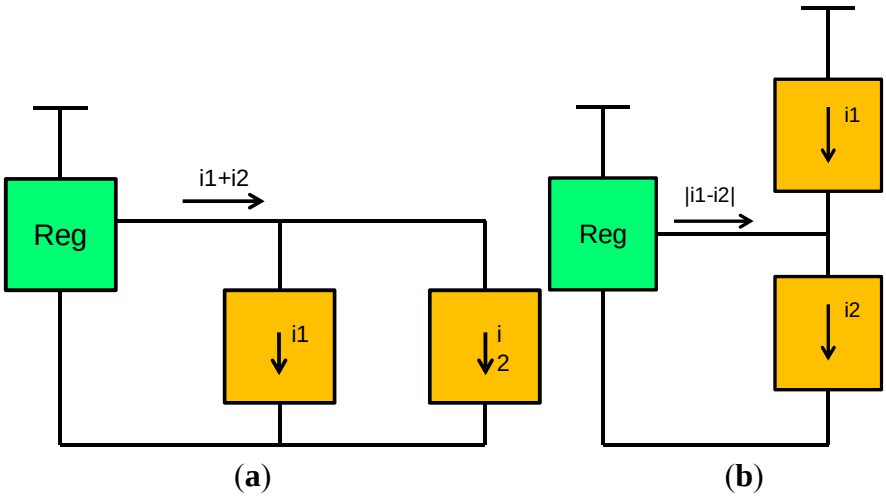
To understand how the 2:1 SC circuit works, consider an example of a slightly imbalanced workload with a supply voltage of V_{dd}, where the current offset pushes V_{out} to droop below 1/2 V_{dd} by ΔV (Figure 3b). In the first phase, as the voltage droops down at the load, flyCap1 begins charging to ΔV voltage above 1/2 V_{dd}, while the voltage on flyCap2 falls below 1/2 V_{dd} by ΔV. In the second phase, through the on-chip switches, flyCap1 and flyCap2 swap places. Since flyCap1 was charged to a higher voltage, it redirects this charge back onto the V_{out} node. This redirection of charge helps

pull the load voltage ΔV above 1/2 V_{dd}. This ripple (2ΔV) is a manifest of the capacitor charging/discharging, and the faster the switching frequency, the lower the ripple.

2.3. Higher Efficiency in Voltage Stacked Regulation

One of the primary loss components in a switched capacitor regulator is the IR loss across the power switches. For a conventional DC-DC converter providing current to parallel cores, the *entire* current needs to flow through the regulator causing large IR drop across the switches. The traditional way of lowering this loss is to design efficient switches with low parasitic resistance. However beyond a limit, it is impossible to reduce the on-resistance of the switches and conductance loss starts to dominate. The uniqueness of voltage-stacked down-conversion lies in the fact that most of the current is reused from the top core and only the *difference* of the currents needs to flow through the regulator (instead of the sum of the currents) as shown in Figure 4b. This accounts for the *higher efficiency in stacked DC-DC conversion.*

Figure 4. (**a**) Conventional DC-DC converter sourcing currents (i1 + i2) to parallel loads; (**b**) Voltage Stacked DC-DC converter sourcing/sinking currents (|i1 − i2|) to stacked loads.



2.4. Power Loss Optimization for Switched Capacitor Circuit

As explained in [10], merely optimizing for the most energy efficient design can be misleading or impractical. Ideally the optimization should try to achieve the minimum energy point subject to some design constraint, for example by using the concept of *hardware intensity* [10]. In our work, we considered a design methodology to maximize the achievable efficiency by tuning different *sensitivity knobs* for a given area constraint [11,12].

In [9] we briefly discussed the design rules for this regulator. For a given current imbalance, while the fly-capacitor size determines the *amount of charge* that can be delivered, the switching frequency sets the *output ripple*. Taking maximum ripple as a design constraint, we chose the capacitor size and switch width as the two tuning knobs. As pointed out in [11], an energy efficient design is achieved when the costs (sensitivity ratios) of tuning the knobs are balanced. Each of the points on the energy-area design space represents percent power loss per percent area for an energy-efficient regulator

design. In order to understand the design space, we developed analytical expressions (sensitivities) for all the tuning variables [12].

Switched capacitor circuit power loss can be categorized into two kinds, the series loss and the shunt loss [4]. Series loss consists of the intrinsic switched capacitor loss and the conductance loss of the switches:

$$P_{series} = P_{Cfly} + P_{Rsw} = I_{Load}\frac{\Delta V}{2} + m.I_{Load^2}\frac{R_{on}}{N.W_{sw}} \quad (2)$$

where R_{on} is the switch resistance density (Ω.m); W_{sw} is the size of each of the switches (we took all the switches to be equal sized); N the number of switches and m is a constant determined by the switched capacitor topology [4].

The additional shunt losses arise from switching the parasitic capacitance of the fly-capacitors and the power switches:

$$P_{shunt} = M_{bott.}V_0^2.C_{bott}.f_{sw} + V_{sw}^2.N.W_{sw}.C_{gate.}\frac{I_{Load}}{C.\Delta V} \quad (3)$$

where M_{bott} is a constant depending on SC topology; V_o and V_{sw} are the bottom plate-capacitor (plate-cap) and gate voltage swings; C refers to the fly-capacitor; C_{gate} refers to the gate oxide capacitance (gate-cap) of the switches and ΔV is the voltage droop.

