

Stepped Supply Voltage Switching for Energy Constrained Systems

S. Khanna, K. Craig, Y. Shaksheer, S. Arrabi, J. Lach, and B. H. Calhoun
Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia
351 McCormick Rd, Charlottesville, Virginia
E-mail: sudhanshu@email.virginia.edu

Abstract

The energy consumed in switching the voltage on the power rail (V_{DD} switching energy) is a significant overhead in systems using Dynamic Voltage Scaling (DVS) and/or power gating. In this work we propose and demonstrate the use of Stepped Supply Voltage Switching (SVS) for reducing V_{DD} switching energy. We show the analysis, benefits, and overheads of using SVS for DSP algorithms implemented with voltage scalable adders and multipliers in simulation and silicon. SVS helps achieve 45% and 60% net savings in V_{DD} switching energy while switching from 0.3V to 1.2V using one and two intermediate steps respectively. Power supply noise, another concern in systems using power gating or DVS, is also analyzed. SVS helps reduce noise by over 40%.

Keywords

DVS, Power gating, Stepped charging, Adiabatic charging, Low power.

1. Introduction and Motivation

Many contemporary and emerging applications like portable multimedia, smart phones, and bio-medical devices impose strict constraints on energy consumption while also needing high performance for short bursts of time. Although technology scaling has provided raw performance gains and lower switched capacitance, increasing the battery life by lowering system energy consumption is still an ongoing effort.

Circuit techniques like power gating and dynamic voltage scaling (DVS) are employed in virtually all energy constrained systems to reduce leakage and dynamic power respectively. Blocks or individual components are power gated by disconnecting them from voltage supply or ground when the block is idle. This is implemented by inserting a PMOS header or footer NMOS between the block and its voltage supply or ground respectively. DVS reduces the voltage supply (V_{DD}) of the block when the block has timing slack prior to any deadline, thus reducing throughput linearly and energy quadratically. However, when the block needs high performance, the voltage rail must be charged back to its original high value. Voltage supplied to a block may be changed by using an on-chip or off-chip regulator. If two different blocks need different voltages at the same time, they have to be controlled by separate regulators or multi-output regulators. Regulators consume significant area and power, which limits the number of voltage islands that a system designer can create. Such an implementation of DVS with different voltage islands [1][2] using their own regulators is commonly referred to as multi- V_{DD} . An alternative is to route a small number of voltage supplies

across the chip and connect all blocks using DVS to each of those supplies with headers, as shown in Figure 1 [3][4][5]. Then, for a given block turn on only the header corresponding the voltage supply needed, and keep the other headers off. [3] refers to this implementation of DVS as Panoptic DVS (PDVS).

Regardless of the type of implementation of DVS, V_{DD} switching energy is an issue that must be addressed. This issue exists in systems using power gating as well. V_{DD} switching energy overhead imposes a restriction on how often DVS or power gating can be used. Since there is energy overhead to bring a block out of its low power state (power gated or low voltage DVS setting), the block must remain in its low power state for a duration long enough that offsets the V_{DD} switching energy overhead. In this work we show how Stepped Supply Voltage Switching (SVS) can help reduce V_{DD} switching energy and therefore allow systems to switch into lower power states more often. We introduce and analyze SVS in Section II and report simulation and measurement results in Section III.

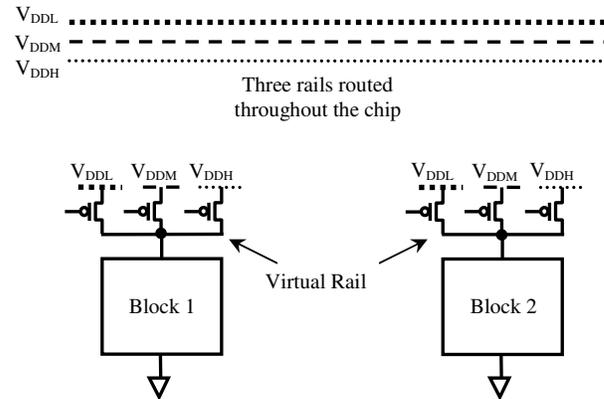


Figure 1: DVS implementing using headers switches and a set of shared V_{DD} s routed throughout the chip. Headers can be toggled dynamically.

