Asymmetric Underlapped FinFET Based Robust SRAM Design at 7nm Node

A. Arun Goud, Rangharajan Venkatesan, Anand Raghunathan and Kaushik Roy
School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA 47907
Email: {arungoud, rvenkate, araghu, kaushik}@purdue.edu

***Abstract*— Robust 6T SRAM design in 7nm technology node, at low supply voltage and rising leakage, requires ingenious design of FinFETs capable of providing reasonable I_{on}/I_{off} ratio and acceptable short channel effects even under new leakage mechanisms such as direct source to drain tunneling. In this work, we explore asymmetric underlapped FinFET design with the help of quantum mechanical device simulations considering both the bit-cell and cache design constraints. We show that our optimized FinFET achieves a significant improvement in on-current over conventional symmetrically underlapped FinFETs. Through circuit simulations using compact models, we demonstrate that when such asymmetric underlapped n-FinFETs are used as bit-line access transistors, read/write conflict can be mitigated with simultaneous reduction in 6T SRAM bit-cell leakage. Improvement in write noise margin as well as access time can also be achieved under iso-read stability condition. Based on these technology and bit-cell models, we have developed a CACTI-based simulator for evaluating asymmetric FinFET based SRAM cache at 7nm node. Using this device-circuit-system level framework and optimized asymmetric underlapped FinFETs, we demonstrate significant energy savings and performance improvements for an 8KB L1 cache and a 4MB last-level cache.**

Keywords—7nm; FinFET; asymmetric underlap; 6T SRAM; noise margin improvement; low leakage; cache; CACTI; scaled interconnect

I. Introduction

The continuous increase in the dataset size of applications and widening of the gap between processor and main memory speeds have led to an ever-increasing demand for large caches. In modern processors, the caches account for a significant fraction of the chip area as well as the power consumption. Traditionally, SRAM has been the workhorse for realizing caches due to its process compatibility and relatively low read and write access times. However, rising leakage and short channel effects such as threshold voltage roll-off, Drain-induced barrier lowering (DIBL), etc., pose significant challenges to SRAM design.

Earlier research efforts have explored a number of different device and circuit level techniques such as multi-V_t and multi-t_{ox} transistors [1], asymmetric source/drain doping [2], body biasing [3], write assist [4], 8T bit-cells [5], etc., to tackle the leakage problem and to improve SRAM stability. However, as the technology scales down to sub-10nm nodes, extremely small channel lengths and close proximity between highly doped source and drain regions introduces newer leakage components such as the direct source to drain tunneling (DSDT) leakage in addition to sub-threshold thermionic leakage and gate leakage. This leads to very high leakage currents as well as degradation of the transistor I_{on}/I_{off} ratio, which impacts the leakage and access time of the SRAM bit-cell. Furthermore, the conflicting design requirements of read and write operations in 6T SRAM bit-cell and limited supply voltage makes it highly challenging to design robust SRAM bit-cells with a low access time, V_{ccmin} as well as good read, write and hold noise margins that are needed for data stability. SRAM design at sub-10nm technology nodes, therefore, requires suitable device geometries that are optimized by taking these challenges into consideration.

In this work, we explore device optimization for 7nm FinFET based SRAM considering the increased short channel effects and leakage at the device level, the stability constraints of the SRAM bit-cell at the circuit level and the cache design requirements at the system level. We investigate the design of asymmetric underlapped FinFET that has a high on-current in one direction and a lower on-current in the other direction. We exploit this asymmetry to mitigate the read-write conflict in 6T SRAM bit-cells by using asymmetric underlapped FinFETs for bit-line access transistors. By systematically choosing the underlap, we also lower bit-cell leakage by reducing the dominant leakage components – direct source to drain tunneling (DSDT), subthreshold diffusion current and direct gate oxide tunneling (in particular, edge-direct tunneling). Further, we realize different levels in the cache hierarchy by suitable engineering of asymmetric FinFETs. For L1 caches, which are closest to processor, we optimize the device to achieve fast access times and lower read/write energies. For last-level caches (LLC), which have large size, the design is optimized to achieve low leakage power consumption. We show that this design outperforms the conventional symmetric underlapped FinFET based SRAM in terms of noise margins, access times, V_{ccmin} and leakage at the 7nm technology node.