The bottom plate-cap loss scales with only frequency while the gate-cap loss scales with both frequency and switch width. At higher load-currents the latter starts dominating, hence we neglected the bottom plate loss in our analysis. Also by using metal-insulator-metal (MIM) or trench-capacitors for the fly-capacitors, the bottom-plate loss can be minimized further. Thus total loss in switched capacitor regulator is given by:

$$P_{loss} = I_{Load}.\frac{\Delta V}{2} + m.I_{Load}^2.\frac{R_{on}}{N.W_{sw}} + V_{sw}^2.N.W_{sw}.C_{gate}.\frac{I_{Load}}{C.\Delta V} \quad (4)$$

2.5. Sensitivity Analysis

As mentioned before, we selected fly-capacitor and switch width as the two tuning variables for studying the power loss vs. area tradeoff-sensitivity analysis of the cost metrics with respect to the tuning variables and equalizing the cost across the design space provides us with optimum design points. We chose the power loss and area consumed by the switched capacitor circuit as the cost metrics:

$$A = A_cC + N.W_{sw} \quad (5)$$

where A is the total area and A_c is the area per unit capacitance—this value will depend on the capacitor technology. In our model we used the value assuming MIM cap density of 3 nF/mm^2. In [11] the sensitivity ratio for knob X is defined as:

$$S_x(X) = \frac{\frac{\partial P}{\partial X}}{\frac{\partial A}{\partial X}} \mid x = X \quad (6)$$

where S_x represents the amount of energy that can be traded-off for area by tuning variable X. The sensitivity of power loss to area due to switch width and fly capacitor are given by:

$$\frac{\frac{\partial P}{\partial W_{sw}}}{\frac{\partial A}{\partial Wsw}} = \frac{V_{sw}^2 C_{gate.}}{C.\Delta V} I_{Load} - \frac{I_{Load}^2 R_{on}}{N^2 {W_{sw}}^2} \quad (7)$$

$$\frac{\frac{\partial P}{\partial C}}{\frac{\partial A}{\partial C}} = \frac{-N.V_{sw}^2.W_{sw}C_{gate}}{\Delta V.A_C C^2} I_{Load} \quad (8)$$

For optimizing the design in the energy/area design space, the costs across the tuning variables should be equal [11]:

$$\frac{\frac{\partial P}{\partial W_{sw}}}{\frac{\partial A}{\partial Wsw}} = \frac{\frac{\partial P}{\partial C}}{\frac{\partial A}{\partial C}} \quad (9)$$

This means that at optimal design points, any marginal gain in energy will result in the same percentage amount of loss in area. The values of R_{on} and C_{gate} are fixed, depending on the technology node, while fly-capacitor values are chosen in accordance with the maximum allowable droop ΔV and I_{Load} max. In this topology, I_{Load} max refers to the worst-case current imbalance between the domains. Plugging-in the values and solving for W_{sw} provides the optimum switch width for a given capacitor size at the minimum energy point. Thus this global optimization process can allow us to look at the entire design space while balancing out the different tuning variables depending on the weight of their individual cost. To verify our claim, we did a Monte Carlo simulation selecting a large number of arbitrary capacitor values and switch widths and plotting them against our result based on sensitivity optimization. As can be seen in Figure 5, our method indeed yields the lowest points in the design space, also known as the Pareto curve. Based on the above optimization, for a given fly-capacitor size, allowable droop, and maximum $I_{Load,}$ we can find the optimum switch widths. In Figure 6, we plugged-in this optimal value of switch width and performed efficiency analysis for different load currents. Unlike for a conventional regulator, the optimization of a voltage-stacked regulator will depend not only on the I_{Load} but also on the difference of current between the domains [6]. In order to have a fair comparison, we kept one of the stacked loads at 200 mA while varied the other load from 0–200 mA. We have plotted this ratio of current imbalance to chip current consumption ($\Delta I/I_{Load}$) along the x-axis in Figure 6b. As can be seen from the plots, stacked regulation can yield higher efficiency, and since the bulk of the current comes from the off-chip supply, the switch widths and fly-capacitor area can be reduced to improve the power density of the circuit. Stacked regulation achieves 81%–95% efficiency (depending on imbalance between stacked domains) compared to 81% for non-stacked load of 200 mA for 1.2 V–0.6 V down conversion. Thus it is fair to say, even in the worst case, stacked regulation is comparable to the best case in conventional SC regulation [9].

2.6. Open-Loop versus *Closed-Loop*

In SC circuits, regulation is performed by modulating the output resistance of the converter in response to changes in load current [13]. The output of a SC converter is given by:

Figure 5. Monte Carlo Analysis of power optimization. Zoomed in view (right) shows that our sensitivity-based optimization gives the lowest points on the curve.