Another concern in designs using power gating or DVS is power supply noise. The sudden current drawn by the local power grid of a block coming out of its power gated state results in $L \frac{di}{dt}$ and dynamic IR voltage drops on the higher level voltage grid. This noise impacts other blocks using the same higher level grid. We show how SVS helps reduce power supply noise with no dedicated circuitry for noise reduction in section IV.

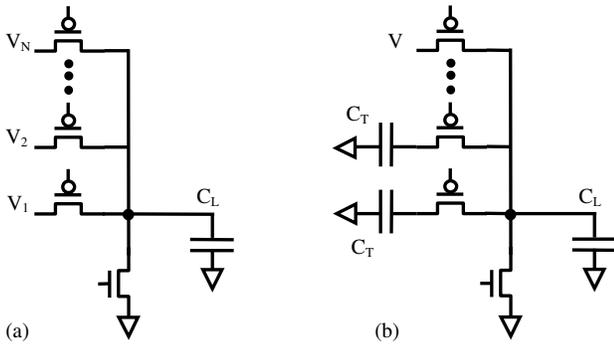


Figure 2: Concept of Stepwise Charging (a) Load Capacitor C_L being charged from zero to V in $(N-1)$ intermediate steps (b) intermediate supplies replaced by tank capacitors

Before going in to the details of SVS, we describe stepwise charging. Stepwise Charging [6], an inductor-free form of adiabatic charging, uses a bank of voltage supplies with uniformly distributed voltages to charge a load capacitor from ground to the final voltage needed. The technique has been proposed as an energy efficient way to design drivers with large capacitive loads. Figure 2a shows a simplified implementation of stepwise charging with a total of N supplies. To charge the load capacitor, each supply is switched on and off in an ascending fashion. The opposite order is taken while discharging. The net energy consumed to charge and discharge the capacitor is $\frac{CV^2}{N}$ [6]. Thus stepwise charging theoretically provides linear saving in energy, which asymptotically reaches zero V_{DD} switching energy consumption. This type of approach has been applied to save power in charge sharing schemes on buses and in SRAM bitlines, for example, but not to power supplies to our knowledge. The above formula assumes that current going into a voltage supply results in energy “recovered”, an assumption that we will revisit later in this work. To avoid having numerous voltage supplies, [6] proposes the use of numerous large tank capacitors. Also, the time overhead of rippling up and down all voltage supplies at *every clock edge* (as [6] is proposed for charging circuit outputs) results in significant performance overhead.

We apply the concept of stepwise charging for charging and discharging the virtual power rail during power gating and DVS related mode changes in SVS. SVS ripples through voltage supplies only during mode changes, thus limiting the performance impact at the system level. We do not use multiple large tank capacitors but address the issue of multiple voltage supplies in an alternative fashion by taking advantage of components already in many DVS designs.

Section II introduces SVS and analyzes its benefits, overheads, and applications in systems. Section III details the simulation and measurement results showing SVS benefits and overheads for adders, multipliers, and DSP data flow graphs. Section IV highlights power supply noise issues, and shows how SVS is an effective low overhead technique to reduce noise. Section V lists the key results and further research directions.

2. SVS: Theoretical Benefits, Overheads, and Applications

The work in [3]-[5] describes implementations of DVS using multiple voltage supplies and headers, also known as PDVS. As mentioned briefly in the previous section, PDVS systems have a small number of voltage supplies routed throughout the chip. Circuit blocks that experience variable workloads or idle modes are connected to each of these supplies using headers. Figure 1 shows the concept behind PDVS using three voltage supplies and two independent blocks that can be assigned to different voltages. Depending upon the current state of the system, a block can be dynamically connected to any one of the voltage supplies. If needed, all headers can be switched off, and the block can be power gated. In this manner, PDVS helps implement DVS and power gating on a per-block basis without the need for expensive regulators for each block. The cost of generating multiple supply voltages is amortized over the entire system. Moreover, systems deploying traditional DVS and multi- V_{DD} also generate and route multiple supply voltages and thus reduce the overhead of implementing PDVS over other schemes [1][2].

