

# Implicit DC–DC Downconversion Through Charge-Recycling

Saravanan Rajapandian, *Student Member, IEEE*, Zheng Xu, *Student Member, IEEE*, and Kenneth L. Shepard, *Senior Member, IEEE*

**Abstract**—This paper describes an energy-efficient means to achieve on-chip dc–dc conversion for digital CMOS circuits. The approach uses balanced voltage islands running at fractions of the off-chip supply voltage. Charge “discarded” by one domain is “recycled” to supply energy for another. When the domains are ideally balanced, all the energy dissipated by electrons in “dropping” to lower potentials is used for active computation. We describe the design and measurement of a prototype system in a 0.18  $\mu\text{m}$  CMOS process that provides active on-chip voltage regulation and controlled dc–dc conversion with this technique.

**Index Terms**—Dc–Dc conversion, power management.

## I. INTRODUCTION

**D**YNAMIC (or adaptive) voltage scaling (DVS) of digital circuits offers the ability to trade-off power and performance through adjustment of the supply voltage,  $V_{DD}$  [1]–[3]. Most DVS systems are based on the idea that multiple power grids are available to be “tapped into” to support multiple voltage operation, which comes at the cost of additional complexity and area [4], [5]. An alternative is to provide for dynamic dc–dc conversion for each voltage island, which would allow for continuous scaling, negate the need for multiple global power grids, but require a very efficient adaptive power supply regulator, preferably one that is small and can be completely integrated on-chip. This on-chip support for multiple supply voltages also allows analog circuits (which in general benefit from larger voltage headroom) access to higher operating voltages.

The most efficient dc–dc converters are buck-type regulators, which generate a reduced dc level by filtering a pulse-width modulated (PWM) signal through a simple  $LC$  filter [6]. By varying the frequency or duty-cycle of the PWM signal, different dc levels can be generated. While buck converters can operate at very high efficiencies ( $> 80\%$ ), they require off-chip filter components, which limits their usefulness. Recent work has considered the prospect of integrating inductors on chip that include magnetic materials [7], but this clearly adds considerable cost and complexity to fabrication. To deal with this limitation, two other types of dc–dc converters are possible which can

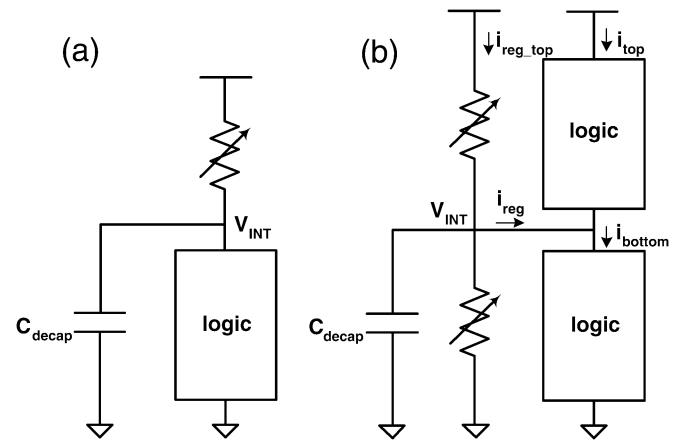


Fig. 1. (a) Linearly regulated reduced voltage requires a large power transistor; (b) this power transistor can be replaced with a logic and only a small push-pull regulator is needed to regulate the node  $V_{int}$ .

be easily integrated on-chip—linear regulators and switched-capacitor power supplies—but these offer comparatively poor efficiencies.

A linear regulator is a power transistor [shown as a variable resistor in Fig. 1(a)] that is controlled by a feedback amplifier to keep the intermediate supply voltage  $V_{int}$  constant with changing load current demands. The efficiency of such a linear regulator is limited to  $V_{int}/V_{DD}$ , which provides for low efficiencies at small values of  $V_{int}$ . In principle, switched-capacitor (SC) supplies allow one to produce lower voltages at higher efficiencies than linear regulators. SC supplies are capacitance dividers, in which the capacitors are periodically “exchanged” as they are discharged by the load current. Because of practical limitations, overall efficiency in generating, for example, a  $V_{DD}/2$  supply voltage with reasonable loading is still quite poor even in carefully designed systems [8] (about 60%–65%). Furthermore, both linear regulators and switched-capacitor supplies consume large on-chip areas for both the power transistors (of the linear regulator) and the capacitors (of the switched-capacitor supply).

In this paper, we explore an alternative, *charge-recycling low-voltage regulation*, that allows for very energy-efficient on-chip generation of reduced supply voltages [9]. The idea borrows, in part, from charge-recycling ideas as applied in buses [10] or the highly capacitive lines of memories [11], [12]. In this case, as shown in Fig. 1(b), we replace the power transistor of the linear regulator by *another* domain of logic. Now, the energy lost by electrons in dropping from potential  $V_{DD}$  to potential  $V_{int}$  is usefully employed in computation in the “top” logic domain.

Manuscript received September 1, 2004; revised November 30, 2004. This work was supported by the DARPA/MARCO C2S2 Center, by the Semiconductor Research Corporation (SRC), by the National Science Foundation under Contract CCR-0086007, and by gifts from IBM and Intel.

The authors are with the Columbia Integrated Systems Laboratory, Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: sara@cisl.columbia.edu; zheng@cisl.columbia.edu; shepard@cisl.columbia.edu).

