

# A Low Voltage SRAM Using Resonant Supply Boosting

Rajiv V. Joshi, *Fellow, IEEE*, Matthew M. Ziegler, *Member, IEEE*, and Holger Wetter

**Abstract**—This paper presents a novel resonating inductor-based supply boosting scheme for low-voltage static random-access memories and logic in deep 14-nm silicon on insulator (SOI) FinFET technologies. The technique combines capacitive (C) and inductive (L) boosting for the first time. Simulation and measured hardware results from a 14-nm test chip show that this new technique is able to improve  $V_{min}$  (down to 0.3 V), functional yield, and access time, when compared with designs with or without capacitive-boosted supplies. Simulations also reveal the optimal combinations of “L” and “C” needed for each  $V_{dd}$  to achieve minimal boost voltage, where the static random-access memory can be rendered fully functional in the absence of any assist circuitry. Furthermore, the resonant supply provides power savings compared with a boosted supply alone.

**Index Terms**—Boosted supply, FinFet, inductors, low-power electronics, low  $V_{min}$ , 8T, static random-access memory (SRAM).

## I. INTRODUCTION

MBEDDED static random-access memories (SRAMs) are still the workhorse of the VLSI industry. Being the smallest in size and the densest structures on a chip, SRAM represents a critical portion of most microprocessor and system on chip designs. Decades of effort has been spent lowering power and maintaining functionality in these designs as the technology scales. There are various techniques listed in the literature to lower the power of SRAMs. Some of the notable ones are static and dynamic dual power supply, usage of multi- $V_t$  devices for the cell, shorter bitlines, usage of write-assist techniques, the addition of transistors in an SRAM cell, and the usage of technology, e.g., silicon on insulator (SOI) versus bulk and nonplanar versus planar. The power supply remains one of the key knobs for reducing power consumption; however, lowering the power supply requires a balancing act for maintaining  $V_{min}$  in order to achieve functionality as well as the performance of SRAM cells and latches. Functionality here can be best described in terms of cell write ability, read stability, and data retention, which strongly depend on process, voltage, and temperature variations. Furthermore, for high-performance processors, it is important to maintain wide range

Manuscript received July 27, 2016; revised September 24, 2016; accepted November 1, 2016. Date of publication December 28, 2016; date of current version March 3, 2017. This paper was approved by Associate Editor Vivek De. This work has been partially sponsored by the DARPA Microsystems Technology Office (MTO), under contract no. HR0011-13-C-0022.

R. V. Joshi and M. M. Ziegler are with IBM, Research, Yorktown Heights, NY 10598 USA (e-mail: rvjoshi@us.ibm.com).

H. Wetter is with IBM, STG, 71032 Böblingen, Germany.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2016.2628772

of operation from low to high  $V_{dd}$  to achieve the optimal power and performance operating point for the given situation.

The operation of SRAM arrays at near threshold voltages still poses a challenge due to variability and functional yield difficulties when lowering supply voltages. Researchers have proposed techniques to improve stability for conventional 6T/8T cells, which include dual static [1], [2] as well as dynamic power supply [3]–[10], boosting techniques using charge pumps [6] and assist techniques [4]–[15] targeting specifically arrays. However, many of these techniques result in area, power, and/or supply voltage penalties. Recent trends show that static and separate power supplies are commonly used for arrays. The approach of using a static dual supply for the SRAM cell, the wordline, and/or the logic comes at the cost of an extra power supply that adds complexity to the design as well as power. As an alternative to static dual supply, a charge pump methodology was proposed to boost only the wordline [6]; however, extending this approach to the entire array would increase area and power. Likewise, the wordline may be dynamically boosted supply with a 50% duty cycle [9], [10], [12]. To alleviate some of the power consumption overhead of a 50% duty cycle boost, “a step-down” dynamic wordline boost is applied to only wordline drivers with a much reduced duty cycle [7]. While this step-down approach reduces power, the lower voltages may lead to SRAM arrays functionality issues, which get worse as the desired beta and gamma ratios of SRAM cells cannot be achieved in scaled technologies and especially in FinFET technologies.

Integrated read- and write-assist techniques utilizing wordline under drive and cell voltage collapse are often used for 6T SRAM macros [5]. However, with these techniques, the demonstrated voltage retention time is reduced to 1.6–4.2 ns depending on the threshold voltage of the pFET. Write-assist techniques utilizing reduced bitline voltages [11] or negative bitlines [13] are often used for achieving write ability, although they do not help to improve performance especially when  $V_{min}$  is lowered. Furthermore, negative boost swings for an individual or a group of bitlines need assist circuits, which come as a power cost [13]. Usage of dynamic voltage collapse-assist techniques also helps write ability; however, cell operation becomes difficult at extremely low voltages [5].

Many of the above-mentioned techniques employ planar technologies. The recent advent of FinFET technology and its application to processors have helped to lower operating voltages slightly [16], [17]. Although FinFET technology is attractive for low-power applications, many new circuit

TABLE I  
COMPARISON WITH PRIOR WORKS

Reference	SRAM Density (Mb/mm <sup>2</sup> )	V <sub>min</sub> (V)	Type/Tech/Size	Area	Comments/Power
[5] Karl <i>et al.</i>	11.1	0.6	6T/14nm/84Mb	7.5mm <sup>2</sup>	Capacitive charge sharing Write assist
[6] Rooseleer <i>et al.</i>	0.0316	1.2	6T/40nm/256Kb	0.19mm <sup>2</sup>	29pJ/access, 454 MHz, 114fJ/bit
[9] Kulkarni <i>et al.</i>	0.403	Improvement by 180mV	8T/22nm/12KB	0.238mm <sup>2</sup>	Voltage collapse+ coupling wordline
[13] Pilo <i>et al.</i>	0.021	0.7	6T/22nm/512Kb	(360 x 267um <sup>2</sup> ) 0.09612mm <sup>2</sup>	1.35Ax1.1V, 8Mb Fine granularity power gating cell
[20] Koo <i>et al.</i>	5.1	0.495 0.56@-10C, 460MHz	8T/14nm/512Kb	2.266mm <sup>2</sup>	2.25Mb, Cell voltage collapse, write assist bitlines
<b>This Work</b>	2.21	0.3	8T/14nm/36Kb	(155um x 105um <sup>2</sup> ) 0.0163mm <sup>2</sup>	Novel Capacitive + Inductive coupling (Full macro)

innovations are still needed to lower  $V_{min}$  for memory as well as logic. Also, many of the prior techniques listed previously are localized to either wordlines or arrays, and do not fully resolve the simultaneous requirements for functionality, performance, yield, and  $V_{min}$  of SRAMs. Although much of the work reported in advanced technologies is on bulk technology, SOI offers many attractive qualities from a technology perspective, including vertical etched fin profiles and soft-error rate advantages over bulk. In this paper, we exploit 14-nm SOI FinFET technology as well as novel circuits and interconnect structures to lower the operating voltage while maintaining performance at higher voltage. We introduce two novel concepts for on-demand power supply boosting: 1) capacitive coupling with interconnects and 2) capacitively coupled inductive boosting. The techniques we propose can be applied to memory as well as logic, and mitigates many of the overheads of the prior art.

Recently, we have proposed a voltage-supply boosting technique that leverages the inherent capacitance in a FinFET device to boost the supply voltage and improve SRAM operation [4]. In this paper, we develop another concept by adding an inductor between a new standard cell-based boosting circuit and the virtual supply voltage grid to the macro. Conventionally, parasitic inductance has been considered problematic for digital circuits, leading to mitigation techniques [18]. More recently, however, techniques such as resonant clocking have explicitly added inductors to induce resonant behaviors for power savings [19]. Our new inductor-based supply boosting technique not only leads to power reduction, but also improves our prior supply boosting technique in terms of  $V_{min}$ , access time, and yield. Table I provides a comparison of our proposed inductor-based supply boosting technique and prior works. While it is difficult to directly compare designs that may differ in target objectives and technologies, it is clear that our proposed technique advances the art significantly in terms of  $V_{min}$  and novel, simplified solutions.

