Pragmatic Design of Gated-diode FinFET DRAMs

Ajay N. Bhoj and Niraj K. Jha
Dept. of Electrical Engineering
Princeton University, Princeton, NJ 08544
{*abhoj, jha*}*@princeton.edu*

***Abstract*— Scaling bulk CMOS SRAM technology for on-chip caches beyond the 22nm node is questionable, on account of high leakage power consumption, performance degradation, and instability due to process variations. Recently, two/three transistor one gated-diode (2T/3T1D) DRAMs were proposed as alternatives to address the SRAM variability problem, with an emphasis on high-activity embedded cache applications. They are highly competitive with an SRAM in terms of performance, while having a smaller power and area footprint at lower technology nodes. The current evolutionary trend in transistor structures is toward an era of multi-gate devices, which makes it necessary to identify design issues and advantages of gated-diode DRAMs implemented in a multi-gate technology.**

In this work, we address gated-diode DRAM design in FinFET technology using mixed-mode 2D-device simulations. We revisit the model of internal voltage gain in bulk gated-diodes and extend it to provide quantitative insight into designing Fin gated-diodes, i.e., gated-diodes in FinFET technology. To this effect, we propose FinFET variants of the bulk gated-diode configuration and identify parameters that are critical to enhancing the retention time and read current in 2T/3T1D FinFET DRAMs. Additionally, we show the superiority of 2T1D FinFET DRAM over 6T FinFET SRAM having pass-gate feedback (6T PGFB) and 2T1D bulk DRAM under the effect of variations using a quasi-Monte Carlo method implemented in FinE, an environment we have developed for double-gate circuit design that integrates Sentaurus TCAD from Synopsys with the Spice3-UFDG double-gate compact model from University of Florida under a single framework. Finally, we present a new tunable threshold gated-diode FinFET amplifier which uses an n-type gated-diode for voltage-boosting, along with a p-type gated-diode for zero-suppression.

I. Introduction

Static random access memory (SRAM) is currently the dominant memory architecture for caches in ICs, often occupying as much as 70% of total chip area [1]. Increased scaling has placed considerable stress on SRAM technology due to the effects of process variations on performance, stability and standby leakage power consumption. In order to circumvent the SRAM scaling/variability problem, researchers have considered replacing bulk SRAM with 2T/3T1D bulk DRAM [2], [3], or switching to a multi-gate implementation, such as FinFET SRAM [8], [9], [10]. 3T1D DRAM was shown to meet the performance requirements of an L1 cache memory in the temporal window of repeated accesses/writes, thereby obviating the need for a static memory, in [2]. It is scalable, robust to process variations, and has a smaller area footprint than 6T SRAM, leading to higher density. However, to our knowledge, there has been no attempt to explore the gated-diode DRAM design space in a multi-gate technology.

In this work, we address the above design problem in FinFET technology, in the light of process variations, by performing mixed-mode 2D device-level simulations in a double-gate design environment, called FinE, that we have developed. The main contributions of this work can be summarized as follows:

- We extend the model of internal voltage gain (ζ) in planar gated-diodes to Fin variants of gated-diodes. It provides a good quantitative insight into optimum device sizing, mode of operation and operating voltage.
- We explore the design space of 2T/3T1D FinFET DRAM cells in an effort to enhance retention time and read current, while minimizing cell leakage.
- We contrast bulk and FinFET 2T1D DRAM cells under variations and compare them with 6T PGFB cells [8].
- We present a new tunable threshold gated-diode sense amplifier that uses an n-type gated-diode for voltage-boosting and a p-type gated-diode for zero-suppression.

The work is organized as follows. Section II reviews planar gated-diodes and develops an augmented model of internal voltage gain. Section III describes the simulation setup. Sections IV and V analyze gated-diode DRAM design in FinFET technology. Section VI draws comparisons between 2T1D bulk, FinFET cells and 6T PGFB cells under the effect of process variations. Section VII deals with a new tunable threshold gated-diode sense amplifier with zero-suppression. Section VIII presents the discussion and conclusions.

II. Operating principle of the Gated-diode

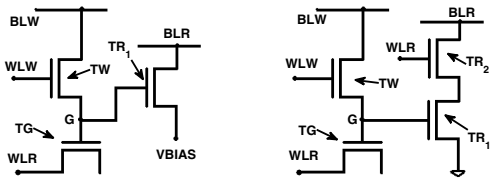
A preliminary analysis of gated-diode operation in planar single-gate technology is available in [3], [4], [5]. In this section, we develop an augmented model of internal voltage gain [3] to aid the design of fin gated-diodes for voltage-boosting, show that it matches well with device simulation in Section V and extend it to zero-suppression using p-type gated-diodes in Section VII.

A. Internal Voltage Gain

A gated-diode (TG, Fig. 1) can be implemented in bulk silicon either by shorting the source and drain of a FET or fabricating a 'partial' FET with a source and no drain [3], to form a two-terminal device. The nonlinear C-V characteristic of TG can be leveraged to obtain internal voltage gain in gain memory cells like 2T/3T1D DRAMs (Fig. 1), permitting low-voltage bitline operation as well as non-destructive read-out.

In Fig. 2(a), node G is initially at V_{HIGH}. On raising the source voltage by V_B, ΔQ_{HIGH} amount of charge is transferred to C_L (raising its voltage to V_{G_f}). This is indicated by the area under the C-V curve. Hence, the voltage boost at node

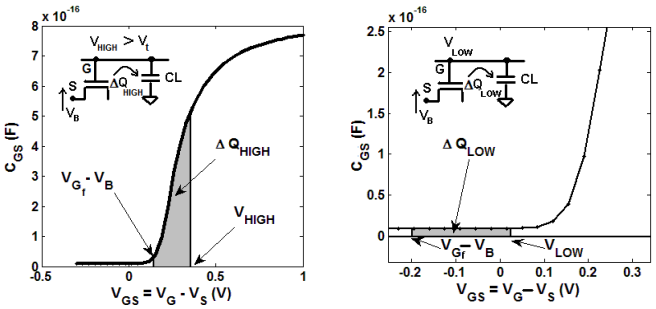
978-1-4244-5028-2/09/$25.00 ©2009 IEEE



(a) Type B1 (b) Type B2

Fig. 1. Bulk gated-diode DRAM cells: (a) 2T1D and (b) 3T1D

G is $\Delta V_{HIGH} = \Delta Q_{HIGH}/C_L$. Fig. 2(b) shows the same with $V_{GS} = V_{LOW}$. As $C_{GS_{LOW}} < C_{GS_{HIGH}}$, $\Delta Q_{LOW} < \Delta Q_{HIGH}$, so that $\Delta V_{LOW} < \Delta V_{HIGH}$. Hence, internal voltage gain enables greater separation between '1' and '0' levels while the stored voltage representing a '1' is much smaller. In Fig. 1, read transistor TR_1, whose gate is tied to G, is significantly overdriven while a '1' is read and does not turn on when a '0' is read, by the above principle. In Fig. 2, when the source voltage is lowered, ΔQ_{HIGH} (ΔQ_{LOW}) charge returns to TG, raising its V_{GS} to V_{HIGH} (V_{LOW}). This results in the highly beneficial non-destructive read-out feature. An important design decision is the choice of V_{HIGH}. It decides the mode of TG operation as well as the internal voltage gain which are quantified below.