Digital Object Identifier 10.1109/JSSC.2004.842861

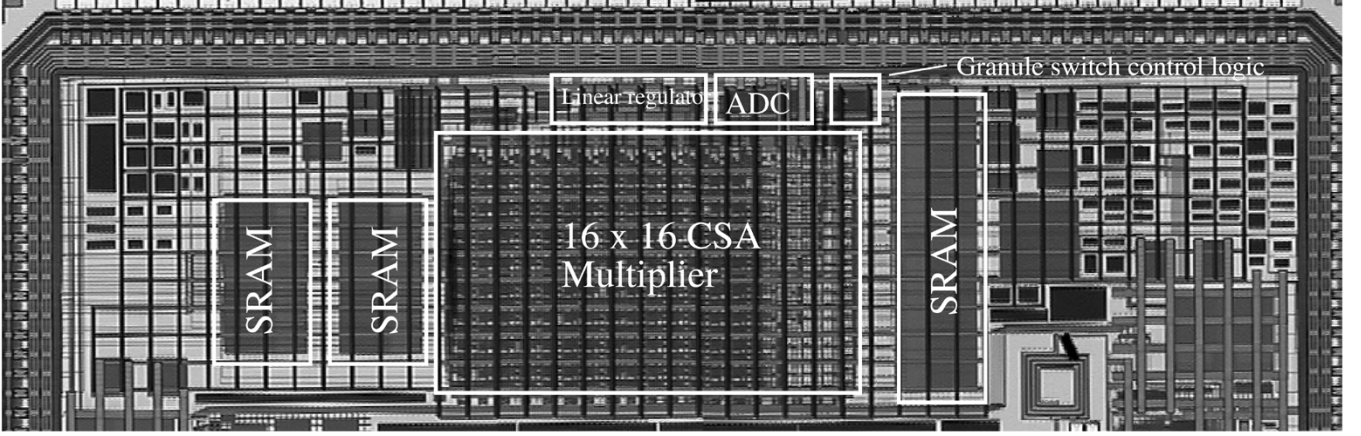


Fig. 2. Die photo of the prototype system.

This charge is then “recycled” to be used in the logic computation of the “bottom” domain. If  $V_{\text{int}} = V_{\text{DD}}/2$ , then this top logic domain is operating between  $V_{\text{DD}}$  and  $V_{\text{DD}}/2$  while the bottom domain is operating between  $V_{\text{DD}}/2$  and ground. Interface circuits, considered in more detail in Section II-B, allow for communication between logic circuits operating across different potential ranges.

To achieve high efficiency, the charge demands of the top and bottom domain must be “balanced,” so that the charge required by the bottom domain can be completely provided by the top domain. Charge imbalances will inevitably come about because of differences in the evaluation node capacitances of the two domains or because of differences in circuit activity in the two domains. A push-pull linear regulator, shown schematically in Fig. 1(b), controls the  $V_{\text{int}}$  supply node by adding or subtracting charge as required. The power transistors [shown in Fig. 1(b) as two variable resistors] can be far smaller than the power transistor required for the linear regulator of Fig. 1(a) because they must only provide the supply-current *difference* between the two domains. Similarly, the decoupling capacitance requirements on the  $V_{\text{int}}$  node in Fig. 1(b) are significantly lower than the requirements in Fig. 1(a).

We note that this approach is general in that voltages other than  $V_{\text{DD}}/2$  may be regulated; for example, one domain may operate at  $V_{\text{DD}}/3$ , while another operates at  $2V_{\text{DD}}/3$ . Similarly, the scheme can be generalized to more than two domains and to regulate down from voltages above  $V_{\text{DD}}$ . For example, one could bring in an off-chip  $3V_{\text{DD}}$  supply and recycle charge through three domains, one operating between  $3V_{\text{DD}}$  and  $2V_{\text{DD}}$ , the second operating between  $2V_{\text{DD}}$  and  $V_{\text{DD}}$  and the third operating between  $V_{\text{DD}}$  and ground. By “stacking” the logic domains  $n$  high, the current demands on the power and ground networks are also reduced by a factor of  $n$  over the current demands of all of the domains running in parallel at the reduced supply.

In this paper, we consider the use of charge-recycling, low-voltage regulation to dynamically operate a large block of digital CMOS either at full  $V_{\text{DD}}$  supply or at a scaled  $V_{\text{DD}}/2$  supply. In Section II, we describe the prototype system that we have designed, fabricated, and tested. Measurement results on this system are presented in Section III. Section IV concludes.

## II. SYSTEM DESIGN

We have designed and fabricated a simple prototype system, shown in the die photo of Fig. 2, consisting of a  $16 \times 16$  carry-save-array multiplier in a TSMC  $0.18 \mu\text{m}$  triple-well process. The multiplier, which has 20 pipeline stages, can be dynamically configured to run at 400 MHz at full supply or 200 MHz at half-supply.<sup>1</sup> Two on-chip SRAMs feed data into the multiplier and a third stores the result.

The overall structure of the voltage-scaled multiplier is shown in Fig. 3. The multiplier logic is partitioned into a set of 16 partitions (or “granules”), representing a total gate capacitance of approximately 8 pF per granule. When running full-rail, all the granules are configured to operate between  $V_{\text{DD}}$  and ground. When running at  $V_{\text{DD}}/2$ , approximately half of the granules are configured to operate between  $V_{\text{DD}}$  and  $V_{\text{DD}}/2$  and the other half to operate between  $V_{\text{DD}}/2$  and ground. The linear regulator, in addition to controlling the  $V_{\text{int}}$  node, monitors the amount of current being sourced or sunk through the output stage. This allows granules to be automatically switched (by the action of the granule-switching control logic) between the top and bottom domains in the presence of long-time-scale mismatch.

### A. Linear Regulator

The linear regulator design, consisting of two single-stage differential error amplifiers and a push-pull output stage (transistors M1 and M2), is shown in detail in Fig. 4. A simple switched-capacitor divider is used to generate the  $V_{\text{DD}}/2$  reference (**half\_vdd\_ref**) for the linear regulator. At 1 mA of dc output load, the regulator (with 13 pF of decoupling capacitance on  $V_{\text{int}}$ ) has an open-loop dc gain of 38 dB and a unity gain bandwidth of 130 MHz with a phase margin of 70 degrees. The amplifier driving transistor M1 is biased with 200  $\mu\text{A}$ , while the amplifier driving M2 is biased with 400  $\mu\text{A}$ . The output stage has a quiescent current of 50  $\mu\text{A}$ . This quiescent current has an unfortunately strong dependence on error amplifier dc offsets which directly add into the gate-to-source bias of output transistors M1 and M2. We attempt to minimize amplifier offset with large input devices and careful layouts. The regulator can source

<sup>1</sup>Unfortunately, the longest pipeline state is approximately 35 fanout-of-four (FO4) delays, limiting the performance.