This paper is organized as follows. Section II provides the proposed scheme and circuits. Simulation results are described in Section III, followed by theoretical corroboration in Section IV. Section V describes hardware measurements from a 14-nm test chip, and Section VI highlights the conclusions.

## II. PROPOSED SCHEMES

In this section, we describe dynamic supply boosting techniques to enable extreme low-voltage operations showcasing a 14-nm 8T SOI FinFET SRAM, including a peripheral logic. We describe two novel concepts for on-demand power supply boosting. First, we introduce a base technique employing capacitive coupling of a FinFET device and interconnects to boost  $V_{dd}$  [4]. Second, we build upon the base technique by adding an inductor to the boosting structure, which is presented for the first time in this paper.

### A. Boosting via Circuit Level Capacitive Coupling

The base technique exploits the unique capacitive coupling effect in a FinFET device to dynamically boost the virtual macro supply voltage during active mode, thus improving the access performance and  $V_{min}$  in the presence of variability. We also utilize interconnects to increase the capacitive coupling and thereby boost the power supply for the full macro. A negative bitline write-assist technique is also incorporated to further improve write-ability yield. The proposed scheme requires only a single supply and exploits the capacitive coupling from the gate and channels of a FinFET to its source as shown in Fig. 1. The basic circuit consists of two opposite polarity FETs in parallel with drains connected to  $V_{dd}$ . The boost transistor consists of an n-type FinFET with its gate controlled by the “BOOST” signal. Their common source forms a virtual  $V_{dd}$  ( $V_{ddv}$ ). In standby, BOOST is “Low”, thus the virtual array supply voltage is at  $V_{dd}$ .

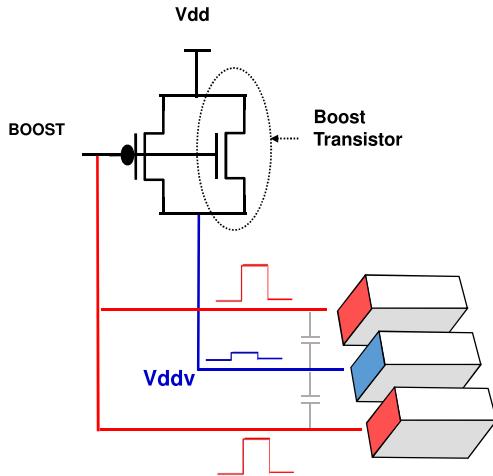


Fig. 1. Transistor boosting and interconnect boosting.

With both its drain and source at  $V_{dd}$ , the fully depleted silicon layer of the booster nFET is at  $V_{dd}$  as well. During active operation, the BOOST signal ramps to “High,” thus turning OFF the pFET header. The ramping up of the gate signal BOOST is capacitively coupled to the source of the booster nFET, thus boosting the  $V_{ddv}$ . The size of the booster is chosen through simulations to yield a  $V_{ddv}$  boost between 0.1 and 0.15 V. Additional boosting can be provided by two BOOST interconnects closely spaced to the virtual power grid [4]. As the BOOST signal switches from “Low” to “High” the capacitive coupling between interconnect further boosts the  $V_{ddv}$  signal.

Fig. 2 shows physics-based technology computer-aided design simulations to demonstrate the boost at short and long pulses with high and low voltages. One key aspect that is evident from Fig. 2 is that the relative  $V_{ddv}$  increase is higher at lower  $V_{dd}$ , making this technique attractive for improving  $V_{min}$  and performance at lower  $V_{dd}$ .

### B. Boosting via Capacitive and Inductive Coupling

Fig. 3(a) shows the new capacitively coupled inductor-based boosting scheme. In this case, the device-based capacitor circuit provides the initial boost to “ $V_{ddv}$ ,” which then gets further boosted by an appropriately sized inductor. The detailed analysis will be shown in Section III. The explicit inductor and inherent macro capacitance creates an *LC* tank that can be used to resonate and overshoot the supply, depending on the “*L*” and “*C*” values.

### C. Standard Cell-Based Boosting Structure

The booster circuit is a novel design employing a standard cell inverter, opening the door to a more general application of supply boosting. In our previous work [4], the boost circuit was designed by modifying the existing pFET header structure in the SRAM array macro to include an additional nFET boost transistor [Figs. 1 and 3(b)], requiring a full custom layout change, but leading to a low area overhead of approximately 4%.

The test chip in this paper, however, employs an SRAM macro that does not include a pFET header, prompting a new approach using a standard cell inverter, where the nFET source connects to  $V_{dd}$  [Fig. 3(c)], leading to a structure identical to the custom booster in Fig. 1. The inverter-based boost cell and prior gain inverters are then wrapped around the SRAM array macro. This new booster cell approach, consisting only of standard cells, provides a solution for boosting any array or logic macro at a reduced design effort cost and even lays a direction for full automation via synthesis tools.

### III. SIMULATIONS

Fig. 4 shows the cross section of the SRAM array we use for simulation-based investigations focusing on an 8T SRAM cell. This cross section represents an array that has 256 cells per bitline and 144 cells per wordline, i.e., the simulated SRAM cell is loaded by 255 cells and 143 cells for the bitlines and wordlines, respectively. Note that this is the same size as the array core from our 14-nm test chip described in Section V. We also include a version of the simulation that adds a negative bitline boost write-assist technique [4], [8] to account for realistic operation at low voltage. All simulations and measurements are at 85 °C. Fig. 5 shows simulated boosted  $V_{ddv}$  waveforms for two voltage and frequency pairs, with and without the new inductor technique. These simulations both use a 4-nH inductor and illustrate the range of  $V_{ddv}$  waveforms that can be generated. For this example, the value of the inductor is so chosen for functionality at ultralow  $V_{min}$ . Similar to other techniques that employ inductors for digital design, e.g., resonant clocking, it is important to size the inductor for the target cycle time range and/or provide multiple inductors for dynamic selection. In resonant clocking, boosting is not the primary motivation, while in this application boosting and recycling charge is a key for functionality and lower power. There are also numerous refinements to the proposed technique, such as, adding a diode to limit undershoot below  $V_{dd}$ , although these variants are beyond the scope of this paper.

Fig. 6 shows how inductor-based boosting with an appropriately sized inductor provides correct operation at low voltages without write-assist circuitry [Fig. 6(c)]. Without boosting, even the write-assisted case leads to incorrect functionality [Fig. 6(a)], while the case of boosting with a 2-nH inductor requires write-assist for correct functionality [Fig. 6(b)]. The 4-nH inductor provides an initial boosted pulse with a higher and wider amplitude than the 2-nH inductor, as Fig. 7 shows, providing functionality even without write-assist. The voltage overshoot provided by a properly sized inductor improves the device strength of the wordline and cell, allowing the writing. Fig. 8 shows an access time comparison between various inductor sizes, where the boost with inductor gives lower access time in general. Out of these inductors, 4 nH shows slightly higher delay for the same boost circuit, but it is capable of lowering  $V_{min}$  to 0.3 V compared with other techniques without any additional circuitry. In terms of power consumption, Fig. 9 shows that, as the inductor size increases, the average power decreases compared with boost alone due to oscillating “*LC*” characteristics as seen from Fig. 7. For 8T cell