In order to evaluate the proposed design and resulting benefits, we have developed a systematic device-circuit-system modeling framework. To account for new leakage mechanisms at the 7nm node such as DSDT, we carry out self-consistent Non-equilibrium Green's function (NEGF) – Poisson simulations [6] assuming ballistic transport model for FinFETs with different combinations of underlap. We then extract compact Verilog-A based lookup table FinFET model from the quantum mechanical device simulation results to simulate a standard 6T SRAM bit-cell circuit with different FinFET device candidates. This is then followed up with system level

This work was sponsored in part by DARPA under the PERFECT program.

simulation of a cache to assess cache access time, read energy and leakage power. For this, we have developed a cache simulator based on CACTI [7] by adding support for 7nm FinFETs and interconnects. While previous studies [8][9] have focused on conventional symmetric FinFET SRAMs in older technologies, the focus and novelty of our work lies in using a device-circuit-system co-design approach, where we optimize 7nm FinFETs with help from physics-based device simulations, for realizing the access, pull-up and pull-down transistors of a 6T SRAM bit-cell in order to attain significant improvements in cache energy and performance metrics.

The rest of the paper is organized as follows – In section II, the device level and bit-cell level challenges at 7nm will be explained. Device simulation approach and asymmetric underlap optimization will be described in sections III and IV, respectively. Generation of compact models will be presented in section V, followed by 6T SRAM SPICE simulation results in section VI showing benefits of optimal asymmetric underlapping. CACTI modifications for FinFETs and projected interconnect parameters will be presented in section VII. Section VIII describes system level evaluation of first and last-level caches and section IX concludes the paper.

II. Design Challenges at 7nm Node

A. Device Challenges

At sub-10nm channel lengths, emergence of new leakage mechanisms such as DSDT is of serious concern as it degrades the subthreshold characteristics severely along with other conventional short channel effects. An effective technique to tackle this leakage component is by using underlap on either side of the gate [10]. By increasing the effective channel length seen by the charge carriers, underlap can reduce both DSDT as well as conventional subthreshold leakage which arises due to diffusion of high energy carriers across the channel barrier in the off-state. The gate oxide tunneling leakage too drops since the field at the gate/drain and gate/source junctions is lowered. For previous technology generations, simulations show that while underlap would cause on-current and off-current to degrade, the associated reduction in gate capacitances due to the undoped underlaps is expected to partially lower the delay [11]. As channel lengths shrink further, the design space available for choosing optimal underlap becomes limited due to the discreteness of the channel semiconductor lattice at such scales. This granularity, which poses new constraints on choice of underlap, needs to be treated correctly using atomistic simulations. Furthermore, the reduced fin thicknesses required to maintain FinFET's immunity to short channel effects will give rise to severe confinement of carriers resulting in changes to the subband structure [12]. This calls for full band quantum mechanical simulations instead of conventional effective mass based simulations.

B. SRAM Challenges

Read/write conflict is a serious challenge faced by 6T SRAM designers. The static read noise margin (RNM) and write noise margin (WNM) of SRAMs, which serve as a measure of its stability, cannot be optimized together. This arises due to the conflicting high on-current for write operation and low on-current for read operation required from the bit-line access transistors (AX). As shown in Fig. 1, during read, AX discharges from the precharged bit-line BL through the pull-down FinFET (PD) connected to the node Q storing a 0. The effective resistance of AX along the path BL-Q is required to be larger (and on-current smaller) than that of PD so that node Q remains close to 0 and thus no read disturb occurs. During write operation, the AX connected to the node storing 1 (QB) is required to pull this node down and hence, must be stronger than the pull-up FinFET (PU). The current conduction through AX for write is along the path QB-BLB and therefore, the access FinFET should exhibit a low effective channel resistance (larger on-current) along this direction.

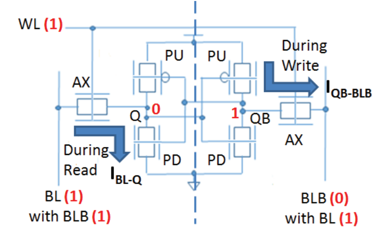


Fig. 1. 6T SRAM cell using symmetric underlapped device for access FinFET (AX). Current flow during read (left half) and during write (right half) are along opposite directions but their magnitude remains equal.

In conventional designs, this conflicting requirement cannot be met as all transistors exhibit symmetrical I-V characteristics. Therefore, the key toward building robust SRAM cells at scaled technologies is to overcome this read/write conflict through device optimization with the help of simulations.

III. Atomistic Device Simulation

In this section, we describe the device simulation framework we have adopted for sub-10nm device level optimizations for SRAMs. Device simulations involve generating current-voltage and charge-voltage characteristics taking into account the unique quantum mechanical effects that would arise at these scales (Fig. 2). The 2D NEGF approach coupled with the 2D Poisson electrostatics equation [6] is an efficient method to capture quantum effects. In this work, we have used the NEMO5 simulator [13] which computes the potential energy profile within the device self-consistently with the quantum mechanical charge and current computed from the sp3d5s* tight-binding based NEGF transport equations. Ballistic approximation has been assumed which is a reasonable approach at such length scales [14]. The fin height *hfin* was assumed to be large enough such that the potential profile along this direction is uniform and hence, all currents and capacitances resulting from the 2D simulation are normalized to the fin height. For circuit simulations, the fin height was chosen to be 10.9nm to calculate total drive current. The other key input parameters required by NEMO5 are listed in Table I