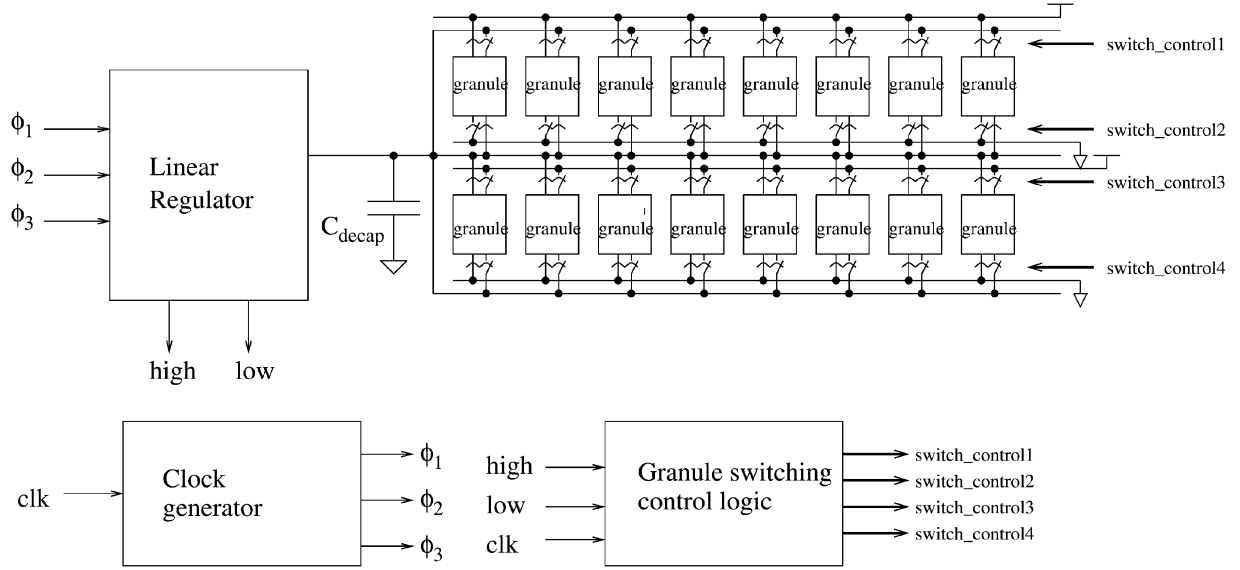


Fig. 3. Charge-recycling prototype system. A linear regulator controls the  $V_{int}$  node. When the linear regulator is sourcing or sinking a large amount of current, the high or low signals trigger the granule-switching control logic to switch granules between domains to remove the imbalance.

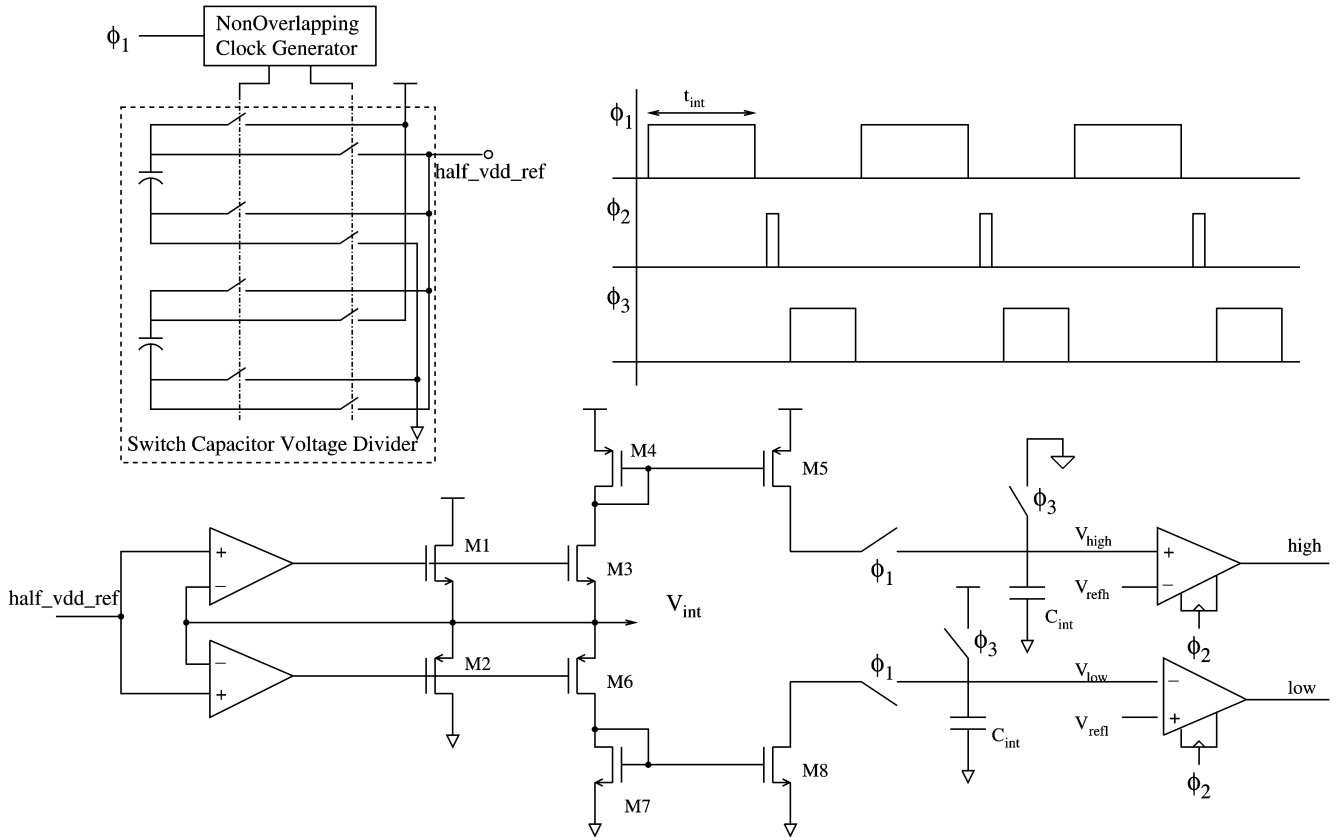


Fig. 4. Linear regulator design with domain charge mismatch monitoring.