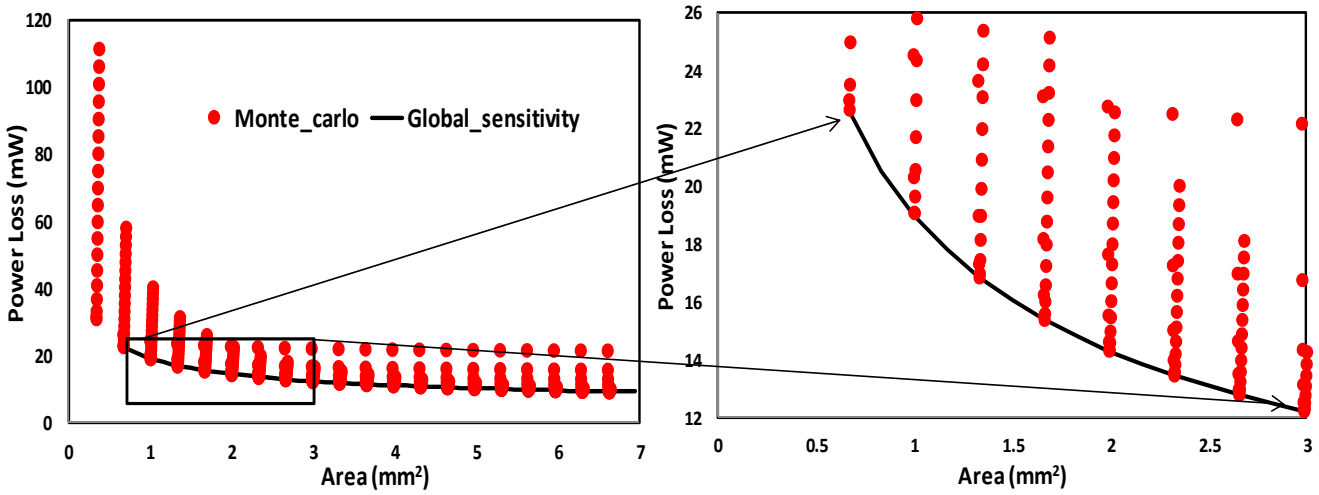
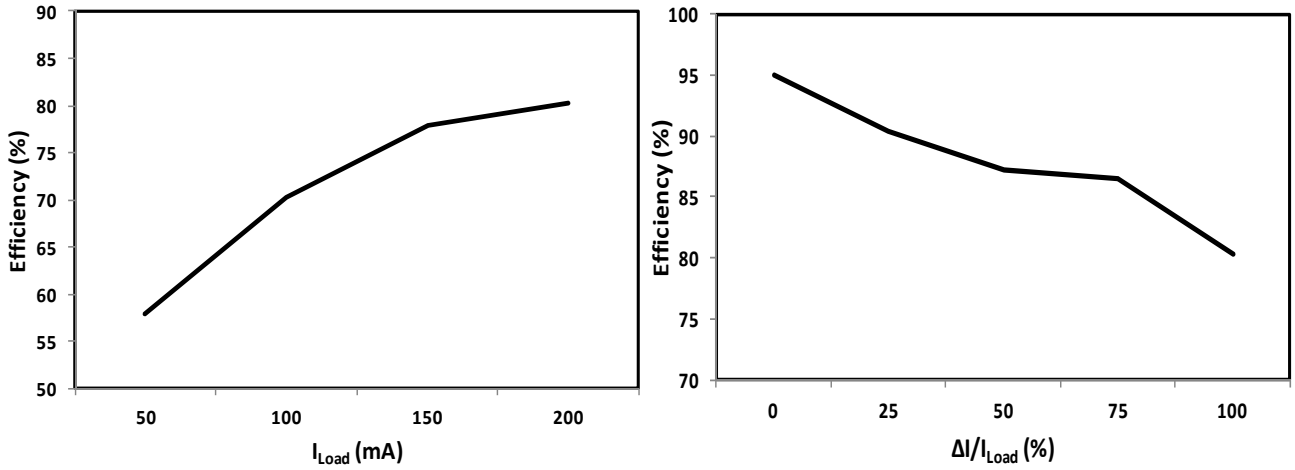


Figure 6. (**a**) Efficiency with varying conventional load (left); (**b**) Stacked load (right) with max I_{Load} = 200 mA. The X-axis indicates relative imbalance (%) between the domains.



(**a**) (**b**)

$$V_{out} = nV_{in} - i_{out} R_{out} (f_{sw}, D_i, G_i) \quad (10)$$

where n is the conversion ratio; f_{sw} the switching frequency; D_i the duty cycle of switching and G_i the conductance of the switches. Each of these variables can be used to control regulation. The conversion ratio (n) is fixed by the number of layers (cores) stacked. One of the main drawbacks of varying Di or Gi of the switches at constant frequency is reduced efficiency at lighter load. However, keeping duty cycle fixed at 50% and by modulating switching frequency with load current, higher efficiency can be achieved, especially at lighter load [4]. The downside of this is the high output voltage ripple as charge transfer is impulsive with slow switching limit of frequency [14]. Hybrid regulation, with two or three control variables together, will bring the highest efficiency, however complicated the control circuitry. In our work, we used modulation of switching frequency as the controlling variable with additional interleaving mechanism to reduce the output ripple. The traditional control method for a SC converter may include a linear feedback loop to control the switching frequency in terms of the

output voltage. However obtaining stability and good transient response over varying load conditions using such a control method is difficult. A nonlinear control can provide superior results; hence we have used a hysteretic feedback scheme with lower and upper bounds to control the regulation (Figure 7) [15]. However our control circuit is different from traditional hysteretic control and the difference comes from the stacked loads as opposed to conventional parallel loads. As explained above, current needs to be either sourced or sunk in this type of load. Consequently the output can go both high and low depending on the current mismatch between the layers, hence higher switching frequency (Clk_high) is needed whenever either of the boundaries is crossed by output voltage while low frequency (Clk_low) can regulate the in-between state. Table 1 explains the states.

Figure 7. Dual-boundary hysteretic feedback control scheme for stacked load. Output clock is pulsed between high and low frequency depending on comparator detected trigger signal.