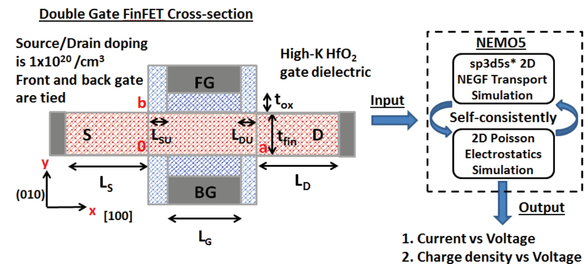


Fig. 2. FinFET device structure and NEMO5 simulation flow. Source side and drain side underlap have been indicated by L_{SU} and L_{DU}.

TABLE I. DEVICE SIMULATION INPUT FOR N/P-FINFET. DIMENSIONS ARE IN MULTIPLES OF SILICON'S LATTICE CONSTANT (0.545NM)

Name	Value	Name	Value
Gate length, L_G	4.9 nm	Source/Drain doping	1e20 /cm^3 N_D (N)/N_A (P)
Source/Drain length, L_S/L_D	5.4 nm	Channel doping	Intrinsic
HfO_2 ($\varepsilon_{r,eff}$=16) thickness, t_{ox}	1.09 nm	Source underlap, L_{SU}	Variable nm
Fin thickness, t_{fin}	2.73 nm	Drain underlap, L_{DU}	Variable nm
Front/Back gate-channel workfunction difference	0.293(N) 0.881(P) eV	Spacer material over underlap	Si_3N_4 ($\varepsilon_{r,eff}$=6.3)

IV. ASYMMETRIC UNDERLAPPED FINFETS FOR SRAMS

A. Asymmetric Source/Drain Underlap

As identified in section II.B, designing access n-FinFETs with asymmetrical I-V characteristics can potentially overcome the read/write conflict allowing for independent optimization of RNM and WNM. This can be achieved through asymmetrical drain and source underlap regions on either side of the gate as shown in Fig. 3. The source and drain terminals here are determined by the current flow direction. When the terminal adjacent to shorter underlap is at a lower potential (source) than the contact adjacent to longer underlap (drain), the gate controllability of the barrier close to the source improves. As per the virtual source injection model [15], this leads to enhancement of drive current (denoted as forward current in Fig. 3). When the potential at the terminals is swapped such that the longer underlap previously on the drain side now appears toward source side, the gate controllability of the source side barrier weakens which consequently reduces the drive current (denoted as reverse current in Fig. 3).

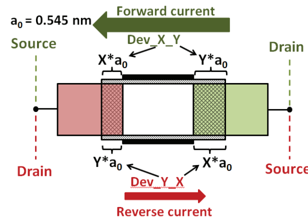


Fig. 3. Naming convention for underlap (*Dev_X_Y/Dev_Y_X*) and current flow direction for devices in Table II.

B. Read/Write Conflict Mitigation

An asymmetric underlap FinFET, by virtue of its unequal currents along BL-Q (or BLB-QB) and along Q-BL (or QB-BLB) can overcome the read/write conflict as shown in Fig. 4.

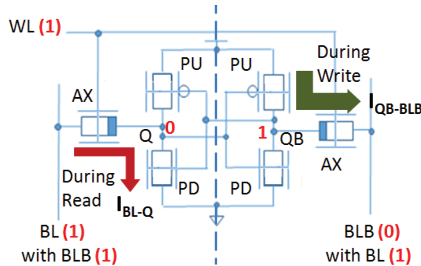


Fig. 4. 6T SRAM cell with asymmetric underlapped device for AX. Longer underlap side (highlighted in blue) is toward node Q and QB.

The read access time, which is a measure of the time required to discharge the bit-line close to the node storing a 0, depends on the total resistance of AX and PD along this discharge path. Although a weak reverse on-current improves RNM, it introduces a read stability/access time tradeoff and thereby imposes a constraint on how weak the AX can be made relative to PD in order to improve the read stability.

Since the leakage components are sensitive to the degree of underlap, the overall leakage trend is also modified by introducing asymmetric underlap. In this work, we lay a greater emphasis on first designing SRAMs with improved stability and low access time at the cost of leakage. This is crucial for small-sized, high performance L1 caches. For last-level caches (LLC) that are dense and power hungry, minimization of leakage power is more critical. This is achieved by lowering the supply voltage V_{DD} at the expense of access time.