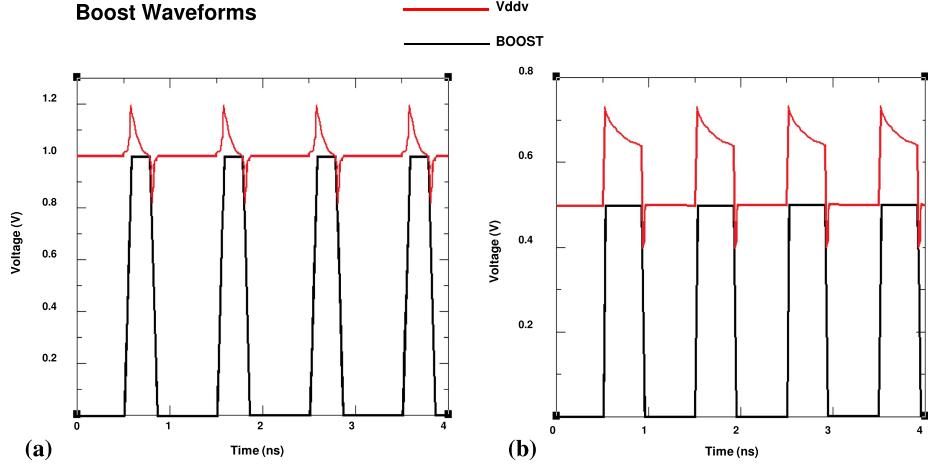


Fig. 2. Boost waveforms at (a) high voltage supply and (b) low voltage supply.

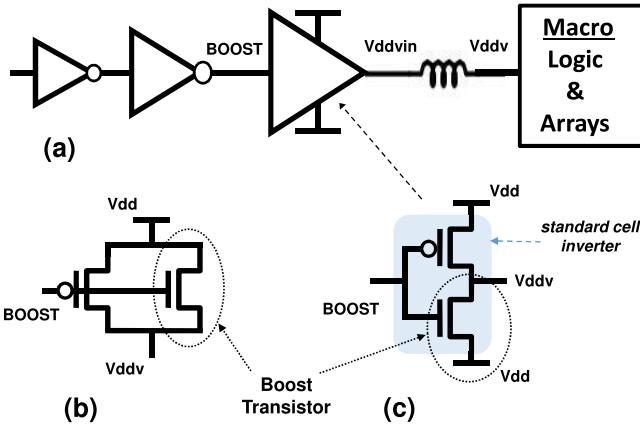


Fig. 3. (a) Inductive-based boosting concept. (b) Custom booster [4]. (c) Inverter-based booster.

functionality, write ability is a key parameter to be evaluated. Writeability failure is defined if the node “0” to be written by “1” does not reach 90% of  $V_{dd}$  during active wordline pulselength. With respect to yield, the writeability yield using a superfast high-sigma technique [21] is shown in Fig. 10. Using key mismatches as the input and the mixture important sampling technique described in [21], the high-sigma regime is explored in Fig. 10. This plot highlights that at low  $V_{dd}$  the 4-nH inductor has over three units higher yield than without boost and 1 unit over boost alone. Usage of the optimal “ $L$ ” and “booster capacitance” allows the boost voltage to be sufficiently higher than the one required for the writeability criterion.

From this yield analysis, it is possible to derive  $V_{min}$  for large arrays. Through the statistical analysis [21], failure probability ( $P_f$ ) per cell can be obtained. Thus estimated fails ( $E_f$ ) as a function of macro size can be given as

$$E_f = P_f \times (\text{SRAM Size}). \quad (1)$$

Fig. 11 shows the predicted cell failures for multiple macro sizes and  $V_{dd}$  voltages using optimal “ $L$ ” and 256 wordlines  $\times$  144 bitline per bank. Based on the redundancy and allowable estimated fails,  $V_{min}$  of a large macro can be estimated. In all

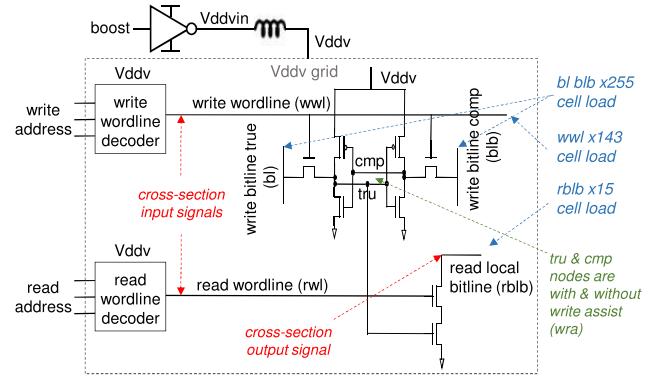


Fig. 4. 8T SRAM macro cross section (all simulations and measurements are at 85 °C).

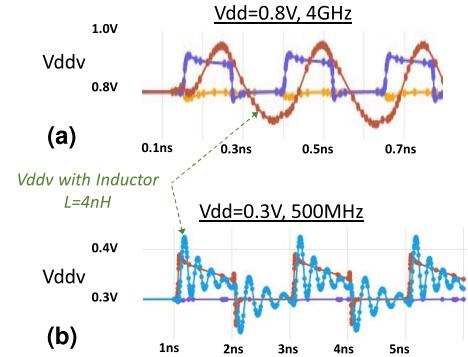


Fig. 5.  $V_{ddv}$  waveforms:  $L = 4$  nH, no inductor. (a)  $V_{dd} = 0.8$  V, 4 GHz and (b)  $V_{dd} = 0.3$  V, 500 MHz.

but the largest array, a minimum  $V_{dd}$  between 0.4–0.45 V is needed to produce a design point that has no expected errors. This supply is provided on demand through optimal usage of “ $L$ ” and booster capacitance. Even in the 64-Mb array macro, additional redundancy and error correction mechanisms would probably allow removing few expected errors for a  $V_{dd}$  of 0.4 V.

In terms of the variation and yield related to the addition of an inductor, Fig. 12 compares yield simulations with and

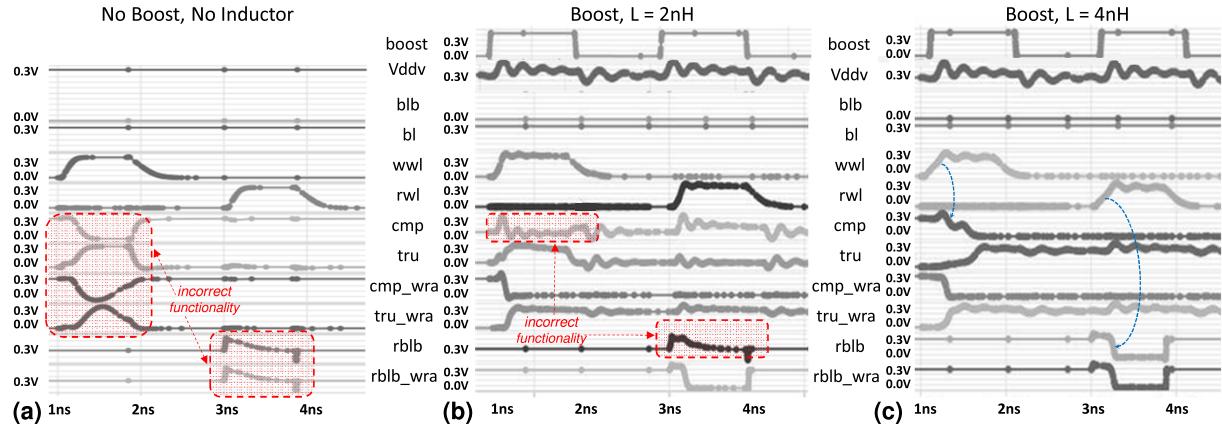


Fig. 6. Simulated waveforms at  $V_{dd} = 0.3$ . (a) No boost, no inductor, (b) boost,  $L = 2\text{ nH}$ , and (c) boost,  $L = 4\text{ nH}$ .

