PPA Scaling of Flip FET Technology Down to A2 Node Enabled by Architecture Innovations: Self-aligned Gate, 2T Design with Embedded Power Rail and Ultra-stacked 4-Tier Transistors

Wanyue Peng†, Haoran Lu†, Jingru Jiang, Rui Guo, Jiacheng Sun, Jianxiang Jin, Yuji Cheng, Shengcheng Zhou, Ziqiao Xu, Chuan Lan, Yanbang Chu, Xun Jiang, Feiyu Teng, Ming Li, Yibo Lin, Xinwei Wang, Runsheng Wang, Heng Wu*, Ru Huang
†These authors contributed equally, School of Integrated Circuits, Peking University, Beijing, China, *Email: hengwu@pku.edu.cn

Abstract: In this work, we carried out a critical examination of Flip FET (FFET) technology from process innovation to circuit design for nodes from A14 to A2. The Fully-aligned FFET (F3ET) featuring ***self-aligned (SA) FS and BS gates*** was experimentally demonstrated, with Common gate (CG) & Split Gate (SG) on fins and multi-stack Nanosheets. The ***Forksheet-based F3ET (F4ET)*** with embedded Power Rail (ePR) was proposed to ***reduce the cell height (CH) to 2T***. With ***ultra-stacked 4-tier transistors***, the Complementary FET (CFET) based FFET (CFFET) shows further area scaling potential. Comprehensive Power-Performance-Area (PPA) evaluation was conducted on both circuit- and block-level. DTCO knobs were also carefully studied on the A7 F3ET$_{NS}$ and A5 F4ET and improved the RO frequency (HP) by 11.3% and 6.2% at iso-leakage respectively. A3 HP F4ET outperforms A14 HP Fin-based FFET (FFET$_{Fin}$) by 38.9% at iso-power @ V_{dd} = 0.7 V. P&R results on 32-bit RISC-V cores shows 44.9% (HP) / 49.8% (HD) area scaling and 20.0% (HP) / 27.9% (HD) frequency improvement from A14 to A5. SRAM scaling down to A2 was also studied in a 256×256 array.

Introduction: With the ending of pitch scaling, the advanced logic development follows two trends: 3D transistor stacking [1-4] and backside interconnects [4-7]. We proposed the FFET [8] previously, combining stacked transistors and symmetric FS-BS signal & power for the first time, with clear parasitics and area benefits demonstrated [9-10]. In this work, we further proposed several crucial process innovations to drive the continuous scaling of FFET from A14 to A2 for both high performance (HP) and high density (HD) applications, as on the roadmap shown in Fig. 1. To solve the issue of misaligned frontside (FS) and backside (BS) gates [11], we introduced and experimentally demonstrated the F3ET with the critical SA gate and its performance gain was studied at A7. For A5 & A3, we combined the F3ET and Forksheet [12] with the ePR into F4ET, pushing the CH to 2T. Circuit- and block-level PPA evaluation validated the scalability of FFET down to A3. Moreover, CFFET, a new architecture with ultra-stacked 4-tier transistors [13] enabled by the unique dual-sided process of FFET, was introduced at A2. At last, the scaling potential of the FFET SRAM was also studied.

Device Fabrication of the F3ET: Fig. 2(a) shows the key steps of the Fin-based F3ET flow. The S/D placeholder (PH) and FS S/D are formed in sequency after the ***SA active and SA DG***, followed by the 1st Flip. Then, BS dummy gate (DG) is stripped after BS S/D epi, followed by the ***CG & SG*** patterning. After the ***dual-sided HK & TiN are deposited together***, the BS RMG and BEOL are formed, followed by the 2nd Flip [14]. At last, the FS RMG and other BEOL processes are finished. Note that the ***N/P S/D epi formed separately on each side*** (the key difference w.r.t the CFET) has no high aspect ratio (AR) challenges as in CFET [8] and ***the dual-sided metal gate is gate-last without thermal budget concerns***. Fig. 2(b) shows the TEMs of the SG (left, oxide on the FS) & CG (right) in F3ET$_{Fin}$. Fig. 2(c) give the SEM & TEM of the ***multi-stack Nanosheets*** with partial channel release, proving the Nanosheet-based F3ET (F3ET$_{NS}$) (Fig. 2(d)).