C. Asymmetric Underlap Optimization

By varying the underlap on either side of the gate, we analyzed the I-V characteristics of several n-FinFETs (Table II). The baseline device for our study was chosen to be the device in Table I with equal source/drain underlap of 1.09nm. The resulting off and on-current meets the ITRS HP roadmap requirement [16]. We also consider source side overlap up to 0.545nm for last two candidates, 4 and 5. The naming convention used for the devices has been described in Fig. 3.

TABLE II. N-FINFETS WITH VARYING DEGREE OF UNDERLAP FOR LG=4.9NM. ALL DIMENSIONS ARE IN MULTIPLES OF SILICON'S LATTICE CONSTANT = 0.545NM

No	Name (N or P type)	Source/ Drain und(ov)erlap	I_{on} (A/um) /I_{off} (A/um) (V_{DD}=0.5V)	SS in (mV/dec)	DIBL (mV/V)
1	**Dev_2_2 (N)**	**1.09/1.09**	**2.23m/73n**	**78.79**	**59.2**
	Dev_2_2 (P)		**2.85m/94n**	**75.89**	**50.7**
2	Dev_1_3 (N)	0.545/1.635	2.35m/87n	79.28	66.5
	Dev_3_1 (N)	1.635/0.545	2.07m/66n	78.46	52.6
3	Dev_0_4 (N)	0/2.18	2.44m/111n	79.86	73.5
	Dev_4_0 (N)	2.18/0	1.91m/64n	78.39	45.8
4	Dev_-1_3 (N)	-0.545/1.635	2.75m/284n	84.18	82.3
	Dev_3_-1 (N)	1.635/-0.545	2.21m/148n	83.17	48.0
5	Dev_-1_4 (N)	-0.545/2.18	2.63m/205n	82.20	81.5
	Dev_4_-1 (N)	2.18/-0.545	1.97m/96n	80.66	43.2

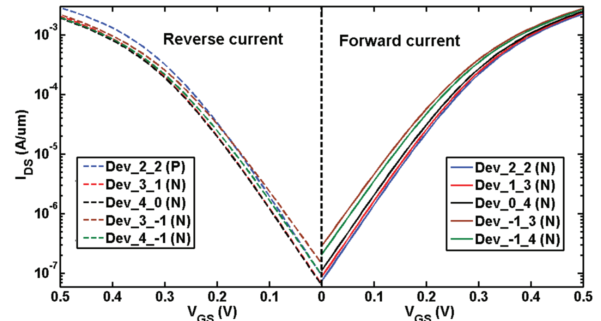


Fig. 5. I-V characteristics for devices in Table II.

It is observed from the forward current characteristics in Fig. 5 that introducing more asymmetry in an n-FinFET, by increasing drain underlap while decreasing source underlap, is desirable since it leads to enhancement of forward on-current which can improve writability. However a slight degradation in forward off-current, subthreshold swing (SS) and drain induced

barrier lowering effect (DIBL) occur, which are likely to impact cell leakage. When the source and drain voltages are swapped, these devices show a progressive decrease in the reverse on and off-currents. The low reverse off-current is beneficial for low cell leakage but if the reverse on-current is too small, the RNM of a 6T SRAM cell might improve but the access delay would degrade. The best asymmetrical underlap device design would therefore be one where the reverse on-current is not too small relative to the on-current of the symmetrical device, such that RNM and access times are reasonably good, while the forward on-current is much larger than the symmetrical device's on-current leading to WNM improvement. With these considerations, it follows that the best access FinFET candidate to realize good WNM, RNM and access time with acceptable leakage would be *Dev_-1_4*.

On the other hand, reducing cell leakage would require a device with low reverse off-current. Thus when leakage is of concern, *Dev_0_4* would be a better choice owing to its very small reverse off-current. We do not increase the drain underlap beyond 2.18nm since the forward on-current is expected to decrease and approach close to the symmetrical device's thereby nullifying the WNM improvement that is being sought. Also the source overlap is restricted to 0.545nm since any further increase in overlap will be accompanied by severe degradation in forward off-current.

V. Compact Model For FinFETs

To estimate SRAM noise margins, leakage power, V_{ccmin} and access times, we carried out bit-cell level circuit simulations using a compact model representation for FinFETs. Device simulation results consisting of bias-dependent current and charge for the chosen candidates were used to generate SPICE compatible Verilog-A based lookup table models. The entire flow is summarized in Fig. 6.