(a) (b)

Fig. 2. ΔQ_{HIGH} (ΔQ_{LOW}) charge transferred when storage node G is at V_{HIGH} (V_{LOW}) and node S is raised by V_B

1) Constrained charge transfer mode: In Fig. 2, when node S is raised by V_B, part of the stored charge is transferred to C_L, raising its voltage to V_{G_f}. We use a step model to approximate the C-V characteristic for TG as well as C_L, i.e., $C_{GS} = C_{GS_{ON}}$, $V_{GS} > V_t$ and $C_{GS} = C_{GS_{OFF}}$, $V_{GS} \leq V_t$; $C_L = C_{L_{ON}}$, $V_G > V_t$ and $C_L = C_{L_{OFF}}$, $V_G \leq V_t$. If the initial voltage at G is V_{HIGH} and final voltage $V_{G_f} = V_{G_{HIGH}} = V_{HIGH} + \Delta V_{HIGH}$, then

$$\begin{aligned} \Delta Q_{HIGH} &= \int_{V_{G_f}-V_B}^{V_{HIGH}} C_{GS}(V)dV \\ &\approx C_{GS_{ON}}(V_{HIGH} - (V_{G_f} - V_B)) \\ &= C_{L_{ON}}(V_{G_f} - V_{HIGH}) \end{aligned} \tag{1}$$

We define internal voltage gain ζ as

$$\zeta = \frac{V_{HIGH} + \Delta V_{HIGH} - (V_{LOW} + \Delta V_{LOW})}{(V_{HIGH} - V_{LOW})} \tag{2}$$

Assuming $V_{LOW} = 0$, $C_{GS_{OFF}} = \eta \cdot C_{GS_{ON}}$, $C_{L_{OFF}} = \kappa \cdot C_{L_{ON}}$ and $C_{L_{ON}} = \chi \cdot C_{GS_{ON}}$, it can be shown that

$$\zeta = 1 + \frac{1}{\left[\alpha + \frac{V_t}{V_B}\right]} \cdot \theta \tag{3}$$

where θ, α are

$$\theta = \left[\frac{1}{1+\chi} - \frac{\eta}{\eta + \chi \cdot \kappa}\right], \quad \alpha = \left[\frac{V_{HIGH} - V_t}{V_B}\right]$$

Our assumption of partial charge transfer implies $V_{G_f} - V_B > V_t$, which imposes the necessary condition:

$$\left[\frac{V_{HIGH} - V_t}{V_B}\right] = \alpha > \left[\frac{\chi}{1+\chi}\right] \tag{4}$$

While the above approximates the true C-V characteristic with a step model at V_t, a better piecewise linear C-V model would capture $\eta(V_{HIGH})$ and $\chi(V_{HIGH})$ at the expense of mathematical complexity.

2) Complete charge transfer mode: In this mode, on raising the source by V_B, all the charge stored in TG is transferred to C_L. Using the earlier formulation, in the V_{HIGH} case:

$$\begin{aligned} \Delta Q_{HIGH} &\approx \int_{V_t}^{V_{HIGH}} C_{GS}(V)dV + \int_{V_{G_f}-V_B}^{V_t} C_{GS}(V)dV \\ &= C_{L_{ON}}(V_{G_f} - V_{HIGH}) \end{aligned} \tag{5}$$

Here, ζ and α have the same form as in Eq. (3) and θ is

$$\theta = \left[\frac{\alpha(1-\eta)+\eta}{\eta+\chi}\right] - \left[\frac{\eta}{\eta + \kappa \cdot \chi}\right] \tag{6}$$

In the complete charge transfer mode, $V_{GS} = V_{HIGH} > V_t$, and on raising node S by V_B, $V_{G_f} - V_B < V_t$, which gives the following necessary condition:

$$0 < \alpha < \left[\frac{\chi}{1+\chi}\right] \tag{7}$$

From the above model, there is a close correlation between the design values chosen for V_B, V_{HIGH} and V_t, and the gated-diode, read FET configurations which dictate the on/off state capacitances and hence, η, χ and κ.

III. Simulation Setup

In this section, we briefly describe our simulation setup. We have developed an environment called FinE (Fig. 3), which integrates Sentaurus TCAD [6] and the Spice3-UFDG [7] model into a single framework, thereby enabling designers to perform high-level experiments with ease. Table

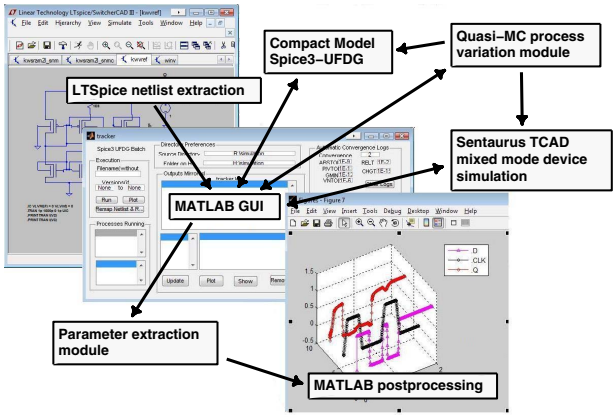


Fig. 3. FinE simulation framework for double-gate circuit design space exploration

I shows the parameters for a typical FinFET device used in later sections, where L_{GF}, L_{GB}, T_{OXF}, T_{OXB}, H_{GF}, H_{GB}, L_{SPF}, L_{SPB}, T_{SI}, H_{FIN}, $L_{UNDERLAP}$, N_{BODY}, Φ_G, N_{SD}, V_{DD} are the front and back physical gate lengths, gate dielectric thicknesses, gate thicknesses, spacer thicknesses, fin width, fin height, gate-drain/source underlap, body doping, gate workfunction, source/drain doping and the operating voltage, respectively. Fig. 4(a) shows the X-Y cross-section that was simulated in TCAD. The heavily doped extended source and extended drain regions ($H_{CON} \times L_{CON}$) aid in forming contacts to the device. They lead into the source/drain regions in the fin where the dopant concentration gradually decreases progressing towards the relatively undoped body region. The V_t of FinFETs is typically tuned by directly adjusting the workfunction of the gate material [11]. From a fabrication standpoint, owing to a high H_{FIN}/T_{SI} aspect ratio, we choose a single gate workfunction for all devices. We also included the effects of using high-k ($\varepsilon_{high\text{-}k} = 20$) gate dielectric in the FinFET structure shown in Fig. 4.