We leverage the presence of multiple supply voltages and headers in PDVS to apply SVS without the use of large tank capacitors or any additional overheads. Figure 3 demonstrates the use of SVS from V_{DDL} to V_{DDH} or from a power gated state to V_{DDH} , the intermediate voltage supply header (V_{DDM}) is pulsed on. While going back to V_{DDL} or to a power gated state, the V_{DDM} header is again pulsed on. It is important to note that the intermediate supply V_{DDM} and associated control signals are not routed especially to enable SVS. Their presence in PDVS systems is simply leveraged by SVS to reduce $V_{DDL} - V_{DDH}$ switching energy.

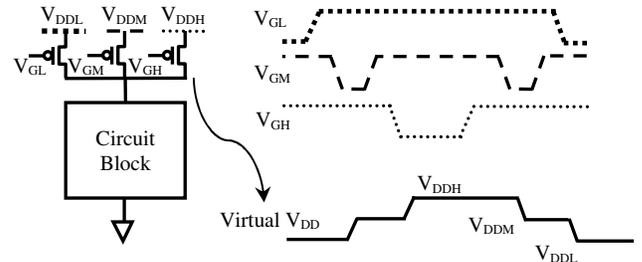


Figure 3: SVS timing based on implementation in Figure 1. V_{GM} pulses during both V_{DDL} to V_{DDH} and V_{DDH} to V_{DDL} supply voltage switches.

We now analyze the energy saved by SVS while switching the block supply from V_{DDL} to V_{DDH} with a single intermediate supply V_{DDM} midway between V_{DDL} and V_{DDH} . The analysis is divided into two sections, depending on whether current going into a supply is assumed to “dissipate” (worst-case) or “recover” (best-case). Take the case when the virtual rail switches from V_{DDH} to V_{DDL} . In this scenario, a current flows from the virtual rail into V_{DDL} . The argument that current going into V_{DDL} can not be recovered has basis in the fact that only rechargeable

batteries have the capability of generating energy when given current. On-chip regulators would simply shunt current flowing into them to ground. On the other hand, it can be argued that the charge flowing towards V_{DDL} will actually not go into the V_{DDL} regulator, but will be stored temporarily in the decoupling capacitor on the V_{DDL} rail or the capacitor at the output of the regulator and will be ultimately used by other circuits operating from the V_{DDL} rail at the same time. In this scenario, the current would be used to do other useful operations, and would actually be recovered in principle. If the components are active, then the actual energy benefit will be closer to the best-case (which we assume for our simulations) since the charge pushed onto the decoupling capacitor can be used by other circuits before it is shunted to ground by an off chip regulator. However, the theoretical analysis that follows is done for both cases and shows the upper and lower bounds of energy saved by SVS.

In equations (1) to (5) for worst-case theoretical savings below, C_L represents the virtual V_{DD} capacitance, which includes the parasitic rail capacitance and the total gate capacitance of all nodes that are '1'. The analysis is based on transitions shown in Figure 3.

$$E_{VDDL-VDDM} = C_L V_{DDM} (V_{DDM} - V_{DDL}) \quad (1)$$

$$E_{VDDM-VDDH} = C_L V_{DDH} (V_{DDH} - V_{DDM}) \quad (2)$$

$$E_{VDDL-VDDM-VDDH} = C_L [V_{DDM} (V_{DDM} - V_{DDL}) + V_{DDH} (V_{DDH} - V_{DDM})] \quad (3)$$

$$E_{VDDL-VDDH} = C_L V_{DDH} (V_{DDH} - V_{DDL}) \quad (4)$$

$$E_{SAVED,L-H} = C_L (V_{DDH} - V_{DDM}) (V_{DDM} - V_{DDL}) \quad (5)$$

$E_{SAVED,L-H}$ is the energy saved by SVS in the transition from V_{DDL} to V_{DDH} . When we transition back from V_{DDH} to V_{DDL} , the analysis is as follows if current going into a supply can be considered recovered (best-case).