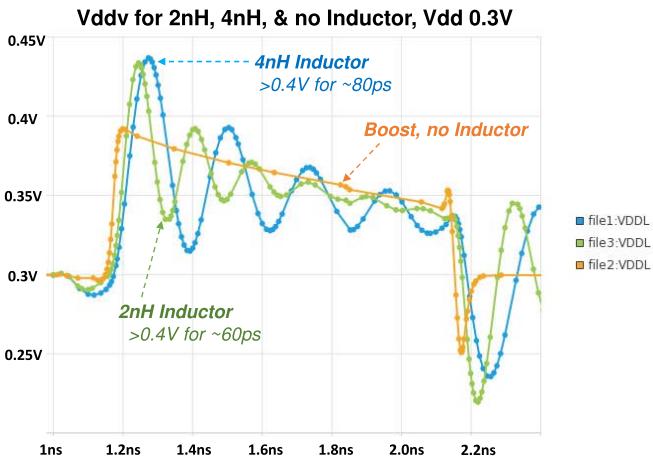


Fig. 7.  $V_{ddv}$  waveforms for  $L = 2\text{ nH}$ ,  $L = 4\text{ nH}$ , and no inductor.

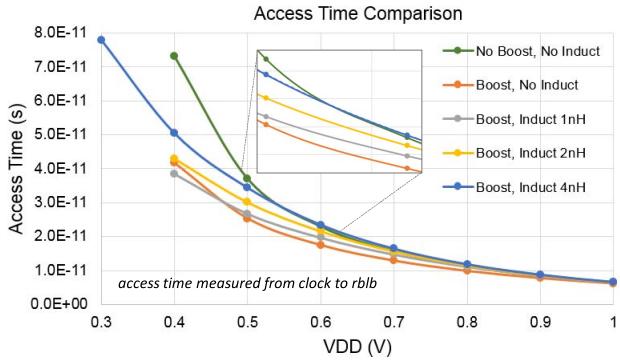


Fig. 8. Simulated cross section access time.

without equivalent series resistance (ESR) variation from the inductor. In Fig. 12, the ESR of the inductor is conservatively modeled as  $5\ \Omega$  and through simulation  $0.5\ \Omega$  is found to be a conservative estimate for  $1\sigma$  variation, leading to a  $6\sigma$  variation of  $3\ \Omega$ . The net outcome from the yield simulation is that the ESR is not expected to have a write-ability yield impact for the proposed inductive boosting technique. Fig. 12 also compares yield related to timing variation, assuming a  $20\text{ ps}/\sigma$

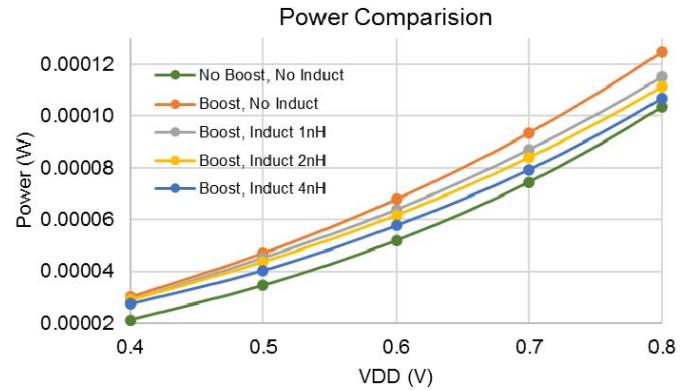


Fig. 9. Simulated power consumption.

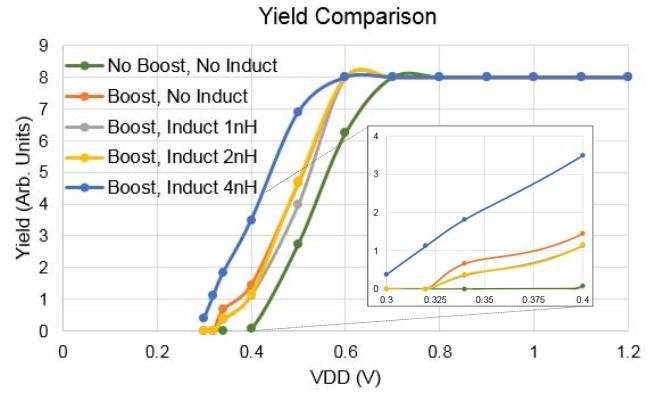


Fig. 10. Simulated yield.

variation of the boost delay signal with respect to wordline. This boost signal timing variation ( $-6\sigma$  to  $6\sigma$ ) also does not show significant degradation of yield. This statistical analysis shows as long as the  $V_{dd}$  voltage droop is above the retention voltage of the cell, the slight droop in  $V_{dd}$  from the inductor oscillations can be tolerated. Hardware results described in Section V support this analysis.

To better illustrate the design considerations for the proposed boosting technique, we use the simplified schematic

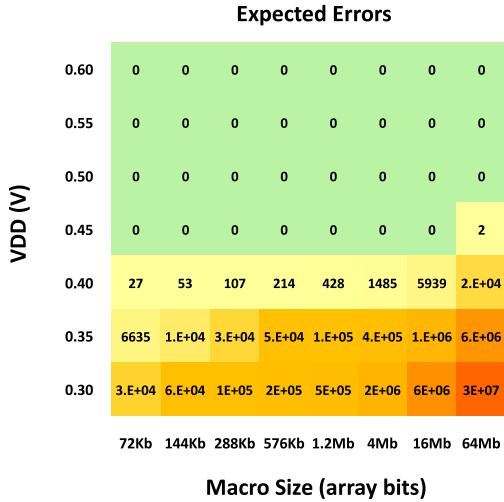


Fig. 11. Estimation of expected errors as a function of macro size and  $V_{dd}$ . Low and acceptable error rates determine minimum  $V_{dd}$ .

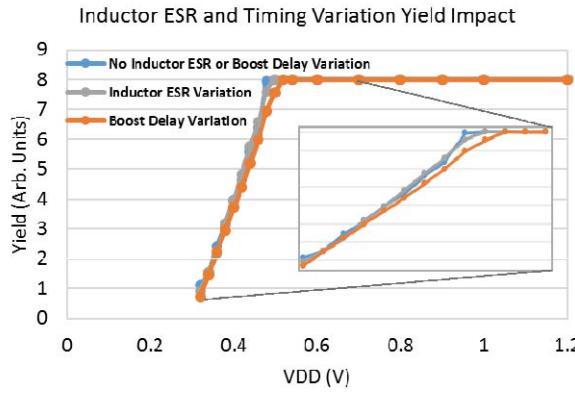


Fig. 12. Simulated yield of boost delay timing and inductor ESR variation.

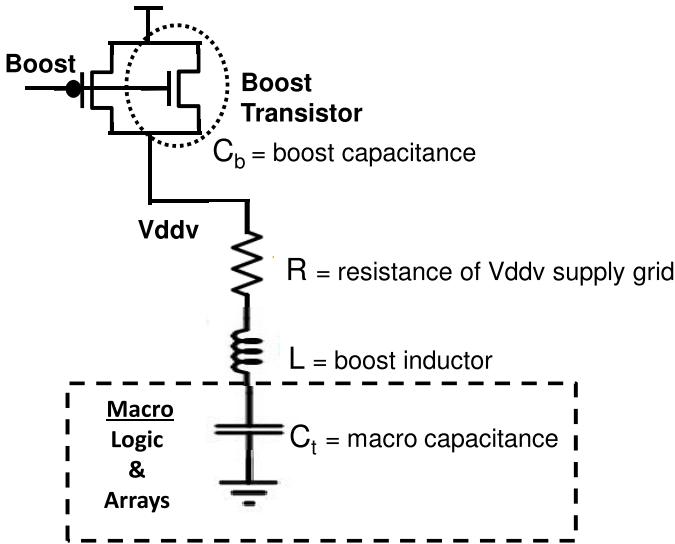


Fig. 13. Simplified schematic with lumped parasitics for generating a theoretical model.

in Fig. 13. Fig. 13 lumps the macro capacitance ( $C_t$ ), i.e., the capacitance seen by  $V_{ddv}$  from the array and logic circuits, as well as the boost capacitance ( $C_b$ ),

i.e., the capacitance created by the boost transistors. In addition, resistance ( $R$ ) and inductance ( $L$ ) are included in the schematic. From a design perspective,  $C_t$  and  $R$  are typically the characteristics of the macro being boosted and not easily modified for the purposes of boosting, whereas  $C_b$  and  $L$  are design parameters for boosting technique. Thus, the design challenge when boosting a macro is: given  $C_t$  and  $R$  values for a specific macro, determine the appropriate  $C_b$  and  $L$  values to reach the desired  $V_{min}$ , yield, performance, and power targets. In terms of  $C_b$ , for a given  $C_t$ ,  $R$ , and  $L$  values, we find that increasing  $C_b$  will increase  $V_{ddv}$  boost voltage seen by the macro, but at some point the  $V_{ddv}$  improvement will reach diminishing returns, at which point a higher  $C_b$  will lead to nonfavorable power increases due to the high number of boost transistors.