TABLE I

FINFET DEVICE PARAMETERS

PARAMETERS	
$L_{GF}, L_{GB} (nm)$	30
$T_{OXF}, T_{OXB} (nm)$	1.2
$H_{GF}, H_{GB} (nm)$	20
$L_{SPF}, L_{SPB} (nm)$	20
$T_{SI} (nm)$	15
$H_{FIN} (nm)$	75
$L_{UNDERLAP} (nm)$	12
$N_{BODY} (cm^{-3})$	10^{15}
$\Phi_G (eV)$	4.4
$N_{SD} (cm^{-3})$	10^{20}
$V_{DD} (V)$	1

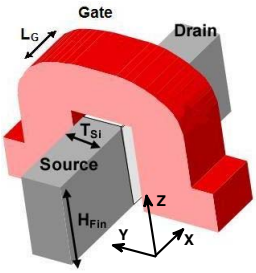
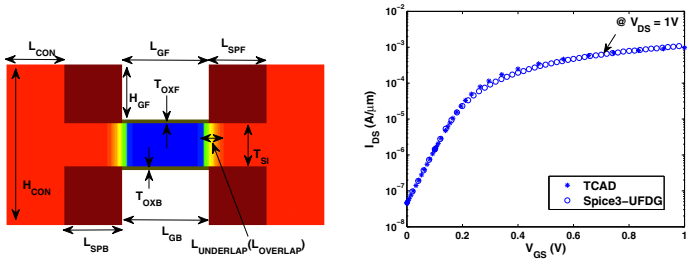


Fig. 4(b) shows the I_{DS} vs. V_{GS} characteristics for the device in Table I predicted using TCAD and the Spice3-UFDG model in FinE. Spice3-UFDG [7] is a physics-based compact model that simulates double-gate devices accurately, and shows excellent agreement with our device simulation in weak as well as strong inversion regions. We have employed device simulations using TCAD in FinE for all subsequent results owing to better convergence behavior in TCAD.



(a) (b)

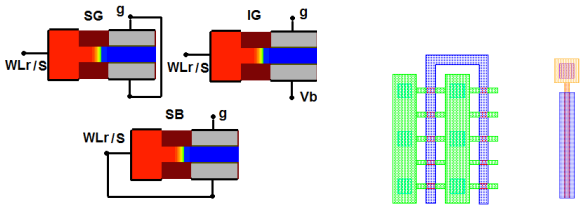
Fig. 4. (a) Two-dimensional (X-Y) cross-section of an n-FinFET simulated in Sentaurus TCAD and (b) I_{DS} vs. V_{GS} for $V_{DS} = 1V$ simulated in Sentaurus TCAD and Spice3-UFDG

With shrinking geometries, the effect of physical parameter variations during fabrication becomes important and is generally modeled through Monte-Carlo simulation experiments. We incorporated a quasi-Monte Carlo (QMC) tool [12] based on Sobol's sequence in FinE to avoid the sample clustering problem encountered in high-dimensional Monte Carlo simulation for time-consuming mixed-mode device simulations.

IV. FIN GATED-DIODES

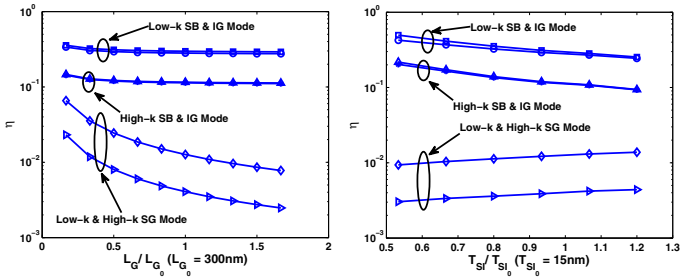
Fin gated-diodes, like their bulk counterparts, can be implemented as a FET with source and drain externally shorted or as a fin with only a source region. There are three possible modes of operation – front and back gates shorted [shorted-gate (SG) mode], back gate tied to source [source-back (SB) gate mode] and back gate biased at V_b [independent gate (IG) mode], as shown in Fig. 5(a).

Using the step C-V models from Section II, we operate TG in constrained charge transfer mode [SG, SB or IG ($V_b = 0$)] with $V_{HIGH} = 0.4V$. We maximize ζ by minimizing η, or equivalently we minimize $C_{GS_{OFF}}$ and maximize $C_{GS_{ON}}$. Since $C_{GS_{OFF}}$ for a single-fin TG, with source and drain externally shorted, is greater than that of a TG with a source region alone, we use the latter configuration hereafter.



(a) (b) (c)

Fig. 5. (a) TG in SG, IG and SB modes, (b) layout of a multi-fin SG-mode TG, and (c) layout of a single-fin SG-mode TG



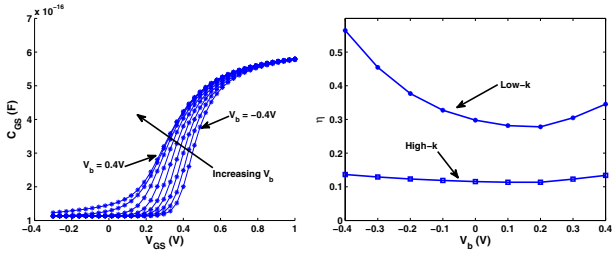
(a) (b)

Fig. 6. (a) Variation of η vs. L_G/L_{G_0}, $L_{G_0} = 300nm$ and (b) variation of η vs. T_{SI}/T_{SI_0}, $T_{SI_0} = 15nm$

There are two ways to implement TG – multi-fin minimum gate length or a single-fin elongated TG, as shown in Fig. 5(b) and 5(c). In the multi-fin case, using $L_G = 30nm$, $\eta \sim 0.1$. In the single elongated fin case, $\eta \sim 0.015$ for $L_G = 300nm$, which is chosen as the nominal TG gate length for subsequent portions of this work. As η is primarily set by the gate length (Fig. 6(a)), using multiple fins with minimum-sized gates is disadvantageous, as η marginally increases in spite of the increase in $C_{GS_{ON}}$, owing to proportionately higher parasitic capacitances. Therefore, using multiple minimum-sized gates results in poor ζ as well as higher layout area. On the other hand, with a single-fin elongated TG, η is small enough to provide good ζ and layout area is much lower.

Fig. 6 shows the dependence of η on L_G and T_{SI}, for SG, SB and IG ($V_b = 0$) mode TGs. For the SG mode, on

decreasing L_G, η increases as the inversion capacitance drops while $C_{GS_{OFF}}$ remains unchanged. η is lower in the high-k case owing to the fact that the gate dielectric has higher ε, leading to higher $C_{GS_{ON}}$, while the rest of the structure is identical to the low-k case. SB and IG modes show nearly an order of magnitude higher η, and variation with L_G and T_{SI} is minimal as $C_{GS_{OFF}}$ scales proportionately with $C_{GS_{ON}}$.