$$E_{VDDH-VDDM} = -C_L V_{DDH} (V_{DDH} - V_{DDM}) \quad (6)$$

$$E_{VDDM-VDDL} = -C_L V_{DDM} (V_{DDM} - V_{DDL}) \quad (7)$$

$$E_{VDDH-VDDM-VDDL} = -C_L [V_{DDM} (V_{DDM} - V_{DDL}) + V_{DDH} (V_{DDH} - V_{DDM})] \quad (8)$$

$$E_{VDDH-VDDL} = -C_L V_{DDH} (V_{DDH} - V_{DDL}) \quad (9)$$

$$E_{SAVED,H-L} = C_L (V_{DDH} - V_{DDM}) (V_{DDM} - V_{DDL}) \quad (10)$$

$$\%E_{SAVED,WC} = E_{SAVED,L-H} / E_{VDDL-VDDH} \quad (11)$$

$$\%E_{SAVED,BC} = (E_{SAVED,L-H} + E_{SAVED,H-L}) / (E_{VDDL-VDDH} + E_{VDDH-VDDL}) \quad (12)$$

Both $E_{SAVED,L-H}$ (5) and $E_{SAVED,H-L}$ (10) get maximized when V_{DDM} is midway between V_{DDL} and V_{DDH} .

The percentage saving due to SVS in the best-case scenario (12) is independent of V_{DDL} and V_{DDH} as long as V_{DDM} is midway between them. With one intermediate step, the saving is 50%, and with two intermediate steps, the saving is a substantial 66%.

The percentage savings due to SVS in the worst-case scenario (11) is dependent on V_{DDL} , V_{DDM} and V_{DDH} even if V_{DDM} is midway. Figure 4 plots the upper and lower

theoretical bounds of energy saved by SVS as a function of V_{DDM} , with V_{DDL} and V_{DDH} set at 0.3V and 1.2V respectively. The curves are based on (11) and (12). The best-case and worst-case scenarios are the upper and lower bounds of energy saved by SVS, assuming no overhead for switching. There are practical limitations imposed by overheads, like the header gate capacitance, which are included in simulation and measurements results in the next section. The energy benefit is normalized with respect to V_{DD} switching energy with no intermediate step.

2.1 Overheads and Limitations

The above analysis demonstrates the theoretical savings of SVS. In a practical scenario, there are overheads and limitations that must be considered. The first (and largest) overhead is from the energy needed to switch the intermediate voltage headers. The number of intermediate voltages available and their values are fixed by DVS related calculations. An additional voltage is only included if it lowers the system energy through DVS. It has been found that a higher a variability of workload results in a higher energy benefit for having an addition rail [4]. The values of the voltage supplies that yield optimal savings in DVS are close to being uniformly distributed [4], and that works well for SVS too, as we found in the theoretical analysis.

SVS also results in linear overhead in V_{DD} switching delay when we ripple through the different voltage supplies rather than taking a single jump. The performance impact of stepwise charging when used in normal circuit operation [6] (every time a circuit output goes high) can be high, but our use of SVS for only V_{DD} switching limits the impact of the performance penalty. Moreover, the fact that there is available slack during a V_{DDL} operation means that performance can be traded off for additional energy savings.

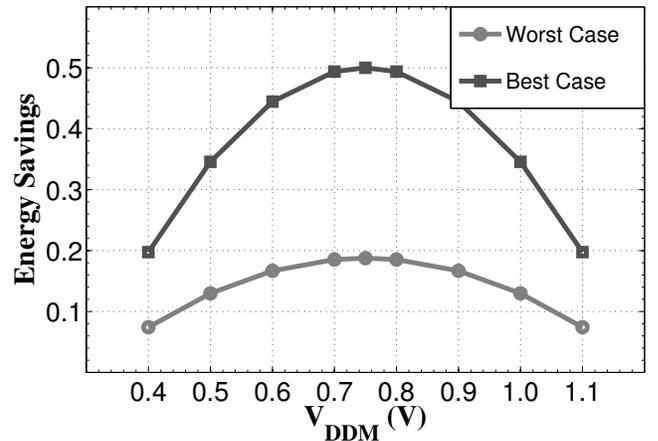


Figure 4: Upper and lower bounds of theoretical energy saved by SVS as a function of V_{DDM} with V_{DDL} and V_{DDH} at 0.3V and 1.2V respectively. Energy saving is normalized with respect to V_{DD} switching energy with no intermediate steps.