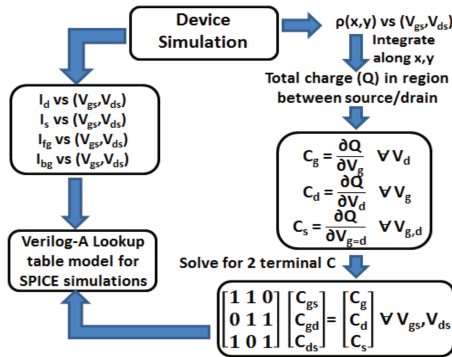


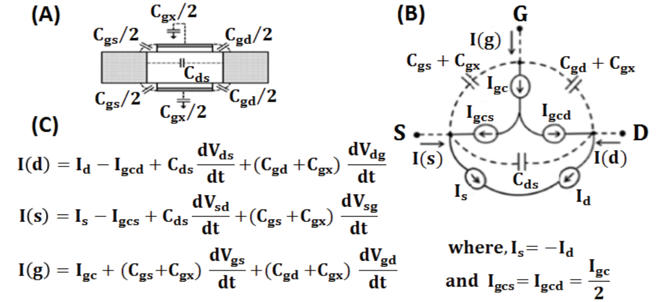
Fig. 6. Flowchart illustrating the device simulation, model extraction and circuit simulation steps.

A. Capacitances

The capacitances treated in our FinFET model (Fig. 7A) were gate-source capacitance (C_{gs}), gate-drain capacitance (C_{gd}), drain-source capacitance (C_{ds}) and gate-substrate capacitance (C_{gx}=0.15fF/um). By using quantum mechanical charge density as the basis for capacitance calculation step, all contributions such as quantum capacitance arising from filling of the channel sub-bands, junction capacitance, fringe/overlap capacitance, etc., are accounted for in C_{gs}, C_{gd} and C_{ds}.

B. Current Source Model for FinFET

The bias-dependent terminal currents and the calculated capacitances were substituted in the voltage-controlled current source model of a MOSFET (Fig. 7B, 7C). No source/drain resistance was included in this model and the two tied gates of the FinFET were folded into one gate terminal due to their equivalent nature. To calculate gate leakage, direct tunneling equations from BSIM4 [17] were adopted for both gates of the FinFET with an effective mass of $0.17m_0$ for HfO_2 [18].



$$I(d) = I_d - I_{gcd} + C_{ds}\frac{dV_{ds}}{dt} + (C_{gd}+C_{gx})\frac{dV_{dg}}{dt}$$

$$I(s) = I_s - I_{gcs} + C_{ds}\frac{dV_{sd}}{dt} + (C_{gs}+C_{gx})\frac{dV_{sg}}{dt}$$

$$I(g) = I_{gc} + (C_{gs}+C_{gx})\frac{dV_{gs}}{dt} + (C_{gd}+C_{gx})\frac{dV_{gd}}{dt}$$

where, $I_s = -I_d$ and $I_{gcs} = I_{gcd} = \frac{I_{gc}}{2}$

Fig. 7. Capacitances (A), voltage controlled current source model (B) and associated equations (C) (shown for n-FinFET only).

VI. SRAM Bit-cell Evaluation

In this section, we utilize the compact model described in the previous section to design a 6T SRAM cell and estimate its noise margin, access time and leakage power.

A. Noise Margin Improvement

Static noise margins such as read (RNM), write (WNM) and hold noise margins (HNM) are determined from the butterfly curves and serve as stability metrics for a bit-cell. They indicate the amount of noise voltage that can be tolerated on the internal nodes of the cross-coupled inverter during each operation such that the cell's contents are not flipped. A comparison of the NMs for a 6T SRAM cell with symmetric underlapped FinFETs (*Dev_2_2*) and cells with asymmetric underlapped FinFET for AX alone is shown in Table III. Here, V_{ccmin} for the bit-cell is defined as the minimum V_{DD} at which RNM and WNM are at least kT/q (=26mV at T=300K) [19].

TABLE III. Noise Margin, Leakage Power, V_{CCMIN} and Access Time Comparison for 6T SRAM Cell at V_{DD} = 0.5V with Symmetric Underlapped n/p-FinFET (Dev_2_2) for Cross Coupled Inverters

Access n-FinFET	HNM (mV)	RNM (mV)	WNM (mV)	Pleak (nW)	V_{ccmin} (mV)	Tac (ps)
Dev_2_2	**189.5**	**105.14**	**60.44**	**3.247**	**251**	**5.77**
Dev_1_3	189.5	107.22	62.56	3.251	240	4.77
Dev_0_4	189.5	108.95	64.00	3.265	219	3.84
Dev_-1_3	189.5	103.19	85.41	3.356	164	2.39
Dev_-1_4	**189.5**	**107.18**	**74.58**	**3.313**	**158**	**2.52**

As explained in IV.C, Dev_0_4 gives the highest RNM improvement with acceptable leakage while *Dev_-1_4* gives the lowest access time with slightly improved RNM, WNM and acceptable leakage. From Table III, it is also observed that the WNM is more sensitive to asymmetric underlapping than RNM. Increasing the drain side underlap of *Dev_-1_4* can potentially improve RNM further and reduce leakage (Fig. 5) but the resulting WNM would be close to *Dev_2_2* based SRAM's WNM.