TCAD and Compact Model: We assumed the nFET/pFET on the FS/BS. Process assumptions of A14 FFET$_{Fin}$, A10 Nanosheet-based FFET (FFET$_{NS}$), A7 F3ET$_{NS}$, A5 & A3 F4ET and A2 CFFET are listed in Table 1 & 2. HP cells have larger CH and stronger current drivability than HD cells. All HP devices used for the ring oscillator (RO) simulation and library characterization have the same I_{off} of 2 nA while for HD devices the I_{off} is 20 pA [15]. BSIM-CMG [16] compact models (CMs) were fitted to TCAD I_dV_g (Fig. 3), I_dV_d and CV curves (not shown).

RO PP roadmaps towards A3: A 15-stage INVD1-based RO with fan-out (FO) of 3 and distributed BEOL loads was used to compare the circuit-level performance from A14 to A3. The BEOL loads were extracted from P&R critical paths (CPs) [14, 17] of 32-bit RISC-V cores, as discussed later in Fig. 8(c). Power-frequency curves of HP FFETs in Fig. 4(a) shows that A10, A7, A5 and A3 HP FFETs outperform A14 HP FFET$_{FIN}$ by 11.4%, 24.8%, 34.1% and 38.9% at iso-power @ V_{dd} = 0.7 V W/O BEOL load. Fig. 4(b) compares the I_{eff} vs C_{eff} relation of both HP and HD FFETs with various BEOL loads at iso-leakage @ V_{dd} = 0.7 V. A5 HP/HD F4ET outperforms A14 HP/HD FFET in speed by 85.6%/85.2% W/ short load. A3 and A5 FFET share the similar frequency while A3 FFET shows less power (-13.5% for HP and -24.2% for HD in average) due to the reduced C_{eff}.

F3ET DTCO: The SA gate in A7 F3ET$_{NS}$ also has great benefits. Fig. 5 (a) shows the F3ET$_{NS}$ has much wider gate connections compared to the FFET W/O SA gate at the same sheet width (W_{NS}) and thus smaller gate resistance (R_{gate}). Besides, the W_{NS} in F3ET$_{NS}$ can be further enlarged thanks to the elimination of gate merge vias [8]. These gives clear benefits in circuits with multistage cascade. For instance, for AND2D1 in Fig. 5(b), design v1 cascades the independent INV and NAND. The other more area-efficient design (v2) uses the CG of the INV to form the NAND, as in Fig. 5(c). Fig. 5(d) shows the circuit speed is affected more by the INV's R_{gate} for the v2 design and the RO results of the two designs. Fig. 5(e) shows the RO frequency of the v2 design is 2.4% higher than v1. Furthermore, thanks to the SA gate in F3ET$_{NS}$, the R_{gate} is reduced by 88.7% at iso-W_{NS}, gaining an extra 1.6% frequency. Also, F3ET$_{NS}$ has larger W_{NS} and thus 7.2% higher frequency (+ 11.3% w.r.t v1) with the R_{gate} considerably small.

F4ET DTCO: For further CH shrinking, Forksheet is further introduced into A5 FFET while keeping the active footprint [12]. Also, power rails at the cell boundary (Fig. 1), limiting the CH scaling, are embedded into the dielectric wall of the Forksheet (Fig. 6(a)). This helps ***reduce the CH from 2.5T [8] to 2T*** in A5 HD F4ET. BS contact (BSC) (Fig. 6(b)) is used to connect the S/D epi and the ePR. BSC optimization such as the low temperature (LT) regrowth of heavily-doped epi [18], replacing BSC metals [18] and using wrapped around contact (WAC) [4] (Fig. 6(b)) help reduce the N/P source resistance by 43.9%/27.4% and improve the RO frequency by 6.2% (Fig. 6(c)). Performance degradation in A5 HP F4ET caused by the IR-drop on the ePR resistance (R_{ePR}) was also evaluated by the RO in Fig. 6(d). Switching the ePR metal from W to Ru [19] and increasing ePR thickness can reduce the R_{ePR} by 68.9% and the frequency degradation w.r.t ignoring IR-drop can be limited to 0.8% with worst-case IR-drop ~ 16 mV (Figs. 6(e-f)).