(a) (b)

Fig. 7. (a) Shift in low-k IG mode C-V characteristics with V_b and (b) variation of η in IG mode vs. V_b

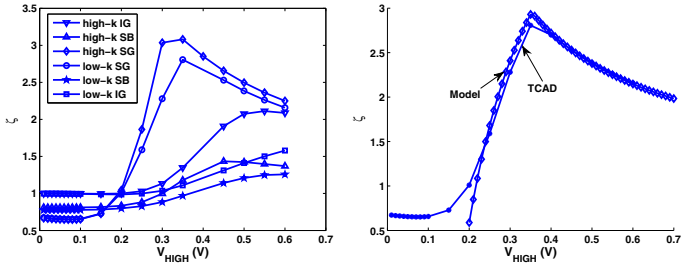
In the IG mode of operation, the C-V characteristic can be tuned by varying V_b, as shown in Fig. 7(a). The C-V curve shifts to the left on increasing V_b from $-0.4V$ to $0.4V$ in steps of $0.1V$, as it progressively gets easier to invert the fin with a higher V_b. The tub shaped η vs. V_b curves in Fig. 7(b) result from low $C_{GS_{ON}}$ at low V_b and high $C_{GS_{OFF}}$ at high V_b. In the next section, we explore internal voltage gain in fin gated-diodes and use them in 2T/3T1D DRAM configurations.

V. Gated-diode FinFET DRAM cells

Over the past decade, a lot of effort has been directed towards integrating one-transistor one-capacitor (1T1C) DRAM as an embedded cache memory. However, read-out in 1T1C cells is destructive, requiring a write-after-read mechanism, which increases the read cycle time. Also, high-speed, low-voltage operation is difficult due to a low signal-to-noise ratio. Gain memory cells like the bulk 2T/3T1D DRAMs described in [3] (Fig. 1) score over 1T1C cells due to their non-destructive read-out feature, low read latency and high signal-to-noise ratio even under low voltage operation. They are 'gain' cells as the amount of charge stored is considerably lower than the charge discharged in the bitline, and the storage capacitor is smaller than the 1T1C trench capacitor. In this section, we study the tradeoffs involved in 2T/3T1D FinFET DRAM design.

Fig. 1 shows a dual-port bulk 2T1D (Type B1) and 3T1D (Type B2) cell – BLW is the write port and BLR is the read port. TW is asserted on raising WLW to write a '0' ($V_{GS,TG} = V_{LOW}$) or a '1' ($V_{GS,TG} = V_{HIGH}$) at node G. In order to perform a read, WLR is raised and a stored '1' is boosted, thereby turning on TR_1. A stored '0' should typically provide no gain. If ζ is high enough, the resulting read current in the '1' case can discharge the highly capacitive bitline very quickly, and provide low-latency read access. An additional transistor can be introduced along the read path of the 2T1D cell to convert it to a 3T1D cell. In the 3T1D case, we can connect the source of TR_1 to ground instead of V_{BIAS} to increase the gate overdrive on TR_1. The read current can be enhanced further, by using a low V_t version of TR_1 [3].

In order to design the optimal 2T/3T1D FinFET cell configuration, we examined ζ under various modes of operation of TG and TR_1. We chose TR_1 as an SG-mode FinFET with parameters in Table I, and used TG in SG, SB and IG ($V_b = 0V$) modes with $L_G = 300nm$ for voltage-boosting ($V_B = 1V$, $V_{DS,TR_1} = 1V$) in the 2T1D configuration. Theoretically, from Section II, $\zeta_{MAX} = 1 + \frac{V_B}{V_{HIGH}} = 3.5$.



(a) (b)

Fig. 8. (a) Variation of ζ with TG in SG, SB and IG ($V_b = 0$) modes vs. V_{HIGH} and (b) ζ vs. V_{HIGH} for low-k SG-mode TG predicted by TCAD and the analytical model

From Fig. 8(a), high-k SG-mode TG offers the best ζ, which approaches ζ_{MAX}, due to the least η, while SB mode shows the worst ζ owing to highest $C_{GS_{OFF}}$ ($V_b = V_B$). While the difference in ζ between low-k and high-k gate dielectric is minor for the SG mode, it is considerable for IG and SB modes. Fig. 8(b) shows that the low-k SG-mode ζ vs. V_{HIGH} measured in TCAD agrees well with the analytical model described in Section II with parameters $\chi \sim 0.18$, $\eta \sim 0.015$, $\kappa \sim 0.44$ and $V_t \sim 0.22V$ extracted from ac simulations. ζ peaks at the boundary of the constrained and complete charge transfer modes. Here, $\alpha = \chi/(\chi+1)$ so that $V_{HIGH}|_{\zeta_{MAX}} = V_t + \alpha V_B \sim 0.22V + 0.15V = 0.37V$. It is important to note that the analytical model in Fig. 8(b) is applicable only when $V_{HIGH} > V_t$ or $\alpha > 0$. The step C-V approximation breaks down as V_{HIGH} approaches the neighborhood of V_t.

TABLE II

Operating Modes vs. Approximate θ

TG	TR_1	η	χ	κ	θ
SG	SG	η	χ	κ	$[1+\chi]^{-1} - [1+\frac{\chi\kappa}{\eta}]^{-1}$
SG	IG	η	$\frac{\chi}{2}$	$\frac{3\kappa}{2}$	$[1+\frac{\chi}{2}]^{-1} - [1+\frac{3}{4}\left(\frac{\chi\kappa}{\eta}\right)]^{-1}$
SB/IG	SG	10η	2χ	κ	$[1+2\chi]^{-1} - [1+\frac{1}{5}\left(\frac{\chi\kappa}{\eta}\right)]^{-1}$
SB/IG	IG	10η	χ	$\frac{3\kappa}{2}$	$[1+\chi]^{-1} - [1+\frac{3}{20}\left(\frac{\chi\kappa}{\eta}\right)]^{-1}$

For the sake of completeness, Table II shows the approximate ζ with TG in SG, SB and IG ($V_b = 0V$) modes and TR_1 in SG and IG ($V_{GBS} = 0V$) modes in the constrained charge transfer case. From Fig. 9, for TR_1 in SG mode, κ remains relatively unchanged at high V_{DS} and decreases as $V_{DS} < V_t$. For TR_1 in IG mode, κ is higher owing to higher $C_{L_{OFF}}$ and increases marginally when V_{GBS} or V_{DS} is increased.

With $\eta_{TR_1,SG} = 0.015$, $\kappa_{TR_1,SG} = 0.44$, $\chi_{TR_1,SG} = 0.18$, and $V_{HIGH} = 0.4V$, we find $\zeta_{TR_1,SG} \sim 2.72$. The IG mode configuration, however, operates in the complete charge transfer mode owing to an increase in V_t, as $V_{GBS,\ TR_1} = 0V$. With $\eta_{TR_1,IG} = 0.015$, $\chi_{TR_1,IG} \sim \chi_{TR_1,SG}/2 \sim 0.09$, $\kappa_{TR_1,IG} \sim$

$1.5\kappa_{TR_1,SG} \sim 0.66$, and extracted $V_{t_{IG}} \sim 0.38V$, using Eq. (6), we obtain an upper bound $\zeta_{TR_1,IG} \sim 1.3$ which is in good agreement with the respective measured ζ curve in Fig. 8(a). For other combinations of TG and TR_1, ζ is lower, as the ΔV_{LOW} component increases. In the IG/SB mode, $\chi = \chi(V_t, V_b, V_{HIGH})$, $V_t = V_t(V_b)$ and, hence, ζ is difficult to model analytically without externally extracting V_t, especially under the complete charge transfer mode. However, the step C-V approximation is still reasonable for relative bounding comparisons. Owing to poor ζ and increased layout area in IG/SB modes, TG is chosen to be in SG mode hereafter. Next, we proceed to the design of gated-diode FinFET DRAM cells using the above insights.