2.2 Applications for SVS

In this work we use adders and multipliers as vehicles to demonstrate SVS. However, SVS can be applied to any digital block using DVS or power gating. A possible application is in [5], where a 167-core processor design is proposed that uses DVS with headers on a per-core basis. For a block as large as a core, the supply rail capacitance (C_L) will be in 100s of pF, and (5) shows larger C_L gives larger savings. We also show in the next section that the SVS benefit/overhead ratio increases as the block size increases.

SVS can also be used to lower energy in memory design. With bit-cell leakage dominating both standby and active power in sub-65nm SRAMs, DVS is often used on SRAM bit-cell arrays. [2] divides the bit-cell array of a SRAM macro into banks, and then only activates the bank being accessed. The other banks remain at a lower V_{DD} , simply retaining their state. However, this is only beneficial if the V_{DD} switching overhead is lower than the leakage energy saved. This results in the notion of a breakeven standby time, and if the SRAM is accessed at a rate faster than the breakeven time, this technique is not beneficial. Using SVS can lower this breakeven time, making voltage scaling useful for more application scenarios.

3. Simulation and Measurement Results

In this section we quantify the benefits and overheads of SVS with simulation and test chip measurement (die photo in Figure 5) results from adders and multipliers connected as shown in Figure 1. A commercial 90nm technology was used for the implementation.

3.1 Implementing SVS on Voltage Scalable Adders and Multipliers

V_{DD} switching energy is a function of the initial and final voltages and the virtual rail capacitance. The virtual rail capacitance is the sum of parasitic wire capacitance of the virtual rail, source capacitance of all PMOSs in the circuit, header drain capacitance, and half of the circuit gate capacitance. The gate capacitance gets included because all circuit PMOSs that are ‘on’ will also pull-up the gates their drains are connected to when we switch the voltage supply.

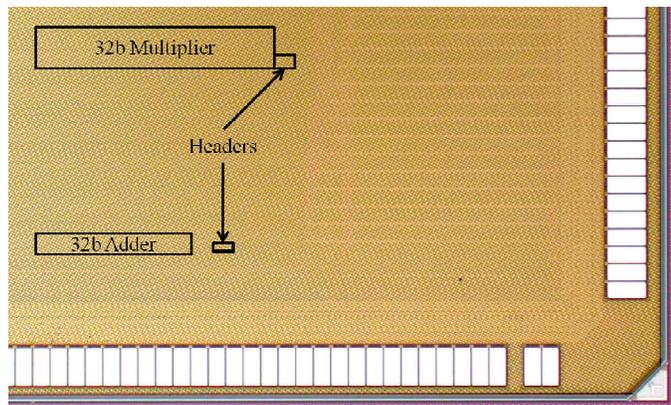


Figure 5: Die photo of the 90 nm test chip.

Figure 6 shows the energy benefit of SVS as a function of V_{DDL} , with V_{DDH} fixed at 1.2V, for a 32b Kogge Stone adder and a 32b Baugh Wooley multiplier. The setup is similar to Figure 3, except that the number of intermediate steps and headers is varied. The circuit block n-well is tied to the virtual rail, thus reducing virtual rail capacitance by shorting the source to n-well capacitance of the PMOSs in the circuit block. Energy saved is normalized with respect to V_{DD} switching energy with no intermediate steps. Intermediate voltages are uniformly distributed between V_{DDL} and V_{DDH} . Unlike the ideal trend indicated by (12) in which the percentage benefit is independent of V_{DDL} , the actual benefit of SVS decreases as V_{DDL} increases from 35% at 0.3V to 15% at 0.7V in the one-step adder case, and from 45% at 0.3V to 30% at 0.8V in the one-step multiplier case. For V_{DDL} greater than 0.8V, there is actually a loss in using SVS for the adder. This is because of the header switching energy overhead of SVS, which is the energy consumed in switching the gates of the additional ‘intermediate’ header(s). As V_{DDL} increases, the V_{DD} switching energy (and the amount saved by SVS) decreases making the header