B. Access Time and Leakage Power

We define access time as the duration between the instant the word line WL has risen to 0.5*V_{DD} and the instant the bit-line BL has discharged to V_{DD} - 50mV. The reduction in access time in Table III for *Dev_-1_3* and *Dev_-1_4* arises due to the reduction in BL capacitance that occurs when they are used as access n-FinFET. This BL capacitance depends on the C_{gd} of the access device that appears between BL and WL. For access time calculation, 8192 bit-cells were made to share BL and their resultant C_{gd} was used to model a realistic BL capacitance. While *Dev_0_4* and *Dev_-1_4* show RNM, WNM, V_{ccmin} and access time improvement, the leakage power has worsened slightly due to their larger off-currents. This can be addressed by lowering V_{DD} to achieve an improvement in leakage power under iso-RNM condition.

C. Iso-Read Stability Comparison

For this comparison, the V_{DD} of the cells with *Dev_0_4* and *Dev_-1_4* as access device (AX) were lowered until their respective RNMs approached the RNM of the all-symmetric underlapped *Dev_2_2* cell (Table IV). With *Dev_0_4* as AX, at V_{DD}=0.486V, a 3mV improvement in WNM, reductions of 1.3ps in delay and 136pW in leakage power per bit-cell can be achieved while with *Dev_-1_4* as AX, at V_{DD}=0.492V, a 13mV improvement in WNM, reductions of 3ps in delay and 24pW in leakage power per bit-cell can be achieved. Most of the power improvement for SRAM bit-cell with *Dev_0_4* as AX at iso-RNM V_{DD} of 486mV arises due to drastic reduction in the DSDT component (Fig. 8B).

TABLE IV. ISO-READ NOISE MARGIN COMPARISON FOR 1ST, 3RD AND 5TH SRAM CELL IN TABLE III BY LOWERING V_{DD}.

V_{DD} (mV)	Access n-FinFET	HNM (mV)	RNM (mV)	WNM (mV)	Pleak (nW)	Tac (ps)
500	**Dev_2_2**	**189.5**	**105.14**	**60.44**	**3.247**	**5.77**
486	Dev_0_4	185.6	105.08	63.35	**3.111**	4.44
492	Dev_-1_4	187.17	104.98	73.61	3.223	**2.8**

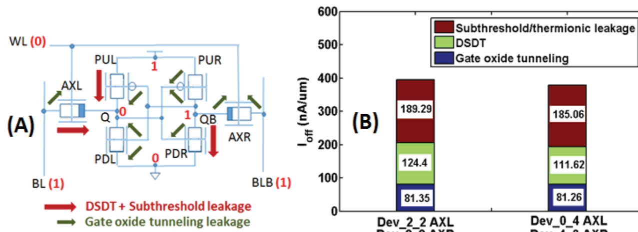


Fig. 8. (A) DSDT and subthreshold leakage components (in red) and gate oxide tunneling component (in green). (B) Comparison of leakage components for 1st and 2nd SRAM cell in Table IV.

To assess system level energy and performance improvements, an L1 and an L2 cache designed using the above baseline and optimized devices were simulated using CACTI cache simulator.

VII. CACTI SIMULATION USING FINFETS

The CACTI simulator is commonly used to assess energy, access time, leakage and area for caches built using planar bulk CMOS down to 32nm technology. To evaluate the benefits of a FinFET based cache at 7nm technology node having bit-cells with different degrees of underlap, CACTI was modified to support the unique FinFET challenges such as quantized transistor widths and bit-cells with different flavors of n-FinFET for the access transistors. Similar to conventional CACTI tool, our cache simulator assumes a hierarchical organization of cache into memory banks, sub-banks, cell matrix (mat), sub-arrays and finally bit-cells (Fig. 9).

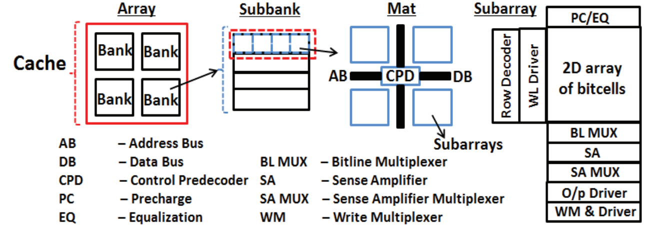


Fig. 9. Cache data/tag array structure used by CACTI [20].

The input for the modified CACTI simulator are the technology parameters of the node followed by the cache specification (cache size, block size, associativity, output width and address width) and configuration of the bit-cell arrays (number of word line divisions N_{dwl} and bit line divisions N_{dbl} which determine the number of sub-arrays N_{dwl} x N_{dbl}, number of sets mapped to a word line N_{spd}, etc.) that is required.