FFET vs CFET W/ BSC: Like FFET, CFET with BSC has been proposed recently [2, 20-22], with better design flexibility [23]. As shown in Fig.7 (a), CFET W/O BSC (v1) and two types of CFETs W/ BSC (CFET W/ TSVM & SA BSC [3-4] (v2.1) and CFET W/ symmetric FS-BS interconnects (v2.2) [22]) are compared with FFET at the same CH at A7. The FFET shows area gain over all the CFETs in cells with SG [8] (Fig. 7(b)). HD F3ET$_{NS}$ and HD CFET v2.2 share similar RO performance and outperform other two CFETs (Fig. 7(c)) for larger W_{NS} (Fig. 7(a)) and reduced parasitics thanks to the dual-sided power [3, 8-10]. The HP F3ET$_{NS}$ gains an extra 7.3% higher frequency at iso-leakage for its smaller gate extension compared to the HP CFET v2.2 without BS COAG [22] (Fig. 7(d)). Furthermore, with input pins on the BS, CFET v2.2 shows worse performance due to the larger BS parasitics (Fig. 7(c)) while F3ET$_{NS}$ doesn't. The relation between CH and M0 pitch among all FFETs and CFETs in Fig. 7(e) shows the FFET has better scalability than all the CFETs thanks to the F4ET's most relaxed M0 pitch.

Block-level PPA Scaling: Physical implementation (including dual-sided P&R) and PP evaluation [24] were carried out on 32-bit RISC-V cores based on A14-A5 FFET libraries @ V_{dd} = 0.7 V. BEOL design rules are listed in Table 3. Post P&R HP core layouts in Fig. 8(a) shows the continuous area scaling down to A5, consistent with Fig. 8(b). Statistics of BEOL layers in CPs are given in Fig. 8(c). Compared to the A14 HP FFET$_{FIN}$, the A10, A7 and A5 HP FFET exhibit continuous scaling in performance (max 20% at iso-power) and energy efficiency (max >15% of average EDP), as in the Figs. 9(a-c). Fig. 9(d) shows the A5 HD F4ET has 49.8% smaller core area, 27.9% higher frequency and 13.1% smaller EDP compared to the A14 HD FFET$_{Fin}$, further validating the PPA benefits.

Ultra-stacked Transistors at A2: ***First of a kind 4-tier transistors with back-to-back-stacked CFETs***, the CFFET is an ultimate extension of the FFET technology (Fig. 10(a)). The dual-sided signal & power, the VHV routing scheme [25] and the super vias enable the aggressive intra-cell routing. Nanosheet, instead of Forksheet, is more suitable for the CFFET due to the needs of super S/D vias on each side of the active. The sheet # (N_{NS}) of the CFFET is limited to 2 considering the high AR constraints. A2 HP CFFET's cell area shows ~8.4% area gain over A3 HP F4ET (Fig. 10(b)), while HD cells are similar (Fig. 10 (c)). Unfortunately, INVD2-based RO results show the A2 HP and HD CFFET's frequency at iso-power degrades by 9.6% and 23.1% @ V_{dd} = 0.7 V w.r.t the A3 ones respectively (Fig. 10(d-e)) as CFFET has fewer N_{NS}, leading to smaller W_{eff} and I_{eff} while C_{eff} is not reduced (Fig. 10(f)) due to the larger parasitcis of the super vias.

SRAM Scaling Down to A2: A SRAM array with 256 rows (WL) and 256 columns (BL) was used for the SRAM analysis. Fig. 11(a) shows the FFET SRAM scaling roadmap towards A2 beyond conventional non-stacked transistors [26-27]. The unique Bipolar SRAM by FFET [8] is a little smaller than the Unipolar SRAM [8] due to the removal of the super vias (Fig. 11(b)). Fig. 11(c) gives the CFFET SRAM, with two stacked CFET SRAMs [28]. Thus, the A2 CFFET SRAM array can be column-folded (256 rows×128 cols) or row-folded (128 rows×256 cols) (Fig. 11 (d)), gaining 50% area at the same memory size. BL & WL RC (Fig. 11(e)) were considered in the worst-case cell simulation, as given in Fig. 12. Bipolar SRAM exhibits smaller write delay than Unipolar SRAM at A14-A5 for the smaller WL C & BL R and smaller read delay at A7-A5 for the reduced BL C. However, the write speed and write margin of Bipolar SRAM degrades greatly at A3 for the significant BL R degradation, which can be recovered for A2 CFFET SRAM with folded row design due to halved BL RC.