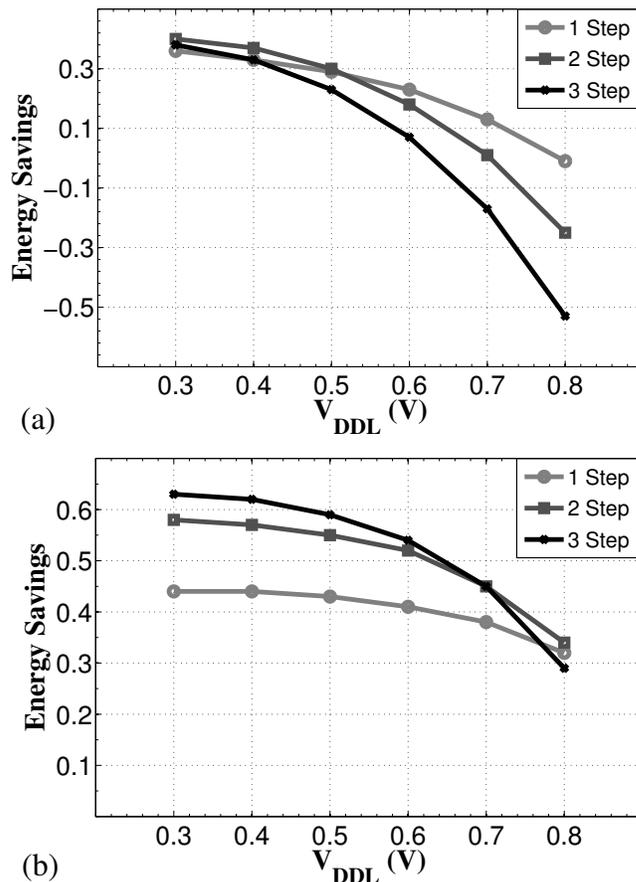


Figure 6: Simulation results showing energy benefit vs. V_{DDL} with V_{DDH} fixed at 1.2V. For one, two and three intermediate steps between V_{DDL} and V_{DDH} . (a) 32b adder (b) 32b multiplier. Energy saving is normalized with respect to V_{DD} switching energy with no intermediate steps.

overhead more significance as a percentage. This energy saved vs. V_{DDL} trend shows that SVS is most useful for drastic mode changes like coming out of power gating, or from a low-voltage DVS mode to a burst of high performance V_{DDH} operation. Also, when going from 0.7V to 1.2V for the adder, SVS is beneficial if one intermediate step is taken, but has loss if two intermediate steps are taken. Again, this is because two-intermediate-step case has higher header switching energy overhead than the one-intermediate-step case. Parasitic layout capacitances for header gates and wires, and virtual voltage rails were extracted from the layout and included in the simulations. For the adder, header and rail parasitic wire capacitors were 40fF and 1.5pF respectively. For the multiplier, these were 40fF and 10pF respectively. Measurements of a 90nm test chip confirm these results. Figure 7 shows the comparison of measured and simulated values of multiplier V_{DD} switching energy with and without SVS. The measured benefit of SVS matches simulated benefit within 5%.

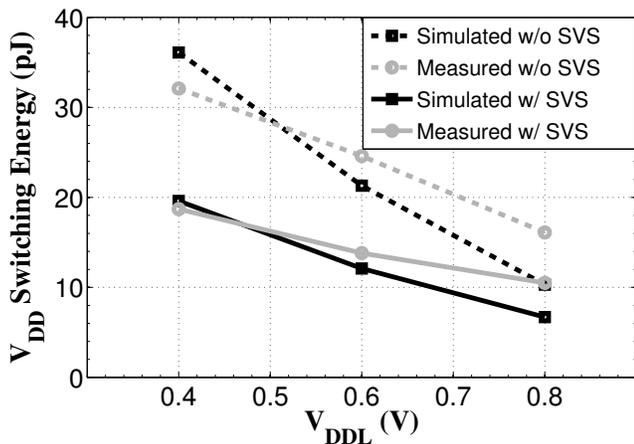


Figure 7: Measured and simulated values of V_{DD} switching energy for 32b multiplier as a function of V_{DDL} . V_{DDH} is fixed at 1.2V. V_{DDM} is midway between V_{DDL} and V_{DDH} .