or sink  $I_{max} = 30$  mA of current before losing regulation, providing current efficiency of 96% for maximum load.

Power transistors M1 and M2 have widths of 600  $\mu\text{m}$  and 1.2 mm, respectively, resulting in subthreshold biasing at zero load current. Transistors M3–M5 mirror out a current proportional to that flowing through transistor M1 for integration onto the capacitor  $C_{int} \cong 0.4$  pF. Similarly, transistors M6–M8 mirror out a current proportional to that flowing through tran-

sistor M2, also integrated onto a capacitor. The current being integrated is 1/800 that flowing through the corresponding regulator output transistor. Clock phases  $\phi_1$  and  $\phi_3$  are used to control the integration, establishing an integration time of approximately  $t_{int} = 150$  ns. After the integration window,  $\phi_2$  clocks the comparators to compare the voltages  $V_{high}$  and  $V_{low}$  with the reference levels  $V_{refh}$  and  $V_{refl}$ , respectively, producing the signals **high** and **low** to the granule-switching control logic.

The use of source-follower transistors M1 and M2 in the output stage of the linear regulator provides better stability and the need for less decoupling capacitance than the more traditional common-source output stage. In many linear regulator applications, the common-source is preferred because of its lower dropout voltage. In this application, dropout voltage is not a concern since we are regulating far from the rails. Assuming equal transconductances ( $g_m$ ) for transistors M1 and M2 and equal dc gain ( $A$ ) and output resistance ( $r_o$ ) for the two error amplifiers, the open loop gain of the regulator is approximately given by

$$H_{\text{open}}(s) \cong \frac{A/2}{(1 + s/\omega_1)(1 + s/\omega_2)} + \frac{A/2}{(1 + s/\omega_3)(1 + s/\omega_4)} \quad (1)$$

where  $\omega_1 = 2g_m/(C_{gs1} + C_{\text{decap}})$ ,  $\omega_2 = 2/C_{gs1}r_{o1}$ ,  $\omega_3 = 2g_m/(C_{gs2} + C_{\text{decap}})$ ,  $\omega_4 = 2/C_{gs2}r_{o2}$ ,  $C_{gs1}$  and  $C_{gs2}$  are the gate-to-source capacitances of output transistors M1 and M2, respectively, and  $C_{\text{decap}}$  is the decoupling capacitance at the output of the regulator. Assuming that the poles  $\omega_2$  and  $\omega_4$  dominate and for frequencies  $\omega \ll \omega_1, \omega_3$

$$H_{\text{open}}(s) \cong \frac{A(1 + s/(\omega_4 + \omega_2)/2)}{(1 + s/\omega_2)(1 + s/\omega_4)}. \quad (2)$$

That is, a zero is introduced between the two poles at  $\omega_2$  and  $\omega_4$ , leading to stable feedback as long as these two poles are dominant. The poles due to the decoupling capacitance ( $\omega_1$  and  $\omega_3$ ) will be large because of the large transconductances of the output source follower transistors.

Decoupling capacitance on  $V_{\text{int}}$  must ensure a low enough impedance beyond the closed-loop  $-3$  dB bandwidth ( $f_{-3\text{dB}} \cong 130$  MHz) of the regulator, where the linear regulator is ineffective. In this case, we are approximating the response time of the regulator by  $T_R \cong 1/f_{-3\text{dB}} \cong 7.5$  ns.

$$C_{\text{decap}} \cong \frac{\Delta i_{\text{reg}}}{\Delta V_{\text{int}} 2\pi f_{-3\text{dB}}} \quad (3)$$

where  $\Delta i_{\text{reg}}$  is the peak-to-peak magnitude of high-frequency-content current transients and  $\Delta V_{\text{int}}$  is the allowable peak-to-peak voltage transient on the regulated node. In our case, for the reasonably balanced domains guaranteed by the presence of the domain switching logic, we find a maximum transient current mismatch of approximately  $\Delta i_{\text{reg}} \cong 1$  mA. For a regulation requirement of  $\Delta V_{\text{int}} \cong 90$  mV, this yields  $\Delta V_{\text{int}}/\Delta i_{\text{reg}} \cong 90\Omega$  (a relatively high-impedance target relative to most regulator applications) and  $C_{\text{decap}}$  of approximately 13 pF, of which 4 pF is provided by explicit on-chip thin-oxide decoupling capacitance, while the remainder is provided by nonswitching circuits and well capacitance. In Fig. 5, we show the simulated output impedance of the regulator alone (without  $C_{\text{decap}}$ ), the impedance of the decoupling capacitance alone, and the combined impedance. Peaking of the output impedance occurs around 250 MHz; since this is well beyond the closed-loop bandwidth of the amplifier, we attribute this to the inductive reactance of the source follower itself as  $C_{gs}$  shorts the gate and source at higher frequencies. This peaking can be damped by adding an equivalent series resistance (ESR)