Conclusion: With architecture innovations, the FFET technology further extends into ***the F3ET with SA gate*** and ***the F4ET with smallest 2T design*** enabled by the ePR. Through comprehensive PPA evaluation, we validated the great scalability of the FFET down to A3 node (better than the CFET W/ BSC). Though ***A2 CFFET with ultra-stacked transistors*** has some area gains but no obvious PP gains for logic application, it shows potential to push the SRAM scaling beyond A2.

Acknowledgements: This work was supported by the National Key R&D Program of China under Grant 2023YFB4402200; and in part by the 111 Project under Grant QYJS-2023-2301-B; and in part by the National Natural Science Foundation of China under Grant 92464206.

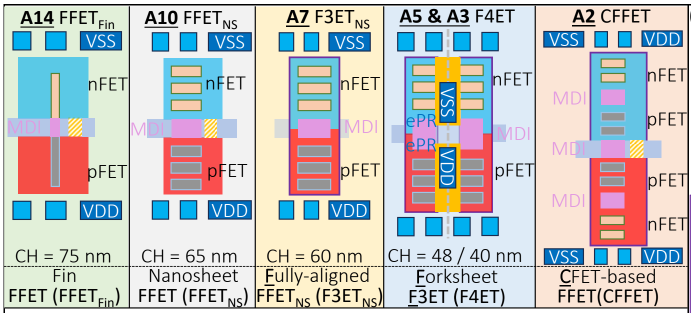

Fig. 1 FFET scaling roadmap from A14 to A2 node. A7 F3ET$_{NS}$ features a unique self-aligned gate. Combining Forksheet together with ePR in the dielectric wall region in A5 & A3 F4ETs pushes the CH to 2T. A2 CFFET with aggressively ultra-stacked 4-tier transistors is also considered for ultimate scaling.

Node	A14	A10	A7	A5 & A3	A2
Device	FFET$_{Fin}$	FFET$_{NS}$	F3ET$_{NS}$	F4ET	CFFET
Struct. stacked	FinFET	Nanosheet	Nanosheet	Forksheet	Nanosheet CFET
Stacked FET #	2	2	2	2	4
CPP (nm)	51	48	45	45	45
M0 pitch (nm)	28	22	20	24/20(A5/A3)	16
M2 pitch (nm)	30	26	24	24/20(A5/A3)	20
Track #	3.5/2.5(HP/HD)	3.5/2.5(HP/HD)	3.5/2.5(HP/HD)	3/2(HP/HD)	4/3(HP/HD)
Process knobs	Baseline	NSFET	SA-Gate	ePR & BSC	Ultra-stacked FETs

Table 1 Design rules and process knobs for FFET from A14 to A2.

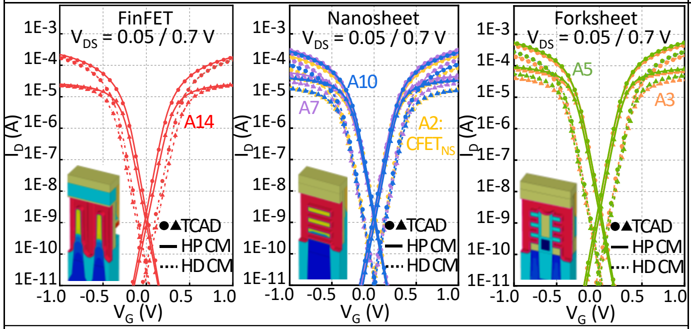

Fig. 3 I_dV_g curves of TCAD and BSIM-CMG compact models fitted.