The adder and multiplier give different energy benefit numbers. Particularly, the multiplier shows much higher savings across the entire V_{DDL} range. The reason is that even though the multiplier has a larger header than the adder, its rail capacitance is much larger than the adder. In other words, the multiplier has smaller percentage header energy overhead than the adder. The header sizing for a block is proportional to the total gate capacitance of the block and the activity factor of the block, which signifies the percentage of nodes that switch every clock cycle. As a block grows in size, it starts incorporating sub-blocks that don't draw current at the same time. In other words, the activity of a block tends to decrease with its size. Hence, while the rail capacitance (C_{rail}) grows at the same rate as the block size, the header size (and header gate capacitance C_{header}) grows at a smaller rate. Thus, the ratio of C_{rail}/C_{header} increases with block size. The benefit of SVS increases with this ratio and hence with the block size, as shown in Figure 8. This figure shows simulation results for energy benefit of

SVS as a function of C_{rail}/C_{header} . Simulation results show that one- intermediate-step SVS has 44% benefit when C_{rail}/C_{header} is 200, but only 27% benefit when the ratio increases to 50.

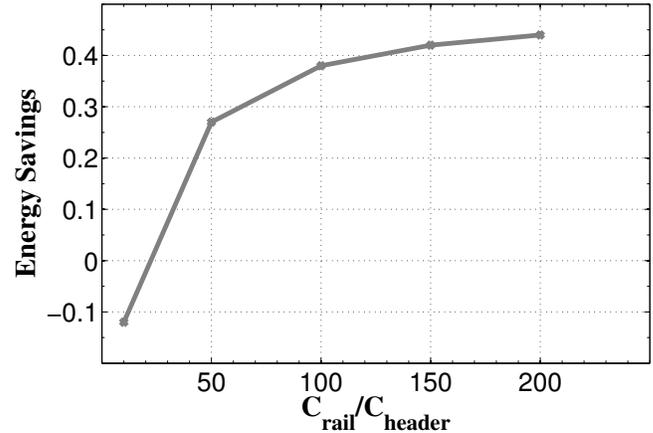


Figure 8: Energy benefit vs. C_{rail}/C_{header} with V_{DDL} , V_{DDM} and V_{DDH} fixed at 0.6V, 0.9V and 1.2V. One intermediate step SVS. Energy saving is normalized with respect to V_{DD} switching energy with no intermediate steps.

Looking at systems using SVS with fine granularity in space and time, we use the SVS adders and multipliers in dataflow graphs (DFGs) that implement FIR, Elliptical Filter, and DiffeQ benchmarks. Components that have slack or are idle are switched to V_{DDL} . The three voltage rails as shown in Figure 1 are at 0.6V, 0.9V, and 1.2V in these simulations. Across the three DFGs, the benefit is 4-6%. The savings are limited because multipliers, which dominate the system energy, are on the critical path and rarely switch from V_{DDL} to V_{DDH} . However, for systems with variable workload all components of the DFG (including the multipliers) would need to switch rates and supply voltages to match the workload. Then, V_{DD} switching energy becomes more significant, and system level savings of SVS are higher. DFGs are examples where DVS is used on a fine granularity. Also, if DVS is applied at a coarser granularity, (e.g. per-core), the benefit of SVS would increase, especially for drastic mode changes like coming out of power gating.

4. SVS and Power Supply Noise

Power supply noise is a well known concern in designs employing power gating [7] and DVS [5]. Whenever a block comes out of power gating, or moves from a lower V_{DD} to a higher V_{DD} , it draws a large current in a short period of time. $L \frac{di}{dt}$ drop across the package inductance and dynamic IR drop across the supply rail resistance result in transient voltage droops on the supply. This can impact other blocks operating on the higher V_{DD} at the same time. Examples of techniques used to reduce power supply noise are split headers [7], where the power gating header is divided into fingers which are turned on in a staggered manner. Another technique is to apply a slow ramp to the header such that the

header turns on slowly. [5] introduces a time delay between the turn-off of one header and the turn-on of the next header.