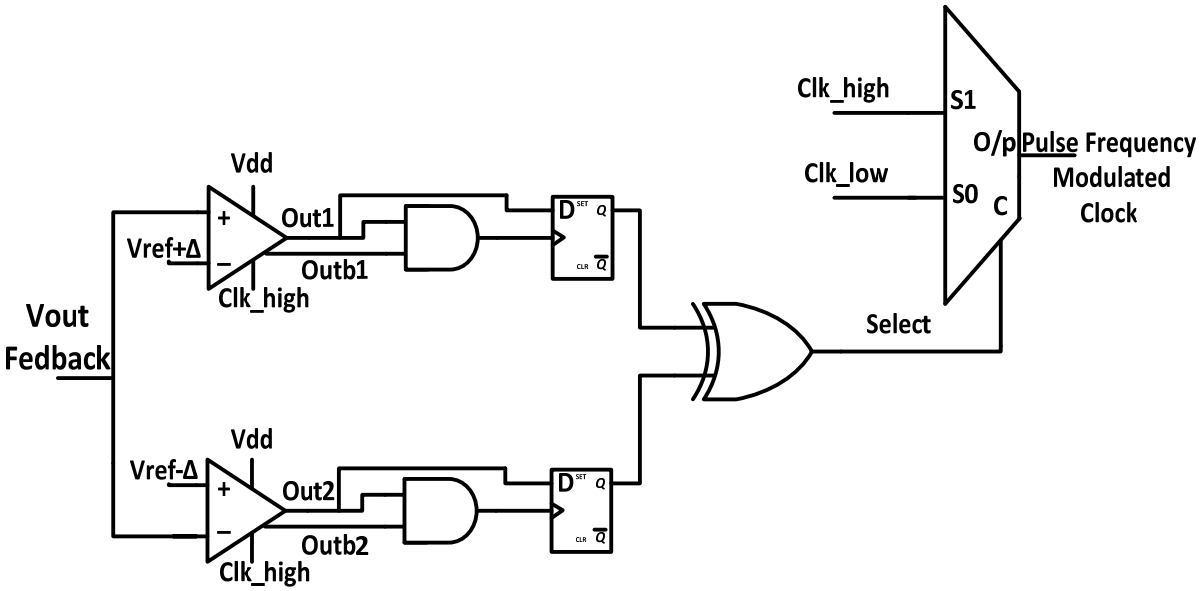


Table1. Different states for the feedback circuit to regulate V_{out}.

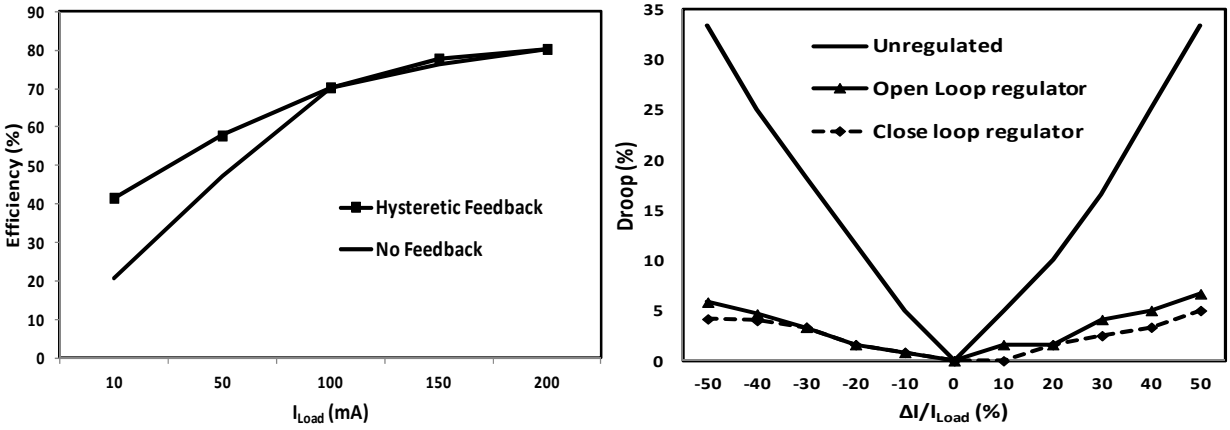
State of O/P	Out1	Out2	Select	O/P Clock
$V_{out} > V_{ref} + \Delta$	Toggle	Low	1	Clk_high
$V_{ref-\Delta} < V_{out} < V_{ref+\Delta}$	Low	Low	0	Clk_low
$V_{out} < V_{ref-\Delta}$	Low	Toggle	1	Clk_high

The feedback circuit consists of two comparators (along with latches, XOR gate and mux) that detect when the output voltage crosses the control boundary. By adding an edge-triggered latch we make sure that only rising edges are associated with a charge transfer [15]. If there were no latches, then the falling edge, which appears because of the clocked comparator and not as a trigger for boundary crossing, will cause an unwanted charge transfer. In order to account for comparator response time, we generated the clock to the latches out of the comparator itself—consequently comparator delay implicitly tracks across all PVT corners. Depending on the "Select" signal which triggers the output mux, low or high clock frequency is applied to the switches. A latch-based voltage sense amplifier was used to reduce power

consumption [16]. The dual-comparator and logic gate controlled feedback circuit power consumption are critical, especially for low power operation. For 2 V–1 V stacked conversion for a load current of 200 mA with the comparator running at 1 GHz, feedback circuit power consumption is 1.2 mW. However this can be scaled down depending on the operational frequency of the load. For the low power (1.2 V–0.6 V, 10 mA) conversion, feedback power scales down to 0.2 mW.

This feedback circuit was used with the SC converter and we have done extensive simulations for 2:1 DC-DC conversion using both parallel loads and stacked loads. Near threshold circuits are typically operated 200 mV above their threshold value [17]. Hence here we considered 0.6 V as NTC V_{out} to be delivered. Conversion efficiency for both low power (0.5 mW–10 mW, 1.2 V–0.6 V) and high power (10 mW–400 mW, 2 V–1 V) loads have been shown for comparison.