Node	A14	A10	A7	A5 & A3	A2
Device	FinFET	Nanosheet	Nanosheet	Forksheet	CFET$_{NS}$
L_g (nm)	18	16	15	15	15
$L_{gate\ metal}$ (nm)	14.4	12.4	11.8	11.8	11.8
N_{fin} / N_{NS}	2(HP)/1(HD)	3	3	3	2

H_{Fin} / W_{NS} (nm)	50	32(HP)
		15(HD)
			36&28(HP)
			21(HD)
				34/28(A3 HP/A2 HP)
				22/16(A3 HD/A2 HD)
					30(HP)
					14(HD)

T_{Fin} / T_{NS} (nm)	6	5	5	5	5
Gate Spacer Thk. (nm)	7	6	5.5	5.5	5.5
Inner Spacer Thk. (nm)	/	5	6.5	6.5	6.5
EOT (nm)	0.95	0.95	0.91	0.91	0.91
I_{off} (nA)	2 (HP) / 0.02 (HD)				

Table 2 Device assumptions for HP & HD FFETs from A14 to A2.

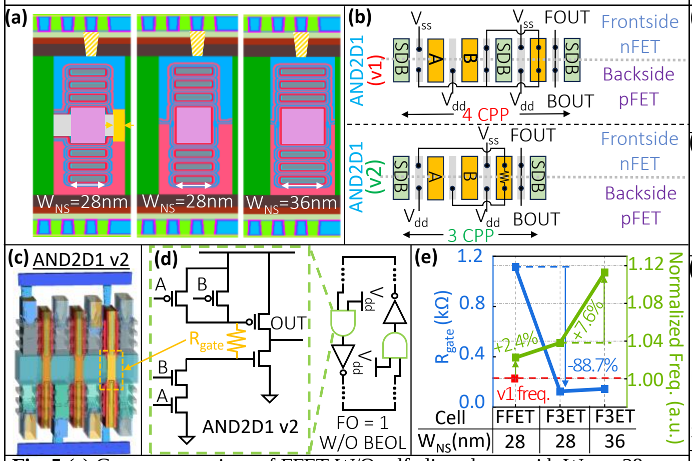

Fig. 5 (a) Gate cross sections of FFET W/O self-aligned gate with W_{NS} = 28 nm, F3ETs with W_{NS} = 28/36 nm, all based on A7 design rule. Thanks to the SA gate, F3ET can support wide $W_{NS.}$ (b) Two AND2D1 designs. The v2 design occupies less cell area than the v1 design by intra-cell routing optimization. (c) 3D schematic of the v2 design. (d) The speed of AND2D1 v2 is affected more by the INV's R_{gate}. (left) A HP AND2D1&INVD1 based RO was used to evaluate the speed. (right) (e) Reduced R_{gate} and enlarged W_{NS} improve the frequency of the F3ET AND2D1.

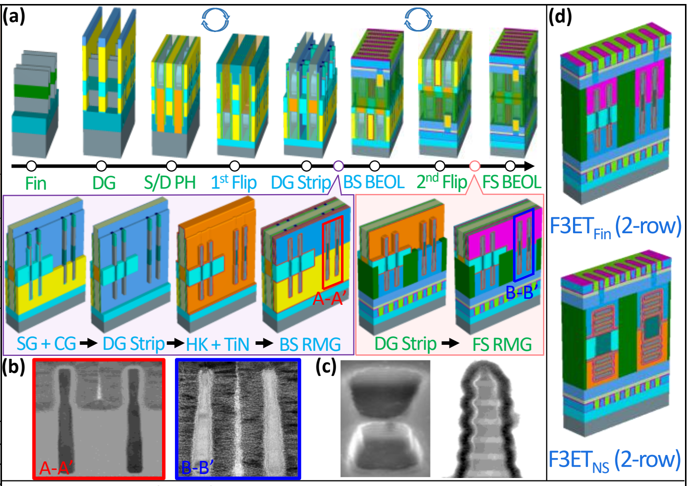

Fig. 2 (a) Process flow of the novel F3ET$_{Fin}$ with SA active and gate. S/D PH & FS S/D are formed after SA active and SA DG, followed by the 1st Flip. BS DG is stripped after BS S/D and CG & SG are then formed. After dual-sided common HK & TiN, BS RMG & BEOL are done, followed by the 2nd Flip and FS RMG. (b) TEMs of SG (left) & CG (right) in F3ET$_{Fin}$. (c) SEMs & TEMs of the partially released multi-stack Nanosheet for F3ET$_{GAA}$. (d) 3D structures of the F3ET$_{FIN}$ & F3ET$_{NS}$.