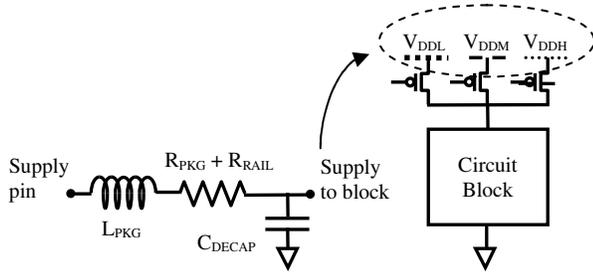


Figure 9: Circuit block with headers and supply rail RLC. While switching from V_{DDL} to V_{DDH} , noise is generated on V_{DDH} .

SVS can be used an effective technique to lower power supply noise. By going from V_{DDL} (or a power gated state) to V_{DDH} in intermediate steps (as shown in Figure 3), SVS not only has inherent staged turn-on, but SVS also lowers noise because the energy (and hence current) being drawn from the supply is lower in the first place (as described in the previous section). Moreover, for a system employing headers for DVS, the benefit of lower power supply noise can be achieved with no additional circuitry (because the headers and rails are needed anyways for DVS).

We evaluated the noise benefit of SVS by simulating a 32b adder along with characteristic package inductance (10nH), rail and package resistance (20ohm), and on-chip decoupling capacitance (10pF), as shown in Figure 9. Parasitic rail and header gate capacitances were included in the simulation. Peak to peak noise values with and without SVS are reported in Table 1. The table shows noise values on V_{DDH} when the adder went from V_{DDL} to V_{DDH} with one step at V_{DDM} , as shown in Figure 3. As before, value of V_{DDM} was taken to be between V_{DDL} and V_{DDH} . “Without SVS” refers to conventional switching from V_{DDL} directly to V_{DDH} . For all values of V_{DDL} , SVS helps reduce power supply noise by over 40%. If more intermediate steps are added, the benefit increases even further.

Table 1: Peak to peak noise on V_{DDH} at when voltage is switched from V_{DDL} to V_{DDH} , and V_{DDL} to V_{DDM} to V_{DDH} . $V_{DDH} = 1.2V$, V_{DDM} midway between V_{DDL} and V_{DDH} .

V_{DDL}	V_{DDL} to V_{DDM} to V_{DDH} (With SVS)	V_{DDL} to V_{DDH} (Without SVS)
0.3V	80 mV	137 mV
0.6V	55 mV	105 mV
0.9V	33 mV	58 mV

5. Key Results and Further Work

V_{DD} switching energy and power supply noise are two critical metrics in systems using DVS and power gating. In this paper we have demonstrated a technique called SVS that leverages existing DVS infrastructure of headers and voltage rails to lower V_{DD} switching energy and power

supply noise. We have shown theoretical analysis that proves the mathematical basis for the energy savings. Simulations and measurement results confirm that V_{DD} switching energy (energy consumed in switching from one V_{DD} to another) is lowered by a factor of 45% to 65% for a 32b multiplier, and by 35% for a 32b adder. As the block size on which DVS is applied grows, the benefits of SVS increase. To take advantage of this trend, application of SVS to per-core DVS in multi-core processors and to memory design is described. Finally, it is shown that SVS helps reduce power supply noise by 40% as compared to conventional power gating or DVS.

6. Acknowledgements

This work was funded in part by a DARPA seedling grant.

7. References

- [1] B. Nam et al, "A 52.4mW 3D Graphics Processor with 141Mvertices/s Vertex Shader and 3 Power Domains of Dynamic Voltage and Frequency Scaling," ISSCC 2007.
- [2] G. Gammie et al, "A 45nm 3.5G Baseband-and-Multimedia Application Processor using Adaptive Body-Bias and Ultra-Low-Power Techniques," ISSCC 2008.
- [3] M. Putic et al, "Panoptic DVS: A fine-grained dynamic voltage scaling framework for energy scalable CMOS design," ICCD 2009.
- [4] D. Chen et al, "Optimal module and voltage assignment for low-power," ASP-DAC 2005.
- [5] D. Truong et al, "A 167-processor 65 nm computational platform with per-processor dynamic supply voltage and dynamic clock frequency scaling," IEEE Symposium on VLSI Circuits 2008.
- [6] L.J. Svensson et al, "Driving a capacitive load without dissipating fCV2," IEEE Symposium on Low Power Electronics 1994.
- [7] S. Kim et al, "Understanding and minimizing ground bounce during mode transition of power gating structures," ISLPED 2003.