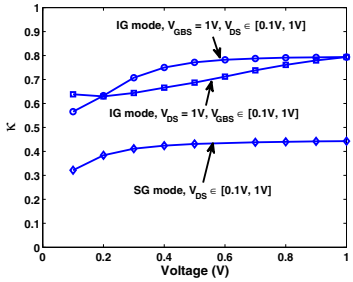
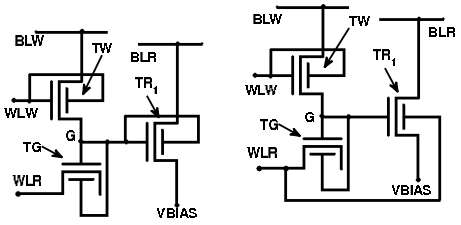
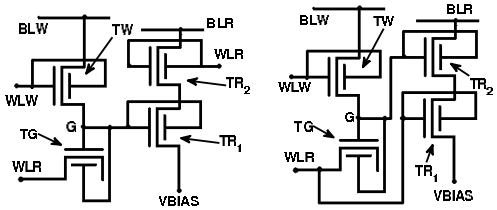


Fig. 9. Variation of $\kappa_{low\text{-}k}$ in SG and IG modes under various biasing conditions, $V_{HIGH} = 0.4V$



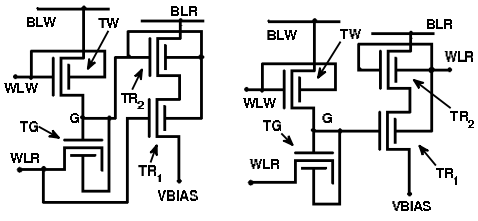
(a) Type 1 (b) Type 2

Fig. 10. Gated-diode 2T1D FinFET DRAM cells



(a) Type 3 (b) Type 4

Fig. 11. Gated-diode 3T1D FinFET DRAM cells



(a) Type 5 (b) Type 6

Fig. 12. Gated-diode 3T1D FinFET DRAM cells

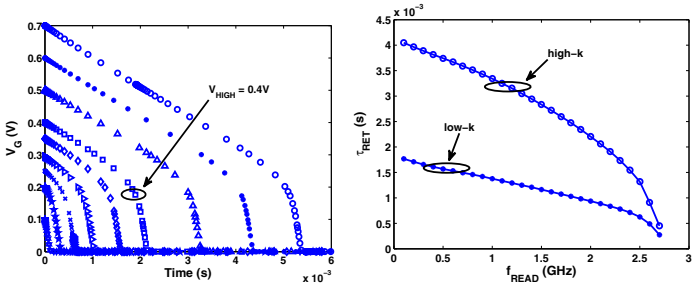
Fig. 10 shows two possible configurations for a dual-port 2T1D FinFET DRAM cell. Figs. 11 and 12 show 3T1D FinFET DRAM cells which are derived from 2T1D cells. Here, TW, TR_1 and TR_2 are instances of FinFET described in Table I and TG is in SG mode with $L_G = 300nm$. The operating voltages for the cells are shown in Table III.

TABLE III
CELL OPERATING VOLTAGES

Parameter	2T1D	3T1D
$V_{BLR}(V)$	1	1
$V_{BLW}(V)$	0.6 to 1	0.6 to 1
$V_{WLW_{low}} \rightarrow V_{WLW_{high}}(V)$	$-0.3 \rightarrow 0.6$	$-0.3 \rightarrow 0.6$
$V_{WLR_{low}} \rightarrow V_{WLR_{high}}(V)$	$0 \rightarrow 1$	$0 \rightarrow 1$
$V_{BIAS}(V)$	0.4	0

The metrics of interest for gain memory cells are standby cell leakage (I_{LEAK}), cell read current (I_{READ}) and retention time (τ_{RET}). The main contribution to I_{LEAK} is sub-threshold leakage along the read path. If V_{HIGH} is high enough, TR_1 is strongly turned on irrespective of WLR, for a stored '1'. In the worst case, all cells of a column store '1', and bitline leakage is significant. In the 2T1D cells, this is alleviated by setting $V_{BIAS} = V_{HIGH}$. If V_{BIAS} is increased further, I_{LEAK} decreases owing to the fact that $V_{GS} < 0$ for TR_1. However, the gate overdrive during a read operation also decreases, thereby reducing I_{READ}. In Types 3, 4, 5 and 6, V_{BIAS} is set to ground as with Type B2.

τ_{RET} is the time period upto which a degraded V_G provides sufficient read current upon voltage-boosting. Fig. 13(a) shows the stored $1 \rightarrow 0$ transition for different V_{HIGH}, for a low-k Type 1 cell. As V_G approaches V_t, the stored charge decreases considerably and $|dV_G/dt|$ increases.

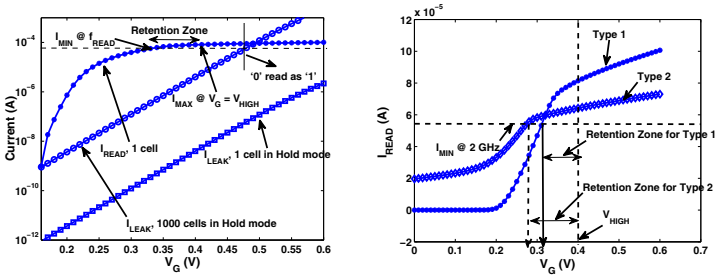


(a) (b)

Fig. 13. (a) Retention period for a stored '1' for different V_{HIGH} at $t = 0s$ in a low-k Type 1 cell with $V_{WLW_{low}} = -0.3V$ and (b) τ_{RET} vs. f_{READ} for low-k and high-k Type 1 cells

It should be noted that τ_{RET} is a function of the read frequency f_{READ}, as f_{READ} sets the maximum sensing time and, hence, determines the minimum I_{READ} or I_{MIN} required to maintain read fidelity. Fig. 13(b) shows the reduction in retention time on increasing f_{READ} for low-k and high-k Type 1 cells. τ_{RET} can be as high as $500\mu s$ even with $f_{READ} > 2.5$ GHz. While the high-k version shows nearly a two-fold improvement in τ_{RET} at low frequencies over the low-k case, the difference erodes at higher frequencies. The read current of the cell under a read operation competes with the total leakage current due to the remaining cells in the column. Fig. 14 shows the above graphically, with the zone of retention determined by I_{MIN} for the operating f_{READ} and V_{HIGH}. Hence, I_{LEAK} should be minimized as a large column