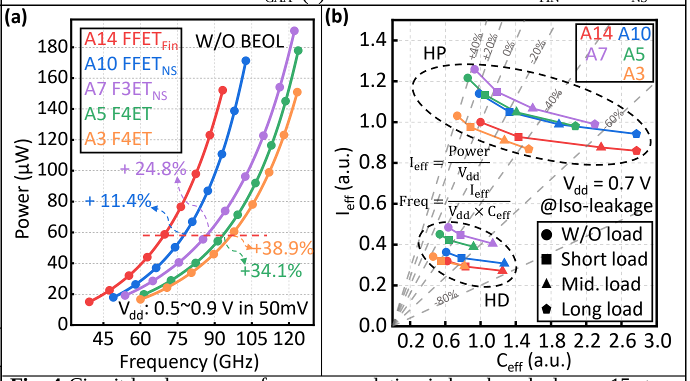

Fig. 4 Circuit-level power-performance evolution is benchmarked on a 15-stage INVD1-based RO with FO 3 and distributed BEOL loads extracted from a 32-bit RISC-V core (discussed later in Fig. 8(c)). (a) Power-frequency curves of the HP FFET W/O BEOL load. A10/A7/A5/A3 HP FFET outperforms A14 HP FFET$_{FIN}$ by 11.4%, 24.8%, 34.1% and 38.9% at iso-power @ V_{dd} = 0.7 V. (b) Effective current (I_{eff}) and capacitance (C_{eff}) relation at iso-leakage @ V_{dd} = 0.7V. I_{eff} and C_{eff} are normalized to A14 HP FFET$_{Fin}$ W/O load. A5 F4ET exhibits superior frequency compared to A14/A10/A7 FFETs across all BEOL loads. HP and HD FFETs share the similar frequency at A5 and A3 while A3 FFET has less power (-13.5% for HP, -24.2% for HD in average) due to reduced C_{eff}.

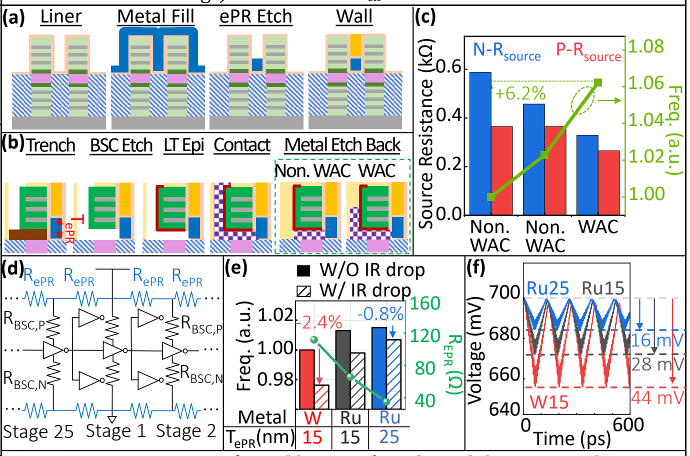

Fig. 6 (a) In F4ET, ePR is formed by a conformal metal deposition and isotropic etch back in the dielectric wall region. (b) Process flow of the F4ET BSC. The WAC increases the BSC surface area. (c) Replacing N-type BSC metal [18] and using WAC reduce A5 HP F4ET N/P source resistances by 43.9%/27.4% and improve the RO frequency by 6.2%. (d) A 25-stage RO (ePR length ≈ 48 CPP [7]) to evaluate ePR IR-drop. (e) ePR metal optimization reduces R_{ePR} and recovers the frequency degradation caused by IR-drop. (f) The worst-case (Stage 13) power transients.