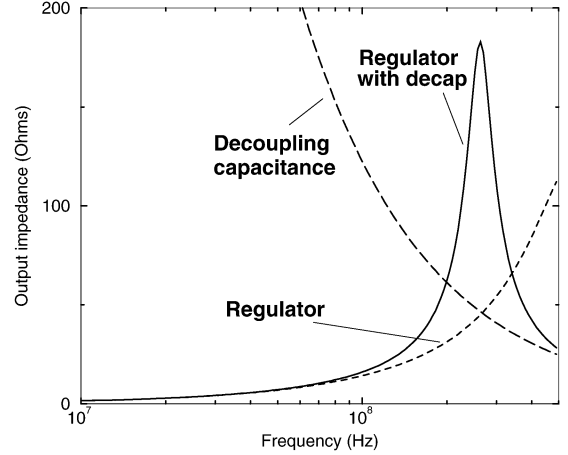


Fig. 5. Simulated output impedance of the regulator alone (without  $C_{\text{decap}}$ ), the decoupling capacitance alone, and the combined impedance.

with the decoupling capacitance or increasing the amount of decoupling capacitance, which is quite undersized in this case.

When the average current sourced (sunk) by the linear regulator exceeds 3 mA over an interval of approximately 150 ns, the **high (low)** signal is asserted to indicate that granules should be switched between domains to remove the imbalance. Since the multiplier consumes an average total current of approximately 22 mA from the 1.8-V supply, with an average current mismatch of 3 mA, the instantaneous efficiency degrades to 80%; granule switching, therefore, assures that the instantaneous efficiency is never worse than 80%. The fraction of the maximum mismatch current ( $3.5 \mu\text{A}$ ) integrated over  $C_{\text{int}} \cong 400$  fF for 150 ns yields appropriate comparator reference voltages of  $V_{\text{refh}} = 1.4$  V and  $V_{\text{refl}} = 0.4$  V. When one of the comparators is triggered, the switching control logic switches logic between domains.

The more current that must be sourced or sunk by the regulator to maintain regulation, the lower the efficiency of the dc-dc conversion. The “instantaneous” efficiency of the system is given by the ratio of the instantaneous power delivered to the load to the total instantaneous power:

$$\eta(t) = \frac{V_{\text{DD}} i_{\text{top}}(t) + V_{\text{int}} i_{\text{reg}}(t)}{V_{\text{DD}} i_{V_{\text{DD}}}(t)} \quad (4)$$

where  $i_{\text{top}}$  and  $i_{\text{reg}}$  are the currents flowing into the top domain from the  $V_{\text{DD}}$  supply and flowing from the output stage of the regulator, respectively, as shown in Fig. 1(b).  $i_{V_{\text{DD}}} (= i_{\text{reg-top}} + i_{\text{top}})$  is the total current flowing out of the  $V_{\text{DD}}$  supply. In addition, one can define a “running” efficiency given by the ratio of the total energy used for logic to the total energy delivered to the system over a time window from  $t_0$  to  $t$ :

$$\eta(t) = \frac{V_{\text{DD}} \int_{t_0}^t i_{\text{top}}(\tau) d\tau + V_{\text{int}} \int_{t_0}^t i_{\text{reg}}(\tau) d\tau}{V_{\text{DD}} \int_{t_0}^t i_{V_{\text{DD}}}(\tau) d\tau}. \quad (5)$$

## B. Logic With Switchable Granules and Full-Rail Interfaces

Each granule shares a set of *granule multiplexer transistors* (as shown in Fig. 3) in both the pull-up and pull-down networks which determine the domain assignment of a particular granule; these switches also allow a domain to be configured for full-rail

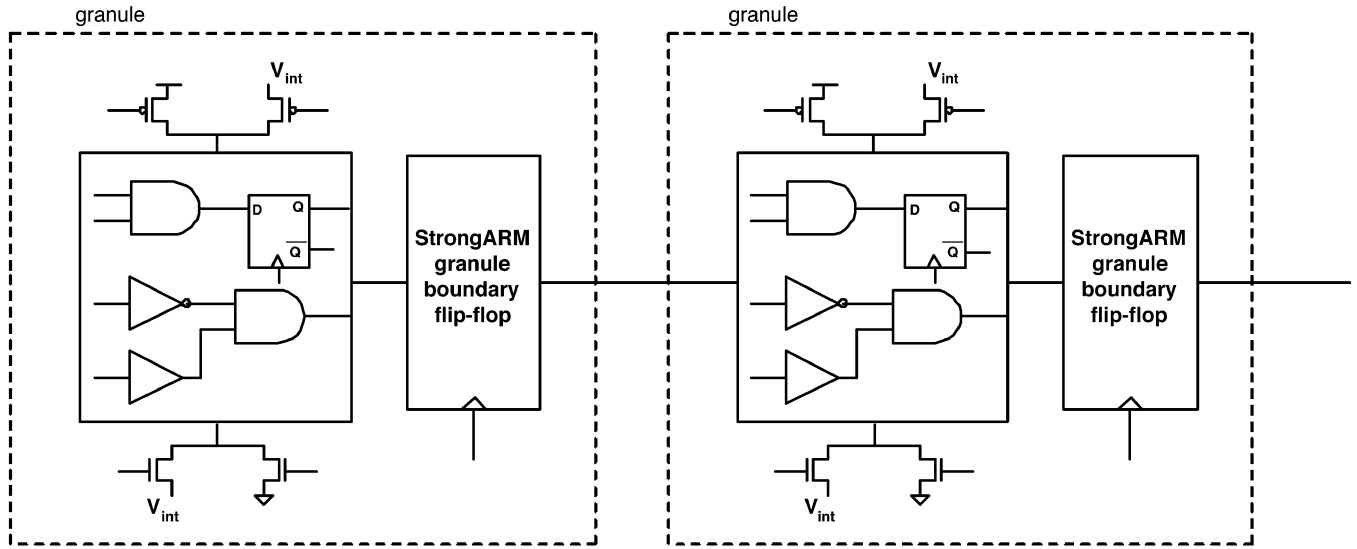


Fig. 6. Two granules, each representing a latch bounded block of logic. Granule multiplexer transistors connect the virtual supply and ground nodes of each granule to  $V_{DD}$ ,  $V_{int}$  or ground. The granule-boundary latches produce full-rail outputs and are implemented with gate-isolated sense-amplifier-based flip-flops.

operation. The switches are similar to the “sleep transistors” which can be employed to control standby power due to sub-threshold leakage and can also be used to perform this function if implemented with “high”  $V_T$  devices [13]. The drain nodes of these multiplexer transistor represent virtual supply and ground nodes.