Figure 8. (**a**) Feedback SC circuit applied to conventional load; (**b**) Comparisons between unregulated, open-loop and closed-loop stacked loads.



(**a**) (**b**)

As Figure 8a shows, the feedback circuit has considerable impact on efficiency for conventional load, especially at lower current. However stacked load has unique characteristics which are explained below with the help of Figure 1a [9].

In Section 2.1 we discussed how the nature of the load can affect voltage stacked DC-DC conversion. Here we discuss how the voltage headroom in each of the stacked cores is going to change with load or activity variation. CMOS load has been considered here, representing the core (Figure 1a).
By charge conservation:

$$I_{top} = I_{bottom} \tag{11}$$

Current consumption of the two cores is given by:

$$I_{top} = \alpha_{top} C_L (V_{dd} - V_{mid}) F_c \qquad I_{bottom} = \alpha_{bottom} C_L V_{mid} F_c \tag{12}$$

$$V_1 = V_{dd} - V_{mid} = \frac{\alpha_{bottom}}{\alpha_{bottom} + \alpha_{top}} V_{dd} \qquad V_2 = V_{mid} = \frac{\alpha_{top}}{\alpha_{top} + \alpha_{bottom}} \tag{13}$$

where α_{top} and α_{bottom} are the top and bottom core activity factors; F_c is the core frequency and V_{mid} is the implicit voltage generated by stacking two cores.

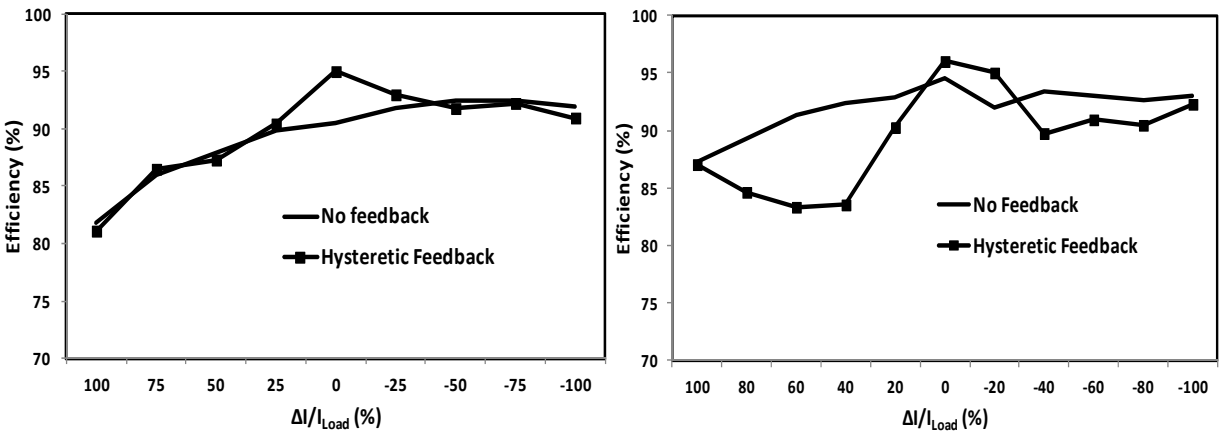
For

$$\alpha_{bottom} = \alpha_{top},\ V_{mid} = 0.5V_{dd} \tag{14}$$

However, if $\alpha_{top} > \alpha_{bottom}$, $V_2 > V_1$; $\alpha_{top} < \alpha_{bottom}$, $V_1 > V_2$. This means that the inherent feedback of voltage stacking forces the voltage headroom to be lower for the core that demands *higher* current, thus acting against the idea of DVFS (that when computing demands are *lower* the voltage should be reduced). Therefore, this internal voltage buildup in voltage stacking can not only add noise to the stacked cores, but also oppose DVFS. This is shown in Figure 8b. The more the imbalance, the more the self-regulation of the system will force the V_{mid} node to go in the opposite direction as seen from the unregulated stacked mid rail. Thus for voltage stacking to work, we need to compensate for this natural feedback tendency and the push-pull scheme is therefore essential for maintaining the charge imbalance within bounds.

However whether adding a feedback loop over the SC converter can bring any benefits or not, is what we try to show in Figure 9a,b. As shown in some of the earlier plots, to demonstrate effectiveness of voltage stacking, efficiency needs to be plotted against the ratio of load current to load imbalance. Thus midpoint on the x-axis for both the plots indicates when the loads are perfectly balanced. Ideally the SC does not need to regulate at that point and it can lower its switching frequency to minimal value to improve energy efficiency. This is shown in Figure 9a,b where the feedback controller reduces the switching frequency around the balanced load condition (midpoint region ofx-axis) and increases the efficiency as compared to the open loop regulator. However when the loads are not in balance, the performance of the closed loop regulator is in fact worse compared to the open loop one for low power loads (Figure 9b). This is because for low power (NTC) stacked loads, the feedback circuit overhead reduces the efficiency. At an imbalance of 50%, closed-loop SC converter suffers an efficiency loss of 10.8% over an open-loop SC converter. Thus for low power DC-DC conversion, voltage stacking can provide an attractive alternative technique and by removing the additional losses in the feedback circuitry, the efficiency can be improved further [9]. However, for closely matched high power loads, hysteretic feedback can still provide higher efficiency as shown in Figure 9a.

Figure 9. (**a**) Comparison of open-loop/close-loop SC circuit for high power: 10 mW-400 mW, 2 V–1 V; (**b**) Low power: 0.5 mW–10 mW, 1.2 V–0.6 V.