A. Quantized Transistor Sizes

The sizing of the word line driver and peripherals plays an important role in determining the access time of the cache. While a planar MOSFET based write driver and peripherals can be sized to any required aspect ratio W/L, for a FinFET based realization, all transistors are required to be sized in multiples of the fin height *hfin* (Note – Since 2D simulation includes contribution from both fin faces, we do not use *2*hfin*). As a result, the minimum transistor width parameter in our FinFET-CACTI is set to *hfin* and transistors of larger sizes are constrained to be multiples of this parameter.

B. Different n-FinFETs for AX and PD

To handle n-FinFETs with different underlaps for access (AX) and pull-down (PD) device in the 6T SRAM cell, we have introduced additional on-current, off-current, effective channel resistance and gate capacitance for the asymmetric access device. Two sets of these parameters (labeled with the suffix *_bq* and *_qb* to denote current flow is from bit-line BL to storage node Q and vice versa, respectively) are required depending on whether AX is in read mode or in write mode. The delay and leakage equations are modified such that the appropriate AX current flow direction and associated technology parameters are involved in the calculation. This modification is carried out only for the tag and data arrays. The cache peripherals (decoders, mux, sense amplifier, etc.) are realized using symmetric underlapped devices and the equations describing these blocks were left unchanged.

C. Interconnect Projection for 7nm

The analytical equations and fitting parameters in CACTI for modeling semi-global interconnects SG (used for routing within mat in Fig. 9) and global interconnects G (used for routing everywhere else) were scaled down to 7nm. The projected parameters are shown in Table V.

TABLE V. INTERCONNECT PROJECTION FOR 7NM TECHNOLOGY

Parameter (nm)	SG	G	Parameter	SG	G
Wire pitch	17.5	56	Wire R (Ω/um)	451	15.6
ILD thickness	210	385	Miller value	1.5	1.5
Wire width	8.75	28	Horizontal K	2.21	1.66
Wire thickness	17.5	61.6	Vertical K	3.9	3.9
Wire spacing	8.75	28	Aspect ratio	2.0	2.2
Barrier thickness	3	3	Fringe C (fF/um)	0.115	0.115
Dishing thickness	0	6.2	Wire C (fF/um)	0.236	0.217

VIII. L1 AND L2 CACHE RESULTS

Our modified CACTI tool was used to simulate two caches - *Cache_SymU* is realized using symmetric underlap FinFET (*Dev_2_2*) for all cache FinFETs while *Cache_AsymU* uses an asymmetric underlap FinFET (*Dev_-1_4* for L1 and *Dev_0_4* for L2) for AX alone and *Dev_2_2* for all other FinFETs.

TABLE VI. L1/L2 CACHE DESIGN SPECIFICATIONS FOR 7NM NODE

Specification	Value	Specification	Value
Cache size	8KB(L1)/ 4MB(L2)	IO bus width	256bits
Block size	64B	Operating Temp	300K
Associativity	4	Cache model	UCA
R/W ports	1	Bank count	1

A. L1 Cache

For L1 caches, low access/cycle times are important. For an 8KB L1 cache with specifications shown in Table VI, it is seen from Table VII that at iso-V_{DD} of 500mV, reductions of 11.3% in access time and 12.8% in cycle time can be obtained with 15.9% reduction in dynamic energy/cycle at improved WNM. The dynamic energy reduction arises mainly due to reduction in BL capacitance from using the asymmetric underlap.

TABLE VII. CACTI SIMULATION RESULTS FOR 8KB L1 CACHE WITH DEV_-1_4 AS AX FOR CACHE_ASYMU

Metric	Cache_SymU at V_{DD}=500mV	Cache_AsymU at V_{DD}=500mV
Access time (ps)	177.08	**157.1**
Cycle time (ps)	276.3	**241**
Edyn/access (fJ)	378	318
Pleak (uW)	116.5	122.1

B. L2 Cache

Energy and performance metrics for the 4MB L2 cache with specifications in Table VI are shown in Table VIII. At iso-V_{DD} of 500mV, it is observed that all metrics can be improved at the same time. Since L2 caches are required to operate at lower energy due to their large size, we scale the V_{DD} and do an iso-RNM comparison.

TABLE VIII. CACTI SIMULATION RESULTS FOR 4MB L2 CACHE WITH DEV_0_4 AS AX FOR CACHE_ASYMU

Metric	Cache_SymU at V_{DD}=500mV	Cache_AsymU at V_{DD}=500mV	Cache_AsymU at V_{DD}=486mV
Access time (ps)	736.1	669.7	722.4
Cycle time (ps)	816.2	748.6	780.5
Edyn/access (pJ)	22.9	18.3	**17.9**
Pleak (mW)	66.38	65.37	**63.3**

It is observed that under iso-RNM, compared to *Cache_SymU* operated at 500mV, *Cache_AsymU* operated at 486mV gives 21.8% reduction in dynamic energy and 4.6% reduction in leakage power along with improved WNM.