Fig. 14(a) shows an example of how sizing the  $C_b$  and  $L$  components can affect the  $V_{ddv}$  boost voltages for a given macro capacitance ( $C_t$ ). Likewise, Fig. 14(b) shows how the inductor size affects  $V_{ddv}$  boost voltages for various macro capacitances. In Fig. 14(b), we have chosen a large enough  $C_b$  such that the boost voltage is nearly maximized, but not so large that we see diminishing returns.

Based on simulations from Figs. 6(c) and 11, we find that for functionality at  $V_{dd} = 0.3$  V, the minimum boost voltage ( $V_{ddv}$ ) is around 0.43 V. While this voltage number may be specific for our technology and simulation models, for our purposes, it does provide a general target that the boost voltage should achieve. Additional margin may of course be added, albeit at the expense of additional power consumption through larger boost circuits or routing resources from a larger inductor. The graph also shows the combinations of  $L$  and  $C$  needed to achieve at least 0.43 V of boost voltage for  $V_{dd}$  of 0.3 V. Extrapolating the plots in Fig. 14, we see that the capacitively coupled inductive boost voltage can be 18% and 9% higher at 0.3 and 1 V, respectively, compared with capacitively coupled boost alone. This information is essential to manipulate  $V_{min}$  of a larger size SRAM.

#### IV. THEORETICAL CORROBORATION

It is important to understand the impact of the inductor-based boosting circuit and corroborate the behavior with a theoretical model. Using the same simplified circuit from Fig. 13, we develop a model that explains the circuit's behavior. The analysis as applied to the macro can be simplified into  $RLC$  circuit with appropriate boundary conditions to evaluate the oscillations. The derivation of the equation is given in the Appendix and is given as follows:

$$V_C(t) = (V_i - V_b)e^{-\alpha t} \left( \cos \omega_d t + \frac{\alpha}{\omega_d} \sin \omega_d t \right) + V_b - St \quad (2)$$

where  $\alpha = (R/2L)$ ,  $\omega = (1/(LC_t)^{1/2})$ ,  $V_b$  is the boost voltage, and  $S$  is the slope of the boost voltage decay with respect to time.

Using this equation, the circuit simulations (oscillation, frequency, and so on) are verified and the number of oscillations and the damping of the oscillations can be predicted.

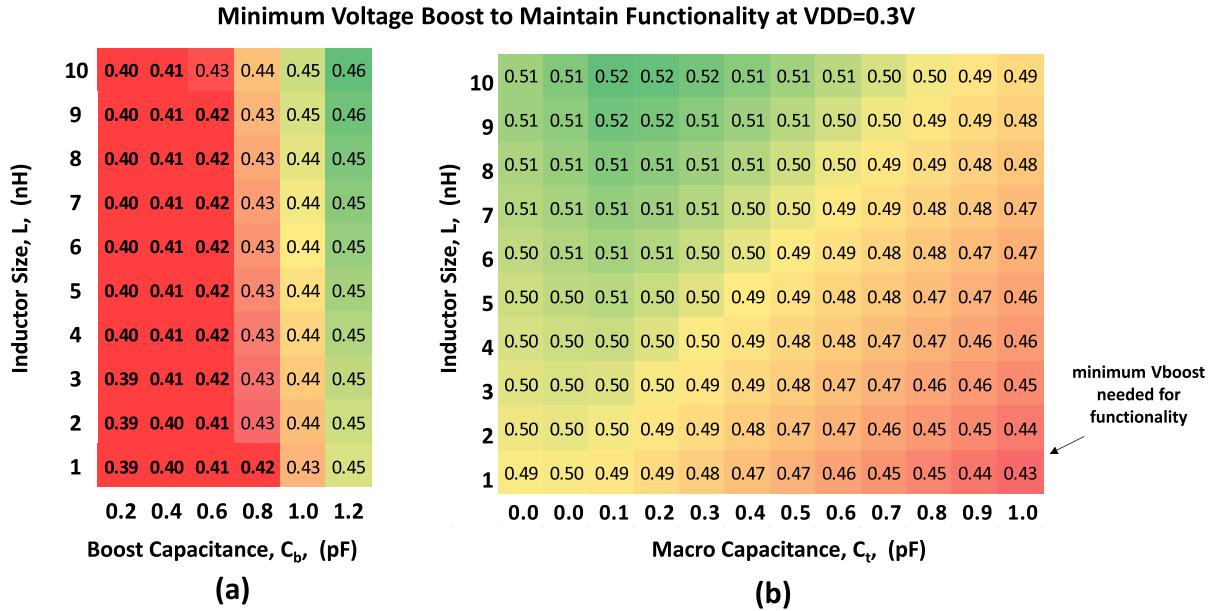


Fig. 14. (a) Inductor size and boost capacitance schmoo. (b) inductor size and macro capacitance schmoo.

Fig. 15 shows the comparison through simulations and theoretical equation and the trends seem to be corroborated well (for  $L = 4$  nH,  $C_t = 0.1$  pF, and  $S = 0.03$  V/ns).

## V. MEASURED HARDWARE RESULTS

The SRAM design in this test chip has a capacity of 36 Kb. It is arranged as 256 entries with 144 b per entry and has one read and one write port. The full array consists of four separate cores with 128 entries by 72 b and two redundant bits per core (Fig. 16). The array employs an eight-transistor SRAM-cell in a 14-track image. On the input side, the data is first stored into latches before it gets written to the array in the subsequent cycle. For the write scheme, the design implements a negative bitline boost write-assist technique to improve write ability at low voltages [4]. For the read scheme, a two-stage approach with both precharged local and global bitlines is used, with 16 cells connected to the local read-bitline. The global read-bitline is saved into a cross-coupled NAND before a mux for the final array data-out.

The fabricated chip in 14-nm FinFET SOI technology is shown in Fig. 17. There are four copies of the design on the chip. The 4-nH inductor is fabricated on the upper two metal layers using a spiral octagon topology with a 50- $\mu$ m radius. Since the inductor  $Q$  and  $L$  values slightly degrade due to eddy current interactions with underlying circuit if placed directly on top of SRAM, we offset the placement of the inductor from an SRAM. However, since front-end-of-line (FEOL) patterning was not required, it could potentially be placed over the array or other logic without consuming FEOL real estate by appropriately modeling the  $Q$  and  $L$  loss.