3. Practical Loads and Capacitor Models

While the SC circuits described above work well for DC current loads, their performance needs to be evaluated with respect to transient response. For this we used two different types of load to analyze the performance of the SC converter. The first load in Figure 10a,b models the big changes in load current during different power modes and performance states in real cores. The transient response in Figure 10c shows that the output remains within a 5% tolerance limit even in the worst case scenario (I_{min} to I_{max}). An additional startup circuitry can reduce the initial voltage droop. With actual applications running on the core, the current load hardly retains the steady state DC value as used in the model. To mimic a more real situation we generated different current profile by running a benchmark application (Dedup, a benchmark from the PARSEC suite) using an accurate full system multicore simulator M5 [18]. These current traces are then included in Cadence ADE as current sources, shown in Figure 11a,b while the output voltage generated from the stacked SC converter is shown in Figure 11c. Adopting the calculation as shown in Equation (1), 2 V–1 V conversion using this stacked approach yields efficiencies as high as 93%. The p-pripple on the V_{mid} (V_{out}) still remains within the tolerance limit. Further reduction in V_{mid} variation can be achieved by increasing the SC converter frequency or by increasing the on-chip decap, however this comes with tradeoffs in terms of efficiency or area.

Figure 10. Current load for different mode change. (**a**) upper stacked loads; (**b**) lower stacked loads; (**c**) 2:1 down converted V_{out}.

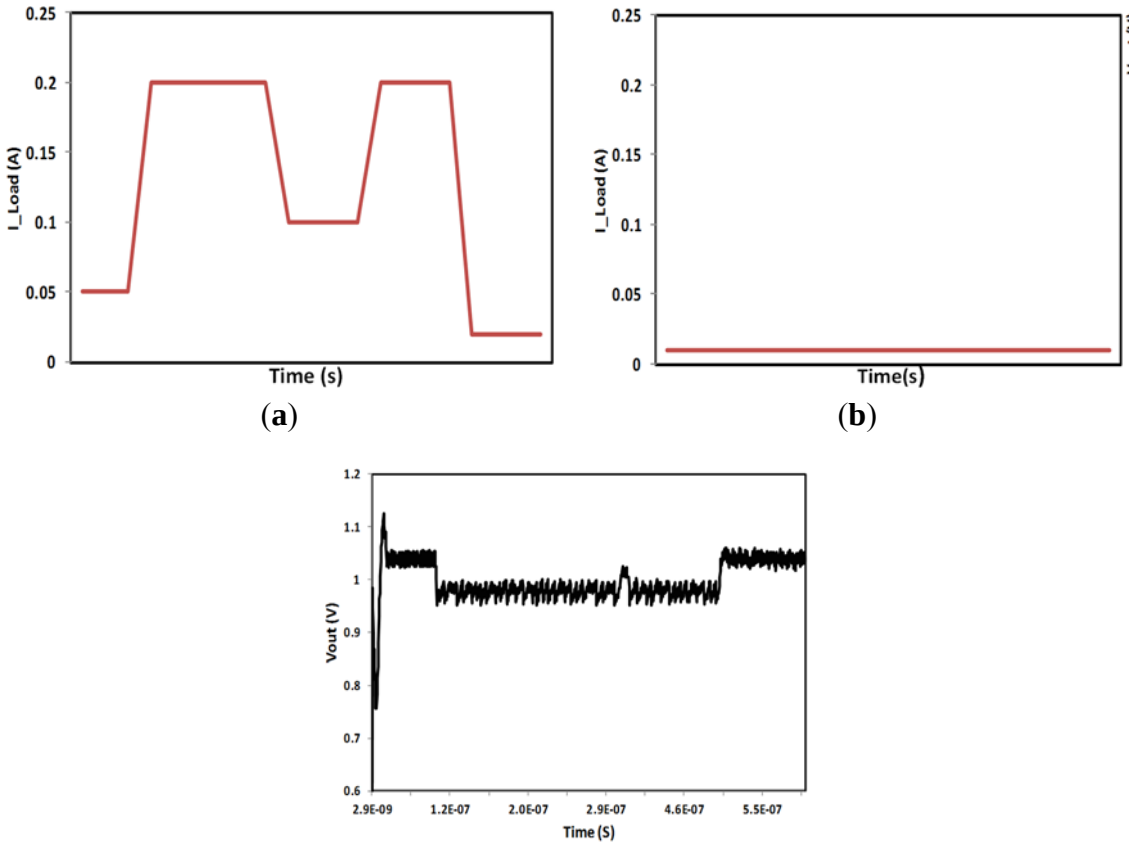
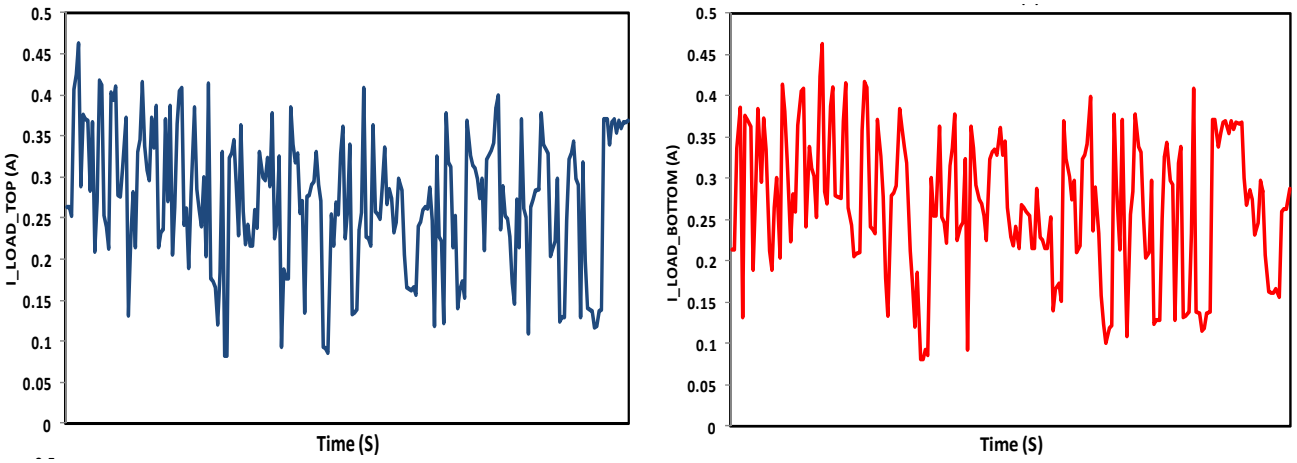
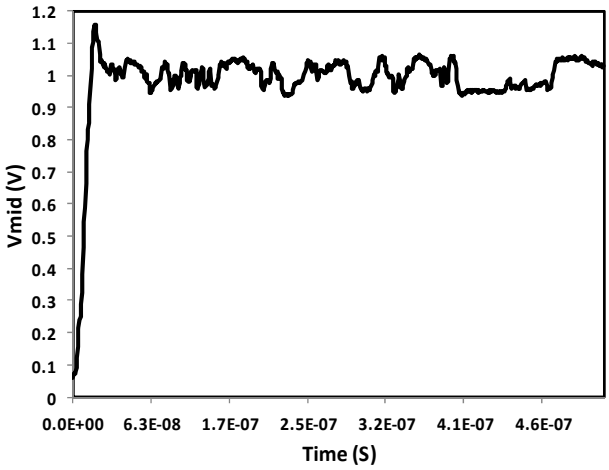