Charge recycling dc-dc conversion does not work well in “normal” bulk CMOS because of body effect. Because their bodies are still tied to ground, nFETs in the upper voltage domains are heavily body affected. This problem is avoided in a triple-well process in which the nFETs are constructed in a p-well within an n-well.<sup>2</sup> These p-wells are then tied to the virtual ground of the granule; similarly, the pFET n-wells are tied to the virtual supply of the granule. The junction capacitance of these wells adds intrinsic decoupling to the virtual supply and ground nodes, improving power supply integrity for a given sleep-transistor width. Because of this well capacitance, each granule has a decoupling capacitance of approximately 0.5 pF on the virtual supply and ground nodes. The nFET (pFET) multiplexer transistors have a total gate width of approximately 400  $\mu\text{m}$  (800  $\mu\text{m}$ ) per granule, which is sufficient to keep  $V_{DS}$  less than 5% of the target supply voltage for the domain.

Granules are chosen to be latch-bounded logic partitions, as shown in Fig. 6. This allows the latches between granules to provide full-rail interfaces, negating the need for more complex level conversion between domains and easily allowing granules to be moved between domains without any changes to the interface circuit requirements. Gate-isolated sense-amplifier-based (GISA) flip-flops, such as those used on the StrongArm processor [14], are used for this purpose. By contrast, latches contained within a granule can be of any type and operate at the reduced supply levels of the granule. The switching energy of these full-rail interface latches will not scale when the logic is

configured to run at reduced supply, reducing the overall achievable power scaling.

### C. Granule-Switching Control Logic

Special consideration must go into the logic that controls the switching of granules between domains to guarantee system stability and ensure (because of the power overhead associated with switching granules) that switching occurs only when the linear regulator is providing too much current for an extended period of time. Switching granules between domains dissipates energy because of the power required to switch the capacitance of the gates of the (large) granule multiplexer transistors. In addition, the decoupling (well) capacitance on the virtual supply and ground nodes must be charged or discharged when a granule switches between domains.<sup>3</sup> The switching control logic is implemented using linear feedback shift register (LFSR) to “randomly” choose a granule to be switched. At most one granule can be “switched” every  $t_{int}$ .

It is interesting to compare the energy wasted due to a degraded efficiency with that consumed in switching a domain. For the maximum allowable mismatch current of 3 mA, 650 pJ is wasted in the 150-ns integration interval. In contrast, we estimate that approximately 3.5 pJ is consumed in switching a granule, including both the capacitance of the multiplexer transistors and the well capacitance that must be charged or discharged. As a result, for optimal system efficiency, much more frequent domain switching could be supported through a lower maximum allowable mismatch current.

The system is designed to continue operating during domain switching. This requires that, when a domain is switched, that the virtual supply and ground nodes charge (or discharge) quickly to their new state (and with comparable time constants) to limit the duration and extent of supply rail collapse. In our prototype system, these time constants are approximately 8 ps.

<sup>2</sup>Silicon-on-insulator (SOI) technology is also an attractive alternative for this technique, since the bodies float to the required voltage by action of the gate, source, and drain.

<sup>3</sup>Fortunately, however, most of the device capacitances and interconnect coupling capacitances between wires of the same domain have the character of floating capacitors, simply translating in voltage as domains are switched.

TABLE I  
SUMMARY OF MEASURED SYSTEM CHARACTERISTICS

Scaling power @ $V_{DD}$ and $f=400$ MHz		158 mW
Scaling power @ $V_{DD}/2$ and $f=200$ MHz		20 mW
Full-rail latch and clock power at $f=400$ MHz		108 mW
Full-rail latch and clock power at $f=200$ MHz		54 mW
Multiplier area		$1.4 \text{ mm}^2$
Switch control logic area		$0.0196 \text{ mm}^2$
Linear regulator	Area	$0.024 \text{ mm}^2$
	Bandwidth	130 MHz
	$I_{max}$	30 mA
	Current efficiency at $I_{max}$	96%

### III. MEASUREMENT RESULTS

Table I summarizes the measured characteristics of the multiplier and the linear regulator. The multiplier functions correctly running at both 1.8-V and at a charge-recycled 0.9-V supply; functionality is not affected by the dynamics of domain switching. In our model system, the regulator and switch overhead add only approximately 3%–4% to the area of the system.

A switching energy of approximately 108 pJ is associated with the full-rail boundary latches and associated full-rail components of the clock distribution that do not scale with supply voltage. This latch “overhead” can be expected to be inversely proportional to the granule size; it is relatively high in this model system because of the relatively small granule size chosen here (8 pF of total gate capacitance). We consider the efficiency of our system independent of this overhead because these components are not regulated at the  $V_{DD}/2$  supply.

In Fig. 7, we show how the measured efficiency of the system degrades with mismatch between the top and bottom domains. In this case, the domain-switching is disabled and granule assignments are “hard-wired.” The pairs of numbers in Fig. 7 denote the number of granules (out of a total of 16) assigned to the “top” and “bottom” domains. “8–8” is an approximately matched configuration, while “2–14” is heavily skewed. We note that the current demands of each granule are not the same, although the assignments noted achieve monotonicity by starting from a balanced configuration and achieving the skewed configurations by moving granules from one domain to the other. Energy efficiency exceeds 85% for the matched configuration in which approximately 10 mA of current are being consumed by both domains and approaches 50% when the domains are out of balance (and the linear regulator must source or sink all of the current for the logic). The linear regulator is easily able to maintain regulation even in the “2–14” case.