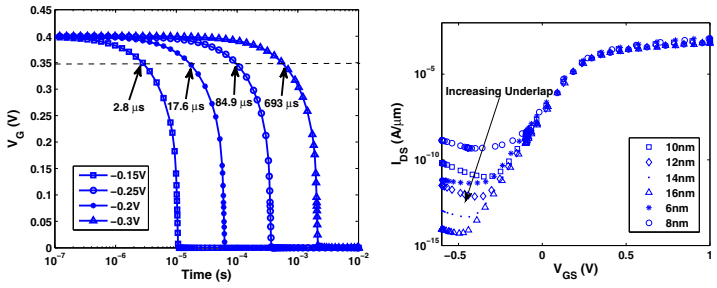
height would limit the maximum operating V_{HIGH}. The main



(a) (b)

Fig. 14. (a) I_{READ} and I_{LEAK} vs. V_G, showing the retention zone for a low-k Type 1 cell and (b) I_{READ} vs. V_G, showing the retention zones for low-k Type 1 and 2 cells at $f_{READ} = 2$ GHz

components of parasitic leakage that discharge node G are sub-threshold leakage in TW and the gate leakage currents through TW and TR_1. While choosing an IG-mode TW with a reverse-biased back gate helps reduce sub-threshold leakage, it increases the write time of the cell. If TW is in SG mode, $V_{WLW_{low}}$ affects the gate leakage as well as sub-threshold leakage in standby mode. Fig. 15(a) captures the dramatic increase in retention time on varying $V_{WLW_{low}}$ from $-0.15V$ to $-0.3V$. Higher reverse biases would imply larger voltage swings on WLW, which is unfavorable from the perspective of power consumption, and eventually gate-induced drain leakage would limit further reduction in I_{DS}. Fig. 15(b) shows that tailoring the underlap in TW is critical to controlling leakage. While a higher underlap yields a lower leakage, I_{ON} degrades as well and, hence, there exists a tradeoff between write time, $V_{WLW_{high}}$, $V_{WLW_{low}}$ and leakage through TW. In

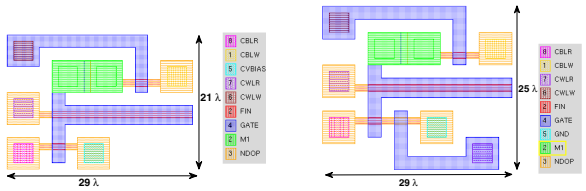


(a) (b)

Fig. 15. (a) Retention time for a stored '1' at different $V_{WLW_{low}}$ in a low-k Type 1 cell with $V_{HIGH} = 0.4V$ and (b) I_{DS} vs. V_{GS} for TW showing the role of underlap in suppressing leakage

order to improve I_{READ}, TR_1 can be upsized. Here, the width quantization property of FinFETs imposes a condition that widths of FETs can be modified only in discrete intervals of a single fin width. Hence, adding additional fins to increase the electrical width increases the cell area considerably, as consecutive fins are separated by a fin pitch. Also, using multiple fins for TR_1 would yield the same ζ only if TG is correspondingly sized up so that χ remains unchanged, as $\Delta V_{HIGH}/V_{HIGH} \propto (1+\chi)^{-1}$.

Fig. 16 shows the layouts of Type 1 and 3 cells where minimum feature size $\lambda = T_{SI}$. Both Type 3 and Type 2 cells suffer a 19% increase in cell area compared to Type 1 cells,



(a) (b)

Fig. 16. (a) Layout of a Type 1 cell and (b) layout of a Type 3 cell

owing to TR_2 and IG-mode TR_1, respectively. Also, a multi-fin TG implementation of the Type 3 cell occupies 1.97× layout area compared to that of Fig. 16(b).

TABLE IV

LOW-k CELL CHARACTERISTICS, $V_{HIGH} = 0.4V$

Type	$I_{READ}(\mu A)$	$I_{LEAK}(nA)$	$\tau_{RET}(\mu s)$	Area (μm^2)
1	81.4	4.0	935 @ 2GHz	0.137
2	64.2	0.03	1185 @ 2GHz	0.163
3	70.0	5.5	1284 @ 2GHz	0.163
4	80.7	3.3	1210 @ 2GHz	0.163
5	80.9	5.0	823 @ 2GHz	0.163
6	69.5	5.5	1847 @ 2GHz	0.163
B1	29.8	1.5	1.7 @ 0.5GHz	0.189
B2	42.3	1.3	4.2 @ 0.5GHz	0.235
6T	60.6	22.6	-	0.260

From Table IV, Type 1 cells show 26% higher I_{READ} than Type 2 cells. Owing to the IG-mode TR_1, Type 2 cells suffer lower standby leakage as the V_t of TR_1 is higher at $V_{WLR_{low}}$. While higher ζ implies better I_{READ}, it does not guarantee a high retention time. This is demonstrated in Fig. 14(b), where a Type 2 cell having lower I_{READ} at $V_G = V_{HIGH}$ shows a larger zone of retention than a Type 1 cell, at $f_{READ} = 2$ GHz. Overall, Type 2 cells appear to strike the best tradeoffs in cell area, I_{LEAK}, τ_{RET} and I_{READ} at high f_{READ}.

VI. COMPARISONS UNDER PROCESS VARIATIONS

In this section, we draw comparisons between the Type 1, Type B1 and 6T PGFB cells [8] under process variations. Due to the time-consuming nature of mixed-mode device simulations, we set the number of QMC samples to 1000 and $3\sigma/\mu \sim 10\%$ for nominal variations of physical parameters in Table I.

We compared Type B1 and Type 1 cells in order to show the huge difference in performance by shifting to FinFETs under approximate iso-area conditions and identical circuit topologies. For Type B1 cells, we used a TG with $L_G = 300nm$, $W = 150nm$ and $V_t \sim 0.22V$. TW was minimum-sized with $L_G = 30nm$, $W = 30nm$ and $V_t \sim 0.3V$. TR_1 was sized with $L_G = 30nm$, $W = 45nm$ and $V_t \sim 0.22V$. From Table IV, we can see that bulk cells fail to compete with corresponding FinFET cells under approximate iso-area constraints owing to poor ζ, I_{READ} and τ_{RET} (the operating f_{READ} is considerably lower at 0.5 GHz compared to 2 GHz for FinFET DRAM cells). Fig. 17 compares the retention times of Type B1 and Type 1 cells. The Type 1 cell shows three orders of magnitude higher retention time at a higher read frequency. The spread in τ_{RET} for Type B1 is very large, with cells having retention time as high as few ms to as low as tens of ps. This is mainly due to the high sensitivity of sub-threshold

leakage and gate leakage through TW to variations. It is also partly due to the difficulty in designing highly scaled planar, bulk gated-diode DRAM cells with cell areas close to their FinFET counterparts and yet have FETs with good electrostatic integrity.