Figure 11. Current traces generated from running benchmark and including them in Cadence ADE as load. (**a**) upper stack load; (**b**) lower stacked loads; (**c**) 2:1 Down converted Vout (Vdd = 2 V, Vout ~ 1 V).



(**a**)

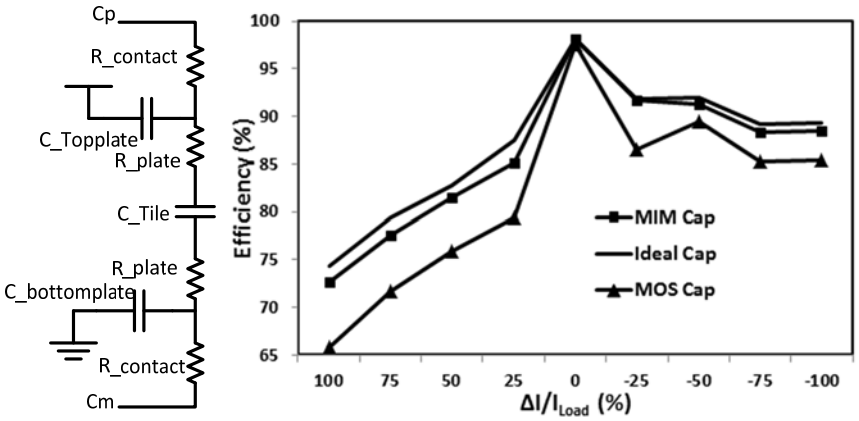
(**b**)



(**c**)

All the simulations so far have used ideal capacitors. However bottom plate parasitic and ESR of the capacitors can have a huge impact on the conversion efficiency and area constraint. Bulk CMOS technology with low ESR can provide capacitive density of up to 12 nF/mm^2, but bottom plate capacitance is highest in MOS Caps among the capacitor technology owing to proximity to substrate (5%–10%). MIM (Metal-insulator-Metal) capacitor with a lower bottom plate parasitic (1%–3%) can provide a good alternative to higher efficiency at the cost of area (3 nF/mm^2) and high ESR (Equivalent Series Resistance) [17]. To account for the parasitic losses, we used models of MIM cap and MOS cap in the simulation including the top and bottom plate capacitive parasitic as well as the contact and plate resistance (Figure 12a). Efficiency for 2:1 down-conversion using MIM cap ranges as high as 72%–91% for a load imbalance of 0–200 mA and maximum load being 400 mA (Figure 12b). Performance of the MOS cap is the worst due to the parasitic. However it gives the area advantage over MIM cap.

Figure 12. (**a**) MIM capacitor model, including all the parasitic; (**b**) Efficiency comparison between real (MIMcap and MOScap) and ideal capacitor.



(**a**) (**b**)

4. Conclusions and Future Work

With aggressive voltage and technology scaling, power delivery is expected to be a major challenge for the semiconductor industry. With increasing power densities, reduction in power pin count and quadratic increase in off-chip power loss, traditional power delivery methods may not be efficient enough. In this work, we exploited the concept of charge recycling through voltage stacking to convert high off-chip voltage to low on-chip voltage at higher efficiency. While voltage stacking eases off-chip power delivery issues, this topology presents a unique challenge of within die noise due to interlayer current mismatch. In our work, we used a push-pull switched capacitor circuit to recycle and redistribute the charge imbalance among the stacked layers, thus reducing voltage noise. We presented a design methodology for optimizing the power loss of the switched capacitor circuit based on global sensitivity analysis. Analytically derived optimum switch width was used in simulating the switched capacitor, both for conventional and stacked load. The high efficiency of the stacked DC-DC conversion justifies the integration of voltage stacking in the 2D IC power delivery network. A novel dual boundary hysteresis-based feedback scheme was developed to regulate the SC converter. However we claim that the self-regulated stacked domains will work better as open-loop regulatora, especially for low power loads. Stacked SC regulator down-conversion (2 V–1 V) was shown at a high efficiency of 93% with current traces generated from real benchmark being used as the load. Even by incorporating the parasitic of the MIM capacitor in simulation, efficiency still ranges from 72%–91% for a load imbalance of 0–200 mA and maximum load of 400 mA. Thus with reduction of the DC-DC converter energy overhead, just-as-needed V_{dd} and near threshold computing can become even more energy efficient.