In Fig. 8, we show how the measured dynamics of the system automatically correcting for charge imbalance through granule switching. Fig. 8 shows the running efficiency of the system as given by (5). Before point A, the multiplier is running a set of pseudorandom input patterns. At point A, the four least significant bits of the multiplier are forced to zero. This causes the bottom domain to consume more charge than the top domain, degrading the running efficiency. The regulator must source approximately 4 mA to regulate the  $V_{int}$  node, mirrored and integrated. The integrated voltage is compared with  $V_{refh}$  and the

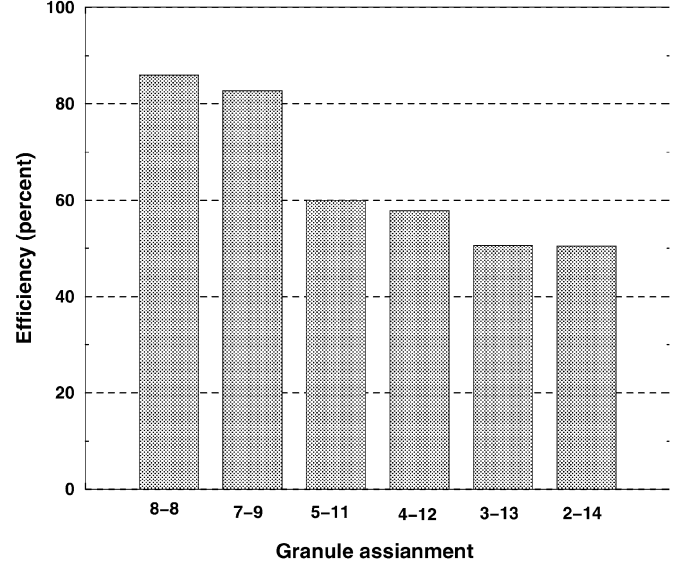


Fig. 7. Efficiency of the charge-recycling dc-dc conversion as a function of charge mismatch. The number of granules assigned to each domain is noted by the pair of numbers.

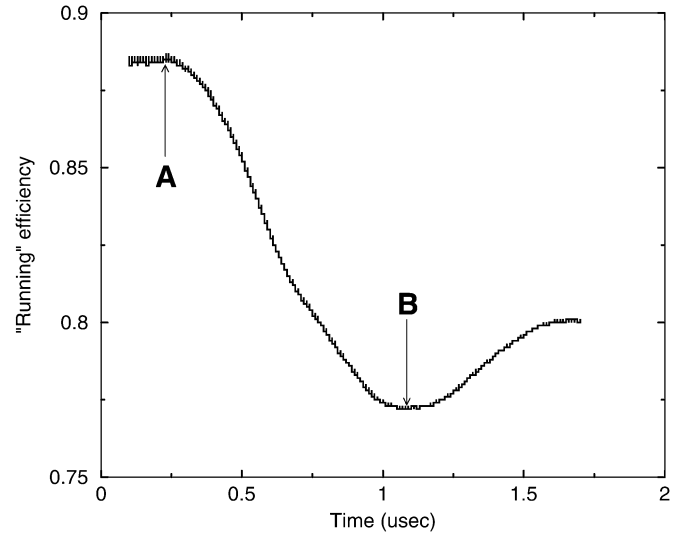


Fig. 8. Dynamics of charge-imbalance correction: “running” efficiency of the system beginning from a charge utilization mismatch (at A) and an “adjustment” through granule switching (at B).

comparator sends the high pulse at point B to activate the controller to switch granule from bottom to the top domain to correct the imbalance. The “running” efficiency over the window subsequently recovers to the 80% target level.

The chip also contains a 40-MHz, 6-bit flash analog-to-digital converter to monitor the regulation of the  $V_{int} = V_{DD}/2$ . Such noninvasive ADC-based on-chip monitoring of power supply integrity is becoming increasingly necessary. Attempting to buffer the power supply voltage for off-chip measurement is difficult because of the bandwidth limitation in the buffer amplifier, the load this amplifier presents to the supply node, and the noise introduced in the buffer. Fig. 9 shows the output of this power-supply-monitoring ADC, showing 10% regulation for the multiplier running with a random input pattern.

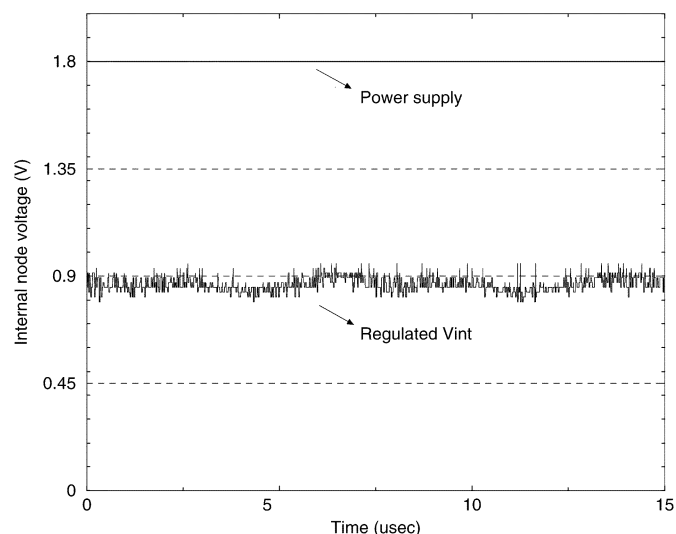


Fig. 9. Regulation of the  $V_{int}$  node as measured by the 6-bit on-chip flash ADC.

#### IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have described a new approach for on-chip dc-dc conversion using charge-balanced voltage islands running at fractions of the supply voltage. Charge “discarded” by one domain is “recycled” to supply energy to another. In a simple prototype system for a dynamic voltage scaling application, we achieved greater than 85% measured efficiency for  $V_{DD}/2$  conversion.