IX. CONCLUSION

In this paper, we analyzed tied double-gate Si FinFETs with varying degree of source/drain underlap using quantum mechanical 2D device simulations for designing robust 6T SRAM bit-cells. We showed that asymmetric underlapping is effective in sub-10nm technology nodes for mitigating read/write conflicts and improving noise margin, leakage, V_{ccmin} and access time of 6T SRAM cells. Using CACTI simulator modified for FinFETs and projected interconnect parameters, a system level comparison of L1 and L2 caches was carried out with symmetrically underlapped and optimized asymmetrically underlapped n-FinFET as bit-cell access transistor. Our systematic device-circuit-system co-design approach shows that for L1 cache, 11.3% reduction in cache access time along with 15.9% improvement in dynamic energy can be achieved. For L2 cache, we achieve 21.8% dynamic energy reduction and 4.6% leakage power reduction using our optimized asymmetric underlapped access n-FinFET.

REFERENCES

[1] B. Amelifard, F. Fallah and M. Pedram, "Leakage Minimization of SRAM Cells in a Dual-Vt and Dual-Tox Technology," IEEE TVLSI Sys., vol. 16, no. 7, pp. 851-860, July 2008.

[2] J. J. Kim et al., "Relaxing Conflict Between Read Stability and Writability in 6T SRAM Cell Using Asymmetric Transistors," IEEE EDL, vol. 30, no. 8, pp. 852-854, August 2009.

[3] C. H. Kim, J. J. Kim, S. Mukhopadhyay and K. Roy, "A Forward Body-Biased Low-Leakage SRAM Cache: Device, Circuit and Architecture Considerations," IEEE TVLSI, vol. 13, no. 3, pp. 349-357, March 2005.

[4] M. Qazi et al., "Challenges and Directions for Low-Voltage SRAM," IEEE Design and Test of Computers, vol. 28, no. 1, pp. 32-43, Jan 2011.

[5] A. Carlson et al., "SRAM Read/Write Margin Enhancements Using FinFETs," IEEE TVLSI, vol. 18, no. 6, pp. 887-900, 2010.

[6] S. Datta, "The Non-Equilibrium Green's Function (NEGF) Formalism: An Elementary Introduction," IEDM, 2002.

[7] Available: http://www.hpl.hp.com/research/cacti/

[8] C. Y. Lee and N. K. Jha, "CACTI-FinFET: An Integrated Delay and Power Modeling Framework for FinFET-Based Caches Under Process Variations," DAC, pp. 866-871, 2011.

[9] A. Shafaei et al., "FinCACTI: Architectural Analysis and Modeling of Caches with Deeply-Scaled FinFET Devices," ISVLSI 2014.

[10] A. A. Goud et al, "Atomistic Tight-Binding Based Evaluation of Impact of Gate Underlap on Source to Drain Tunneling in 5nm Gate Length Si FinFETs," IEEE DRC, p. 51,52, June 2013.

[11] S. K. Gupta and K. Roy, "Spacer Thickness Optimization for FinFET-based Logic and Memories: A Device-Circuit Co-design Approach," ECS Trans., vol. 50, no. 4, pp. 187-192, 2013.

[12] N. Neophytou and H. Kosina, "Atomistic Simulations of Low-field Mobility in Si Nanowires: Influence of Confinement and Orientation", Phys. Rev. B, 84, 085313, August 2011.

[13] Available: https://nanohub.org/resources/13117

[14] S. Gupta et al., "7-nm FinFET CMOS Design Enabled by Stress Engineering Using Si, Ge and Sn," IEEE TED, vol. 61, no. 5, May 2014.

[15] A. Khakifirooz and D. Antoniadis, "Transistor Performance Scaling: The Role of Virtual Source Velocity and Its Mobility Dependence," IEDM, pp. 1-4, December 2006.

[16] Available: http://www.itrs.net/Links/2011ITRS/Home2011.htm

[17] K. M. Cao et al., "BSIM4 Gate Leakage Model Including Source-Drain Partition," IEDM, pp. 815-818, December 2000.

[18] Y. C. Yeo, et. al., "Direct tunneling leakage current and scalability of alternative gate dielectrics," Appl. Phy. Lett., vol. 81, no. 11, Sep 2002.

[19] R. Pandey, et. al., "Impact of Single Trap Random Telegraph Noise on Heterojunction TFET SRAM Stability," IEEE EDL, vol. 35, no. 3, pp. 393-395, March 2014.

[20] Available: http://www.hpl.hp.com/techreports/2008/HPL-2008-20.pdf