References

1. Dreslinski, R.G.; Wieckowski, M.; Blaauw, D.; Sylvester, D.; Mudge, T. Near-threshold computing: reclaiming moore's law through energy efficient integrated circuits. *Proc. IEEE* **2010**, *98*, 253–266.
2. Yu, P.; Xin, Z.; Huang, J.; Muramatsu, A.; Nomura, M.; Hirairi, K.; Takata, H.; Sakurabayashi, T.; Miyano, S.; Takamiya, M.; *et al*. Misleading Energy and Performance Claims in Sub/Near Threshold Digital Systems. In Proceedings of 2010 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Jose, CA, USA, 7–11 November 2010; pp. 625–631.
3. Patounakis, G.; Li, Y.W.; Shepard, K.L. A fully integrated on-chip DC-DC conversion and power management system. *IEEE J. Solid-State Circuits* **2004**, *39*, 443–451.
4. Hanh-Phuc, L.; Sanders, S.R.; Alon, E. Design Techniques for fully integrated switched-capacitor DC-DC converters. *IEEE J. Solid-State Circuits* **2011**, *46*, 2120–2131.
5. Mazumdar, K.; Stan, M. Breaking the Power Delivery Wall Using Voltage Stacking. In Proceedings of the Great Lakes Symposium on VLSI, Salt Lake City, UT, USA, 3–4 May 2012; pp. 51–54.
6. Lee, S.K.; Brooks, D.; Wei, G.-Y. Evaluation of Voltage Stacking For Near-Threshold Multicore Computing. In Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design, Redondo Beach, California, USA, 30 July–01 August 2012; pp. 373–378.
7. Rajapandian, S.; Zheng, X.; Shepard, K.L. Implicit DC-DC downconversion through charge-recycling. *IEEE J. Solid-State Circuits* **2005**, *40*, 846–852.
8. Kanev, S. Motivating Software-Driven Current Balancing in Flexible Voltage-Stacked Multicore Processors. Bachelor's Thesis, Harvard University, Cambridge, MA, USA, May 2012.
9. Mazumdar, K.; Stan, M.R. Charge Recycling On-Chip DC-DC Conversion for Near-Threshold Operation. In Proceedings of the Subthreshold Microelectronics Conference (SubVT), Waltham, MA, UK, 9–10 October 2012; pp. 1–3.
10. Brodersen, R.W.; Horowitz, M.A.; Markovic, D.; Nikolic, B.; Stojanovic, V. Methods for True Power Minimization. In Proceedings of the 2002 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), New York, NY, USA, 10–14 November 2002; pp. 35–42.
11. Zhenyu, Q.; Ziegler, M.; Kosonocky, S.V.; Rabaey, J.M.; Stan, M.R. Multi-dimensional Circuit and Micro-architecture Level Optimization. In Proceedings of the 8th International Symposium on Quality Electronic Design(ISQED), San Jose, CA, USA, 26–28 March 2007; pp. 275–280.
12. Mazumdar, K.; Stan, M.R. Breaking the 3-D IC Power Delivery Wall. In Proceedings of the 2012 Asilomar Conference on Signals, Systems and Computers, Asilomar, CA, USA, 4–7 Novmber 2012.
13. Seeman, M.D. A Design Methodology for Switched-Capacitor DC-DC Converters. Ph.D. Thesis, University of California, Oakland, CA , USA, 2009.
14. Ng, V.W.S. Switched Capacitor DC-DC Converter: Superior Where the Buck Converter has Dominated. Ph.D. Thesis, University of California, Oakland, CA, USA, 2011.
15. Van Breussegem, T.M.; Steyaert, M.S.J. Monolithic capacitive DC-DC converter with single boundary-multiphase control and voltage domain stacking in 90 nm CMOS. *IEEE J. Solid-State Circuits* **2011**, *46*, 1715–1727.

16. Jain, R.; Sanders, S. A 200 mA switched capacitor voltage regulator on 32 nm CMOS and regulation schemes to enable DVFS, In Proceedings of the 14th European Conference. Birmingham, UK, 30 August–1 September 2011; pp. 1–10.
17. Fick, D.; Dreslinski, R.G.; Giridhar, B.; Gyouho, K.; Sangwon, S.; Fojtik, M.; Satpathy, S.; Yoonmyung, L.; Daeyeon, K.; Liu, N.; *et al.* Centip3De: A 3930 DMIPS/W Configurable Near-Threshold 3D Stacked System with 64 ARM Cortex-M3 Cores. In Proceedings of the 2012 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, USA, 19–23 Feberary 2012; pp. 190–192.
18. Binkert, N.L.; Dreslinski, R.G.; Hsu, L.R.; Lim, K.T.; Saidi, A.G.; Reinhardt, S.K. The M5 simulator: Modeling networked systems. *IEEE Micro* **2006**, *26*, 52–60.