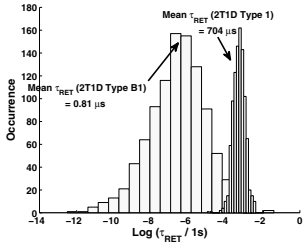
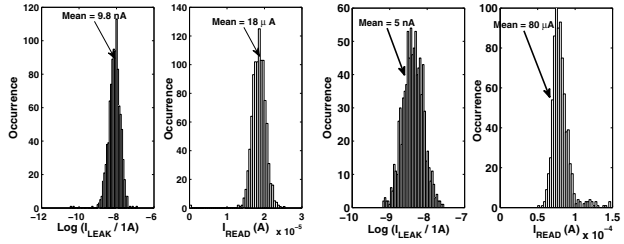


Fig. 17. Effect of process variations on τ_{RET} of Type B1 cells at $f_{READ} = 0.5$ GHz and Type 1 cells at $f_{READ} = 2.3$ GHz

Figs. 18 shows the effect of variations on I_{LEAK} and I_{READ} for Type B1 and Type 1 cells, respectively. While the spread in I_{LEAK} is similar, Type 1 cells show 4× larger mean I_{READ}. Hence, gated-diode FinFET DRAMs outperform their bulk counterparts under iso-area conditions.



(a) (b)

Fig. 18. (a) Effect of process variations on I_{LEAK} and I_{READ} for Type B1 cells and (b) effect of process variations on I_{LEAK} and I_{READ} for Type 1 cells

In order to make a comparison with 6T PGFB cells, we used the FinFET described in Table I (except $L_{UNDERLAP} = 6nm$ to improve I_{READ}, tipping the scales in favor of 6T PGFB) to simulate the pull-down and access FETs and complementary p-channel devices with identical parameters for the pull-up FETs. Fig. 19 shows the layout, with cell area 1.9× that of Type 1 cells. Note that 6T PGFB cells can only be implemented using FinFETs, not bulk single-gate FETs.

Fig. 20(a) shows the variation in I_{READ} and I_{LEAK} vs. $\Phi_G \in (4.4eV, 4.8eV)$ for the 6T PGFB cell. I_{READ} is maximum at $\Phi_G = 4.4eV$ and progressively decreases with increasing Φ_G. Hence, we used $\Phi_G = 4.4eV$, in an attempt to match I_{READ} with that of Type 1 cells. This is not optimal from a static noise margin (SNM) and I_{LEAK} perspective. Also, 6T PGFB yields higher SNM than most FinFET SRAMs in the literature, with minimal decline in I_{READ}, for a given Φ_G. With $L_{UNDERLAP} = 12nm$, $I_{LEAK} = 5.4nA$ and $I_{READ} = 34.5\mu A$, which is insufficient for operation at $f_{READ} = 2$ GHz. Decreasing $L_{UNDERLAP}$ to $6nm$ improves I_{READ}, but increases I_{LEAK} as well. From Fig. 20(b), even with $\Phi_G = 4.4eV$, the mean I_{READ} under nominal variations is 25% smaller than that of Type 1 cells, suggesting that gated-diode FinFET DRAMs can outperform 6T FinFET SRAMs in terms of read current as well as cell area.

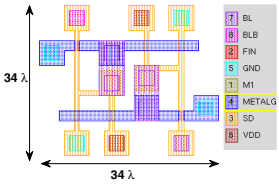
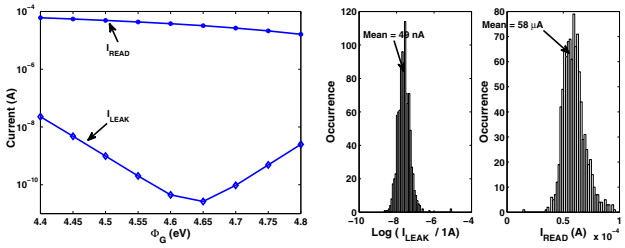


Fig. 19. Layout of an all-single-fin 6T FinFET SRAM with pass-gate feedback. Cell area $\sim 0.26\mu m^2$

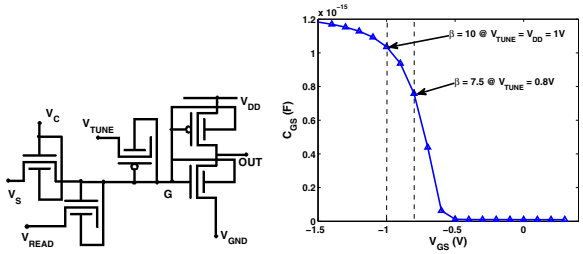


(a) (b)

Fig. 20. (a) Variation of I_{READ} and I_{LEAK} with Φ_G for 6T PGFB cells and (b) effect of process variations on I_{LEAK} and I_{READ} for 6T PGFB cells

VII. Gated-diode FinFET amplifier

Gated-diode DRAMs employ gated-diode sense amplifiers, which are designed based on the same principles of voltage-boosting discussed in Section II. Fig. 21(a) shows a gated-diode FinFET amplifier that modifies the configuration used in [4], [5]. The amplifier consists of a pass-gate transistor and gated-diode(s) for voltage-boosting/supression, which feed an inverter whose output can be latched.

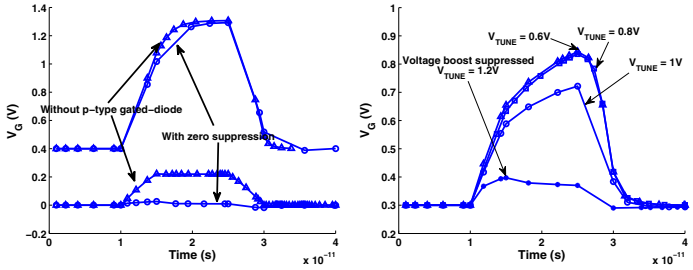


(a) (b)

Fig. 21. (a) Gated-diode FinFET sense amplifier and (b) p-type gated-diode C-V characteristic. Operating points are shown for $V_G = 0V$

In the absence of the p-type gated-diode shown in Fig. 21(a), the sensing mechanism works by allowing V_S to charge the n-type gated-diode on enabling V_C. Next, V_{READ} is asserted and the boosted voltage at G trips the inverter if $V_G = V_{HIGH}$.

A significant problem with voltage-boosting at close-to-zero levels is the inability to adequately suppress zeros from being boosted. This can be resolved by connecting a p-type gated-diode at G with its source terminal connected to V_{DD} for the static case, and V_{TUNE} for the tunable threshold case. Applying the step model to the C-V curve in Fig. 21(b), we set $C_{GS} = C_{P_{ON}}$, $V_{GS} \leq -|V_{t_p}|$ and $C_{GS} = C_{P_{OFF}}$, $V_{GS} > -|V_{t_p}|$. In the absence of the p-type gated-diode, $\chi^0 = C^0_{L_{ON}}/C_{GS_{ON}}$, $\kappa^0 = C^0_{L_{OFF}}/C^0_{L_{ON}}$. With the p-type gated-diode, $C_{L_{ON}} = C^0_{L_{ON}} + C_{P_{OFF}}$ and $C_{L_{OFF}} = \kappa^0 \cdot C^0_{L_{ON}} + C_{P_{ON}}$. Setting $\Omega = C_{P_{OFF}}/C_{P_{ON}}$ and $\beta = C_{P_{ON}}/C^0_{L_{ON}}$, we have $\chi = \chi^0(1 + \Omega\beta\chi^0)$ and $\kappa = (\kappa^0 + \beta)/(1 + \Omega\beta)$. Therefore, θ,



(a) (b)