The measured histogram of chips in Fig. 18 shows that the new boosted inductor technique improves  $V_{min}$  by 80 mV over techniques without boost and 50 mW with boost only. These measurements even include a few functional cells at 0.29 V. Moreover, Fig. 19 shows how optimizing the boost signal

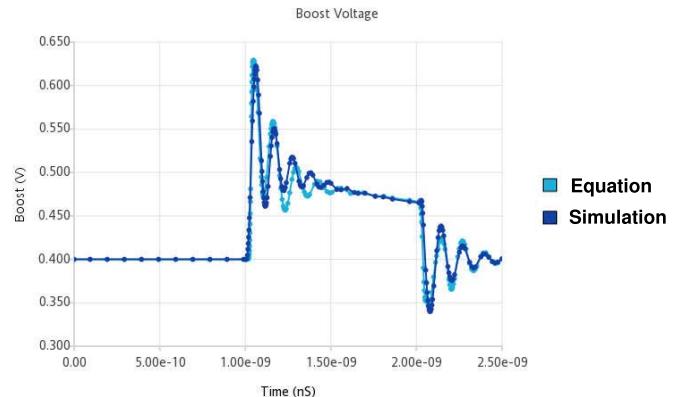


Fig. 15. Corroboration of generated equation and simulations.

timing sequence with respect to the macro clock can further reduce  $V_{min}$ . Thus at low voltages, there is sufficient margin between boost and clock signal (i.e., internally twodline signal) and corroborates our predictive yield analytics. Also, full functionality at such low voltages for various write/read patterns indicate that smaller voltage droops in the picosecond range can be handled without an issue.

In terms of access time, the boosted inductor design is over 10% faster than without boost and over 5% faster than with boost alone, especially at low voltages (Fig. 20). The boosted inductor technique also achieves a lower  $V_{min}$ . Note that Fig. 20 shows access times as measured by the tester, which includes IO pads and on-chip buffers to bring signals to/from the array, thus the boosted inductor speedup percentage of the array alone would be quite higher.

The total measured average power plotted for the padpage in Fig. 21 shows a 10%–12% increase for the boosted inductor case over baseline case of no boost and no inductor. Most of the additional power comes from the buffers driving the

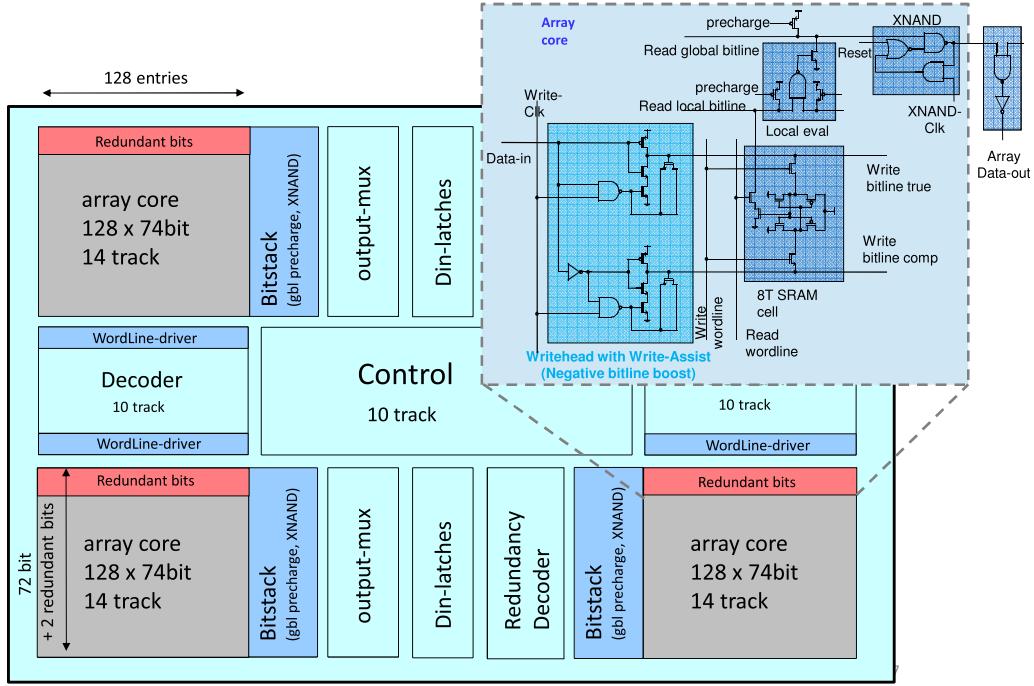


Fig. 16. SRAM macro block diagram.

### Prototype Chip Photograph

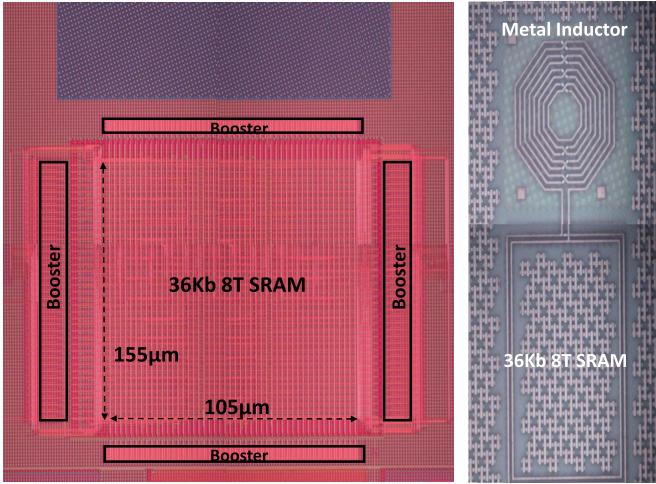
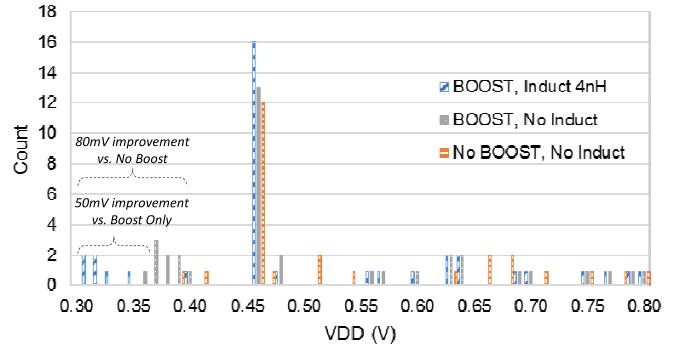


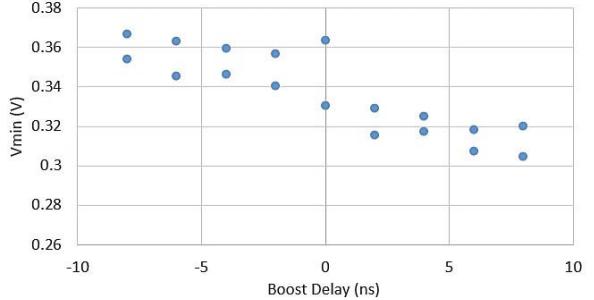
Fig. 17. Die photos.

boost circuit, which can be further optimized and have lower overhead at low voltage. Fig. 21 (inset) shows a slightly lower power for boost with the inductor versus capacitive boosting alone. The simulations in Fig. 5 corroborate this observation. The inductively coupled boost technique improves the boost voltage by 15% without significantly increasing the power, by relying on energy recycling. While a fine grain power breakdown at low voltages is difficult to measure, the measurements and simulations have relative agreement. Although the proposed boosting technique may incur a power overhead relative to the size of the macro being boosted, this technique provides the potential for chip-wide power reduction by reducing  $V_{min}$  of the entire chip by  $\sim 100$  mV or more.

### Measured $V_{min}$ with and without Boost and Inductor

Fig. 18. Measured  $V_{min}$ .

### Boost Delay Impact on $V_{min}$

Fig. 19.  $V_{min}$  impact of fine-tuning boost signal delay (two SRAM cells measured).

Furthermore, for SRAM macros composed of banks of arrays, like the macro in the test chip, there is the potential for finer grain boosting, where only the banks being accessed during a given cycle are boosted, leading to a reduced power overhead.