Future implementations will improve the linear regulator design to improve bandwidth and reduce the decoupling capacitance on internal nodes and will introduce true level-shifting circuits for domain interfaces.

#### ACKNOWLEDGMENT

The authors gratefully acknowledge T. Karnik and P. Hazucha of Intel for many helpful discussions.

#### REFERENCES

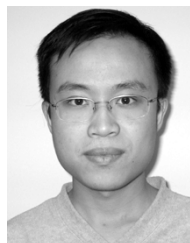
- [1] G. Qu *et al.*, “Energy minimization of system pipelines using multiple voltages,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS’99)*, 1999, pp. 362–365.
- [2] G. Semeraro, G. Magklis, R. Balasubramonian, D. H. Albonesi, S. Dwaradasa, and M. L. Scott, “Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling,” in *Proc. Int. Symp. High Performance Computer Architecture (HPCA-8)*, Feb. 2002, pp. 29–40.
- [3] T. D. Burd and R. W. Broderson, “Design issues for dynamic voltage scaling,” in *Proc. Int. Symp. Low Power Electronics and Design (ISLPED)*, 2000, pp. 9–14.
- [4] K. Usami and M. Horowitz, “Clustered voltage scaling technique for low-power design,” in *Proc. Workshop on Low Power Design*, 1995, pp. 3–8.
- [5] D. E. Lackey, P. S. Zuchowski, T. R. Bednar, D. W. Stout, S. W. Gould, and J. M. Cohn, “Managing power and performance for system-on-chip designs using voltage islands,” in *Proc. IEEE/ACM Int. Conf. Computer Aided-Design*, 2002, pp. 195–202.
- [6] G.-Y. Wei and M. Horowitz, “A fully digital, energy-efficient adaptive power-supply regulator,” *IEEE J. Solid-State Circuits*, vol. 35, no. 4, pp. 520–528, Apr. 2000.
- [7] D. Gardner, A. M. Crawford, and S. Wang, “High frequency (GHz) and low resistance integrated inductors using magnetic material,” in *Proc. IEEE Int. Interconnect Technology Conf.*, June 2001, pp. 101–103.

- [8] G. Patounakis, Y. W. Li, and K. L. Shepard, “A fully integrated on-chip dc-dc conversion and power management system,” *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 443–451, Mar. 2004.
- [9] S. Rajapandian, X. Zheng, and K. L. Shepard, “Charge-recycling voltage domains for energy-efficient low-voltage operation of digital CMOS circuits,” in *Proc. Int. Conf. Computer Design*, 2003, pp. 98–102.
- [10] H. Yamauchi, H. Akamatsu, and T. Fujita, “An asymptotically zero power charge-recycling bus architecture for battery-operated ultrahigh data rate ULSI’s,” *IEEE J. Solid-State Circuits*, vol. 30, no. 4, pp. 423–431, Apr. 1995.
- [11] B.-D. Yang and L.-S. Kim, “A low-power ROM using charge recycling and charge sharing techniques,” *IEEE J. Solid-State Circuits*, vol. 38, no. 4, pp. 641–653, Apr. 2003.
- [12] K. W. Mai, T. Mori, B. S. Amrutur, R. Ho, B. Wilburn, M. A. Horowitz, I. Fukushi, T. Izawa, and S. Mitarai, “Low-power SRAM Design using half-swing pulse-mode techniques,” *IEEE J. Solid-State Circuits*, vol. 33, no. 11, pp. 1659–1671, Nov. 1998.
- [13] J. T. Kao and A. P. Chandrakasan, “Dual-threshold voltage techniques for low-power digital circuits,” *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 1009–1018, Jul. 2000.
- [14] J. Montanaro *et al.*, “A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor,” *IEEE J. Solid-State Circuits*, vol. 31, no. 11, pp. 1703–1712, Nov. 1996.



**Saravanan Rajapandian** (S’04) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2001 and the M.S. degree in electrical engineering from Columbia University, New York, in 2003. He is currently working toward the Ph.D. degree at Columbia University.

His research interests include dc-dc converters, high-speed linear regulators, and low-power digital circuits.



**Zheng Xu** (S’04) received the B.S. degree in electrical engineering from the Cooper Union for Advancement of Art and Science, New York, in 2002 and the M.S. degree in electrical engineering from Columbia University, New York, in 2004. He is currently working toward the Ph.D. degree at Columbia University.

His research interests include low-power high-performance digital circuits and low-jitter low-power clock generation and distribution.



**Kenneth L. Shepard** (S’85–M’92–SM’03) received the B.S.E. degree from Princeton University, Princeton, NJ, in 1987 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988 and 1992, respectively.

From 1992 to 1997, he was a Research Staff Member and Manager in the VLSI Design Department at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he was responsible for the design methodology for IBM’s G4 S/390 microprocessors. Since 1997, he has been with

Columbia University, New York, where he is now an Associate Professor. He also served as Chief Technology Officer of CadMOS Design Technology, San Jose, CA, until its acquisition by Cadence Design Systems in 2001. His current research interests include design tools for advanced CMOS technology, on-chip test and measurement circuitry, low-power design techniques for digital signal processing, low-power intrachip communications, and CMOS imaging applied to biological applications.

Dr. Shepard received the Fannie and John Hertz Foundation Doctoral Thesis Prize in 1992. At IBM, he received Research Division Awards in 1995 and 1997. He was also the recipient of an NSF CAREER Award in 1998 and IBM University Partnership Awards from 1998 through 2002. He was also awarded the 1999 Distinguished Faculty Teaching Award from the Columbia Engineering School Alumni Association. He has been an Associate Editor of IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS and was the technical program chair and general chair for the 2002 and 2003 International Conference on Computer Design, respectively. He has served on the program committees for ICCAD, ISCAS, ISQED, GLS-VLSI, TAU, and ICCD.