Fig. 22. (a) Voltage-boosting for $V_{LOW} = 0V$, $V_{HIGH} = 0.4V$, with and without zero-suppression and (b) tuning the voltage-boosting threshold with V_{TUNE} for $V_{HIGH} = 0.3V$

under the complete charge transfer mode, is

$$\theta = \left[\frac{\alpha(1-\eta)+\eta}{\eta+\chi^0(1+\Omega\beta\chi^0)}\right] - \left[\frac{1}{1+\left(\frac{\chi^0(1+\Omega\beta\chi^0)(\kappa^0+\beta)}{\eta(1+\Omega\beta)}\right)}\right]$$
$$\approx \left[\frac{\alpha(1-\eta)+\eta}{\eta+\chi^0}\right] - \left[\frac{1}{1+\left(\frac{\chi^0(\kappa^0+\beta)}{\eta}\right)}\right], \quad \Omega\beta \ll 1 \tag{8}$$

Comparing Eq. (8) with Eq. (6), we see that the ΔV_{LOW} component is suppressed as $\kappa \rightarrow \kappa+\beta$, while the ΔV_{HIGH} term remains unchanged. Fig. 22(a), shows the above for $V_{LOW} = 0V$ and $V_{HIGH} = 0.4V$, with and without zero-suppression. Typically, $\beta \sim 10$ and $\Omega \sim \eta \sim 0.01$, so that voltage-boosting for the zero level is virtually eliminated. Furthermore, by varying V_{TUNE}, different operating points on the p-type *C-V* curve are chosen (Fig. 21(b)), resulting in different β and thresholds for voltage-boosting. While V_{TUNE} sets β, $\Omega\beta$ is independent of V_{TUNE}, and hence, from Eq. (8), if $\Omega\beta \ll 1$, the ΔV_{HIGH} component is unchanged on varying V_{TUNE}. Fig. 22(b) shows the effect of changing V_{TUNE} on voltage-boosting at node *G*. At $V_{TUNE} = 0.6V$-$0.8V \Rightarrow V_{GS,\ p\text{-}type} = 0.3V$-$0.5V$, the voltage-boosting is considerable enough to trip the inverter. This is consistent with the low capacitance seen in Fig. 21(b). For $V_{TUNE} = 0.8V$-$1.2V$, the operating point shifts to the inverted region in Fig. 21(b), resulting in a high β. Therefore, from the step *C-V* model, the voltage boost is suppressed if $V_{HIGH} + |V_{t_p}| < V_{TUNE}$. Since zero-suppression is controlled by β, it is possible to use *TG*s with poor η as long as χ is low enough to ensure that ΔV_{HIGH} is high. Hence, IG/SB mode *TG*s, which have poor η, can be used in voltage-boosting applications with the option of tuning *C-V* characteristics, and have a reasonably high ζ.

The above style of using an additional p-type gated-diode is not specific to either bulk or FinFET technology – owing to the requirement of a second long gated-diode with $\Omega \sim \eta$, a FinFET implementation would be the most area-efficient. The methodology can be adapted to other types of circuits as well, where greater separation between high and low voltage levels is required due to a low signal-to-noise ratio.

VIII. Discussion and Conclusions

Gated-diode FinFET DRAMs are an attractive choice for low-power, high-activity cache memories of the future. Fin gated-diode structures have a better chance of scaling, as they can easily implement the storage capacitances needed for high internal voltage gain and retention time, under tight area constraints. 2T/3T1D FinFET DRAMs offer a variety of possible cell topologies (based on SG and IG modes of FinFET operation), some of which have been explored in this work. While SG-mode fin gated-diodes yield the highest voltage gain, in the IG mode, the *C-V* curve can be tuned by the back gate bias to vary gain. Also, the read access time can be traded off with cell leakage along the read path, by statically or dynamically biasing the back gate of the read FET.

Overall, gated-diode FinFET DRAMs demonstrate excellent robustness to variations unlike their planar counterparts, and show more than two-fold higher read current per unit cell area, in comparison to 6T PGFB FinFET SRAMs, under similar conditions. The retention time of gated-diode FinFET DRAMs can approach millions of cycles, so that the amortized impact of refreshes on memory performance is negligible. Gated-diode FinFET amplifiers (and their variants) with a tunable voltage-boosting threshold and zero-suppression can be used to combat variability in sensing low-voltage swings under low signal-to-noise ratio scenarios. With dual-port operation, non-destructive read-out, large retention time, and high read current, gated-diode FinFET DRAMs constitute a versatile class of memories that have the potential to replace SRAMs in mainstream applications as well, with appropriate support at the architecture level.

IX. Acknowledgments

This work was supported in part by SRC under Contract no. 2007-HJ-1602 and in part by NSF under Grant no. CNS-0719936

References

[1] S. Balasubramanian, "Nanoscale thin-body MOSFET design and applications," Doctoral Dissertation, Dept. of Electrical Engineering and Computer Sciences, Univ. of California at Berkeley, Fall 2006.

[2] X. Liang, R. Canal, G. Wei, and D. Brooks, "Process variation tolerant 3T1D-based cache architectures," in *Proc. Int. Symp. Microarchitecture*, pp. 15-26, Dec. 2007.

[3] W. Luk and R. Dennard, "A novel dynamic memory cell with internal voltage gain," *IEEE J. Solid State Circuits*, vol. 40, no. 4, pp. 884-894, Apr. 2005.

[4] W. Luk and R. Dennard, "Gated-diode amplifiers," *IEEE Trans. Circuits and Systems-II: Express Briefs*, vol. 52, no. 5, pp. 266-270, May 2005.

[5] W. Luk et al., "A 3-transistor DRAM cell with gated diode for enhanced speed and retention time," in *Proc. IEEE Symp. VLSI Circuits*, pp. 184-185, 2006.

[6] Sentaurus TCAD manuals, http://www.synopsys.com, Synopsys Inc.

[7] J. Fossum et al., "A process/physics-based compact model for nonclassical CMOS device and circuit design," *Solid State Electronics*, vol. 48, pp. 919-926, Jun. 2004.

[8] Z. Guo et al., "FinFET based SRAM design," in *Proc. Int. Symp. Low Power Electronics and Design*, pp. 2-7, Aug. 2005.

[9] R. Joshi et al., "A high performace, low leakage, and stable SRAM row-based back gate biasing scheme in FinFET technology," in *Proc. Int. Conf. VLSI Design*, pp. 665-672, Jan. 2007.

[10] H. Ananthan, A. Bansal, and K. Roy, "FinFET SRAM - device and circuit design considerations," in *Proc. Int. Symp. Quality Electronic Design*, pp. 511-516, Mar. 2004.

[11] D. Ha et al., "Molybdenum gate HfO_2 CMOS FinFET technology," in *Proc. IEEE Int. Electron Devices Meeting*, pp. 643, Dec. 2004.

[12] A. Singhee and R. Rutenbar, "From finance to flip-flops: A study of fast quasi-Monte Carlo methods from computational finance applied to statistical circuit analysis," in *Proc. Int. Symp. Quality Electronic Design*, pp. 685-692, Mar. 2007.