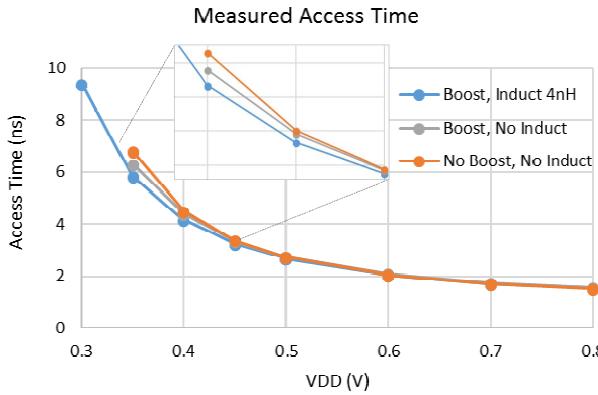


Fig. 20. Measured access time (65 chips).

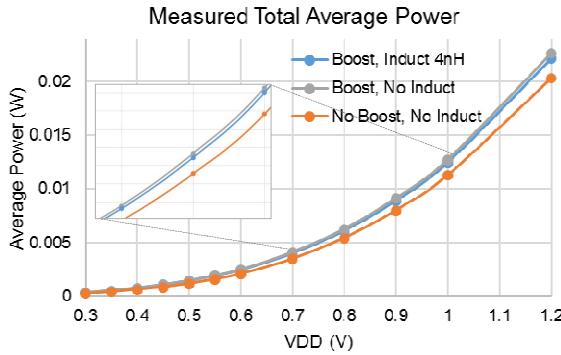


Fig. 21. Measured power (65 chips).

## VI. CONCLUSION

In summary, the new inductor-based boosting technique shows promising results for improving  $V_{min}$ , access time, and power consumption. This novel resonant supply boosting concept was explored through theoretical modeling, simulation, and measured hardware. The 14-nm FinFET SOI test chip measurement results verify simulations, suggesting improvements over the capacitive boosting technique alone. Although this paper provides only an initial investigation into resonant supply boosting, there are numerous avenues to further optimize the proposed technique, which may be beneficial for future low power processors, accelerators, and IoT applications.

## APPENDIX

Using Kirchoff's law for Fig. 13

$$V_f = iR + V_C + \frac{Ldi}{dt} \quad (A1)$$

$$i = C_t \cdot dV_c/dt \quad (A2)$$

$$V_f = RC_t \cdot \frac{dV_C}{dt} + V_C + LC_t \frac{d^2V_C}{dt^2} \quad (A3)$$

$$\frac{d^2V_C}{dt^2} + \frac{R}{L} \frac{dV_C}{dt} + \frac{V_C}{LC_t} = \frac{V_f}{LC_t} \quad (A4)$$

where  $V_f$  is a weak function of time as the boost voltage decays due to leakage. Thus (A4) needs to be solved as a

superimposition of a homogenous,  $V_{CH}(t)$  and nonhomogeneous equation,  $V_{CNH}(t)$

$$V_C(t) = V_{CH}(t) + V_{CNH}(t) \quad (A5)$$

The steady state  $[V_{ss}(t)]$  is constant and the resultant solution is  $V_{ss} = V_f$ , where  $V_f$  is the final value of the voltage and few millivolts above  $V_{DD}$

$$V_f = V_b - S \times t \quad (A6)$$

where  $S$  is the slope of decaying boost voltage  $V_b$  and  $t$  is the time for the homogeneous transient solution

$$\frac{d^2V_{CH}}{dt^2} + \frac{R}{L} \frac{dV_{CH}}{dt} + \frac{V_{CH}}{LC_t} = 0 \quad (A7)$$

Guessing the solution for  $V_{CH}(t) = Ae^{st}$  and simplifying the algebra, two solutions are obtained for (A6)

$$S_2 = -\frac{R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_t}} \quad (A8)$$

$$S_1 = -\frac{R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC_t}} \quad (A9)$$

$$V_{CH}(t) = Ae^{S_1 t} + Be^{S_2 t} \quad (A10)$$

For underdamped condition, i.e., for  $(R/2L) \ll (1/(LC_t)^{1/2})$  the roots of the equation are complex.

Using (5) and (7),  $V_{CH}(t)$  can be expressed as

$$V_{CH}(t) = Ae^{S_1 t} + Be^{S_2 t} \quad (A11)$$

Using the Euler Identity, the (A11) can be simplified to

$$V_{CH}(t) = e^{-\alpha t} (A' \sin \omega_d t + B' \cos \omega_d t) \quad (A12)$$

where  $A'$  and  $B'$  are constants and will be evaluated later.

The nonhomogeneous solution can be found by evaluating the time-dependent nature of (A6).

The nonhomogenous term can be assumed as

$$V_{CNH}(t) = A_1 - A_2 \times t \quad (A13)$$

Equation (A13) leads to the solution of  $A_1 = V_b + RCS$  and  $A_2 = S$ .

Superimposing homogenous and nonhomogenous solution yields

$$V_C(t) = V_{CH}(t) + V_{CNH}(t) \quad (A14)$$

Thus

$$V_{C_t}(t) = e^{-\alpha t} (A' \sin \omega_d t + B' \cos \omega_d t) + V_b + RCS-St \quad (A15)$$

RCS is negligible compared with all the terms and can be neglected.

Boost voltage,  $V_b$ , is given by the boost mechanism and can be expressed in the following way:

$$V_b = V_{DD}(C_t + 2C_b)/(C_t + C_b) \quad (A16)$$

Using the boundary conditions,  $A'$  and  $B'$  can be found out.

$$V_C(0) = V_i \quad (A17)$$

Another boundary condition is

$$\frac{dV_C(t)}{dt}|_{t=0} = \frac{i_L(0)}{C_t} \quad (\text{A18})$$

Since  $i_L(0)=0$  as the BOOST nFET is OFF and floating the initial current is OFF.

Thus

$$\frac{dV_C(t)}{dt}|_{t=0} = 0 \quad (\text{A19})$$

Using the boundary above-mentioned conditions the equation can be written as

$$\begin{aligned} V_{C_t}(t) &= (V_i - V_b)e^{-\alpha t} \left( \cos \omega_d t + \frac{\alpha}{\omega_d} \sin \omega_d t \right) \\ &\quad + \frac{S}{\omega_d} \sin \omega_d t + V_b + \text{RCS-St} \end{aligned} \quad (\text{A20})$$

where  $\omega_d = (\omega^2 - \alpha^2)^{1/2}$ ,  $\alpha = (R/2L)$ , and  $\omega = (1/(LC_t))^{1/2}$

Neglecting RCS and  $S/\omega_d$

$$V_{C_t}(t) = (V_i - V_b)e^{-\alpha t} \left( \cos \omega_d t + \frac{\alpha}{\omega_d} \sin \omega_d t \right) + V_b - \text{St} \quad (\text{A21})$$

#### ACKNOWLEDGMENT

This work has been partially sponsored by the DARPA Microsystems Technology Office (MTO), under contract no. HR0011-13-C-0022. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

#### REFERENCES

- [1] J. Davis *et al.*, "A 5.6GHz 64kB dual-read data cache for the POWER6 processor," in *ISSCC Dig. Tech. Papers*, Feb. 2006, pp. 622–623.
- [2] R. V. Joshi, R. Kanj, S. Nassif, D. Plass, Y. Chan, and C.-T. Chuang, "Statistical exploration of the dual supply voltage space of a 65nm PD/SOI CMOS SRAM cell," in *Proc. Eur. Solid State Device Res. Conf. (ESSDERC)*, Sep. 2006, pp. 315–318.
- [3] R. V. Joshi, R. Kanj, and V. Ramadurai, "A novel column-decoupled 8T cell for low-power differential and domino-based SRAM design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 5, pp. 869–882, May 2011.
- [4] R. V. Joshi, M. Ziegler, H. Wetter, C. Wandel, and H. Ainspan, "14nm FinFET based supply voltage boosting techniques for extreme low  $V_{min}$  operation," in *Proc. Symp. VLSI Circuits*, Jun. 2015, pp. 268–269.
- [5] E. Karl *et al.*, "A 0.6 V, 1.5 GHz 84 Mb SRAM in 14 nm FinFET CMOS technology with capacitive charge-sharing write assist circuitry," *IEEE J. Solid State Circuits*, vol. 51, no. 1, pp. 222–229, Jan. 2016.
- [6] B. Rooseleer and W. Dehaene, "A 40 nm, 454MHz 114 fJ/bit area-efficient SRAM memory with integrated charge pump," in *Proc. Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2013, pp. 201–204.
- [7] H. Morimura and N. Shibata, "A step-down boosted-wordline scheme for 1-V battery-operated fast SRAM's," *IEEE J. Solid State Circuits*, vol. 33, no. 8, pp. 1220–1227, Aug. 1998.
- [8] O. Hirabayashi *et al.*, "A process variation tolerant dual power supply with SRAM with  $0.179 \text{ mm}^2$  cell in 40 nm CMOS using level programmable wordline driver," in *ISSCC Dig. Tech. Papers*, Feb. 2009, pp. 458–459.
- [9] J. Kulkarni, B. Geuskens, T. Karnik, M. Khellah, J. Tschanz, and V. De, "Capacitive-coupling wordline boosting with self-induced VCC collapse for write  $V_{min}$  reduction in 22-nm 8T SRAM," in *ISSCC Dig. Tech. Papers*, Feb. 2012, pp. 234–235.
- [10] M. M. Khellah, A. Keshavarzi, D. Somasekhar, T. Karnik, and V. De, "Read and write circuit assist techniques for improving  $V_{ccmin}$  of dense 6T SRAM cell," in *Proc. Int. Conf. Integr. Circuit Design Technol. (ICICDT)*, 2008, pp. 185–188.
- [11] L. Hsu, R. V. Joshi, F. Assaderaghi, and M. Saccamango, "Method and system for improving the performance on SOI memory arrays in an SRAM architecture system," U.S. Patent 6,549,450, Apr. 15, 2003.
- [12] K. Takeda *et al.*, "Multi-step word-line control technology in hierarchical cell architecture for scaled-down high-density SRAMs," in *Proc. Symp. VLSI Circuits*, Jun. 2010, pp. 101–102.
- [13] H. Pilo *et al.*, "A 64 Mb SRAM in 32 nm high- $K$  metal-gate SOI technology with 0.7 V operation enabled by stability, write-ability and read-ability enhancements," in *ISSCC Dig. Tech. Papers*, Feb. 2011, pp. 254–255.
- [14] K. Nii *et al.*, "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," in *Proc. Symp. VLSI Circuits*, Jun. 2008, pp. 212–213.
- [15] K.-L. Cheng, M. Cao, and G. H. Chang, "A 20nm 112Mb SRAM in high- $K$  metal-gate with assist circuitry for low-leakage and low- $V_{min}$  applications," in *Proc. ISSCC*, Feb. 2013, pp. 616–618.
- [16] E. J. Nowak *et al.*, "Turning silicon on its edge [double gate CMOS/FinFET technology]," *IEEE Circuits Devices Mag.*, vol. 20, no. 1, pp. 20–31, Jan./Feb. 2004.
- [17] R. V. Joshi *et al.*, "FinFET SRAM for high-performance low-power applications," in *Proc. Eur. Device Res. Conf.*, Sep. 2004, pp. 69–74.
- [18] Y. Massoud, J. Kawa, D. MacMillen, and J. White, "Modeling and analysis of differential signaling for minimizing inductive cross-talk," in *Proc. Design Autom. Conf. (DAC)*, 2001, pp. 804–809.
- [19] S. C. Chan *et al.*, "A resonant global clock distribution for the cell broadband engine processor," *IEEE J. Solid State Circuits*, vol. 44, no. 1, pp. 65–72, Jan. 2009.
- [20] K.-H. Koo, L. Wei, J. Keane, U. Bhattacharya, E. A. Karl, and K. Zhang, "A  $0.094 \mu\text{m}^2$  high density and aging resilient 8T SRAM with 14nm FinFET technology featuring 560mV  $V_{min}$  with read and write assist," in *Proc. Symp. VLSI Circuits*, Jun. 2015, pp. 266–267.
- [21] R. Kanj, R. V. Joshi, and S. Nasif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *Proc. Design Autom. Conf. (DAC)*, 2006, pp. 69–72.



**Rajiv V. Joshi** (F'01) received the B.Tech. degree from the IIT Bombay, Mumbai, India, the M.S. degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, and the Dr.Eng.Sc. degree from Columbia University, Manhattan, NY, USA.

He is currently a Research Staff Member at the T. J. Watson Research Center, IBM, Yorktown Heights, NY, USA. He has authored or co-authored over 185 papers and given over 40 keynote/invited talks. He holds 58 invention plateaus and has over 215 U.S. patents and over 350 including international patents. His novel interconnects processes and structures for aluminum, tungsten, and copper technologies are widely used in IBM for various technologies from sub-0.5  $\mu\text{m}$  to 14 nm. He has led successfully pervasive statistical methodology for yield prediction and also the technology-driven static random-access memory at IBM Server Group. He commercialized these techniques.

Dr. Joshi is a member of IBM Academy of technology. He was a recipient of the three Outstanding Technical Achievement, three highest Corporate Patent Portfolio Awards for licensing contributions. He received the 2013 IEEE CAS Industrial Pioneer Award and the 2013 Mehboob Khan Award from Semiconductor Research Corporation. He is inducted into New Jersey Inventor Hall of Fame in 2014 along with pioneer Nikola Tesla and he was also a recipient of 2015 BMM Award. He served as a Distinguished Lecturer for IEEE CAS and EDS Society. He is a fellow of ISQED and a distinguished alumnus of IIT Bombay. He is in the Board of Governors for IEEE CAS. He serves as an Associate Editor of IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS. He served on committees of ISLPED (Int. Symposium Low Power Electronic Design), IEEE VLSI design, IEEE CICC, IEEE Int. SOI conference, ISQED and Advanced Metallization Program committees. He served as a General Chair for IEEE ISLPED. He is an industry liaison for universities as a part of the Semiconductor Research Corporation. Also he is in the industry liaison committee for IEEE CAS Society.



**Matthew M. Ziegler** (M'04) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 2004.

He is currently a Research Staff Member at T. J. Watson Research Center, IBM, Yorktown Heights, NY, USA. He has authored or co-authored over 50 technical publications and has over 15 patents granted and/or filed. His current research interests include VLSI design productivity, design space exploration, optimization, and power reduction. This work has led to design methodologies and design automation systems for productivity enhancement and optimization used throughout IBM processor designs. Dr. Ziegler was a recipient of several Technical Accomplishment Awards in the areas of processor design, design automation, and low power design. He has directly participated in the design of IBM's Power Systems, z Systems, and BlueGene families of products. He has served on the technical program committee of ISLPED the last few years, along with program committees for other conferences previously. He is involved with the SRC as a TAB member. He is currently the Chair of the IBM VLSI PIC (professional interest community), a community fostering research collaboration between IBM and external researchers.

**Holger Wetter** received the Dipl.-Ing. Elektrotechnik degree in electrical engineering from Leibniz University, Hannover, Germany, in 1995.

In 1995, he joined the Server and Technology Group, IBM, Böblingen, Germany. He has been involved in several generations of IBM P- and Z-series microprocessors since 1995. He holds 16 patents. His current research interests include custom circuit design, static random-access memories-design, logic design, and synthesis.