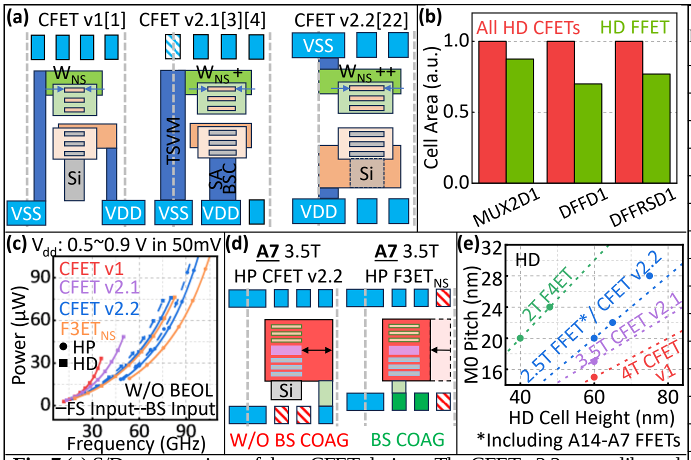

Fig. 7 (a) S/D cross sections of three CFET designs. The CFET v2.2 was calibrated to [22]. (b) The FFET shows area gain over all the CFETs in cells with SG, taking three HD cells as examples. (c) Power-frequency curves comparing $F3ET_{NS}$ and the three CFETs in (a) at A7. $F3ET_{NS}$ shows the best performance due to its largest W_{NS} and smallest parasitics, especially with BS input pin. (d) 3.5T HP CFET v2.2 needs larger gate extension due to the lack of BS COAG, causing larger gate cap. (e) M0 pitch vs HD CH. The FFET (F4ET) shows the best area scalability.

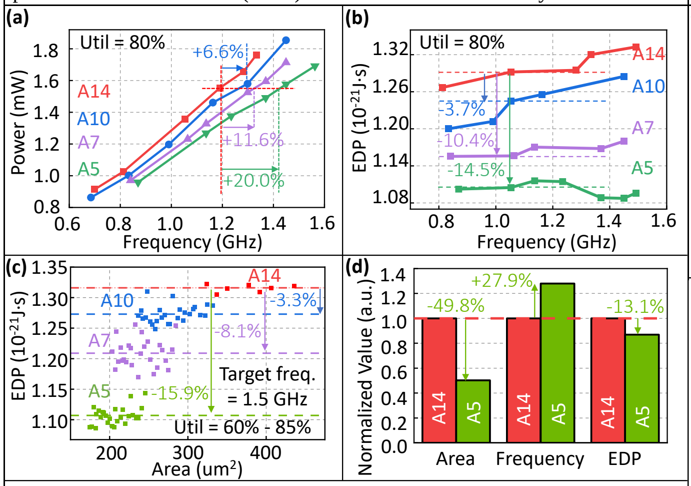

Fig. 9 Block-level PPA results from A14 to A5 FFET RISC-V cores: (a) Power-frequency plot of HP FFET cores at utilization = 80%. (b) EDP vs achieved frequency in HP FFET cores at utilization = 80%. (c) EDP vs core area in HP FFET cores at target frequency = 1.5 GHz and utilization from 60% to 85%. (d) Area, achieved frequency and EDP comparison between the A14 HD $FFET_{Fin}$ and A5 HD F4ET at the same target frequency = 800 MHz.

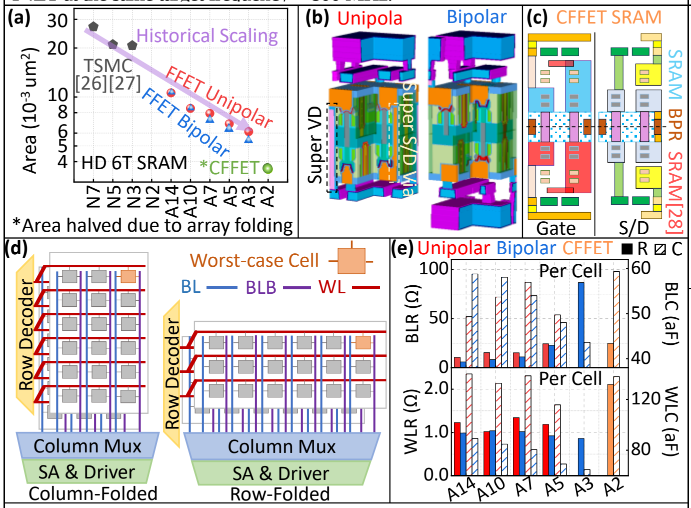

Fig. 11 (a) The SRAM scaling roadmap down to A2. (b) 3D structures of the FFET Unipolar and Bipolar SRAM (smaller due to no super vias). (c) Schematics of the CFFET SRAM in gate and S/D cross sections. BPR is shared by both sides. (d) Two folding schemes for the CFFET SRAM array. (e) WL & BL RC per bit cell at A14-A2. A3 Bipolar SRAM BL R degrades greatly for its narrowest BL.

Layer	Pitch (nm)			
	A14	A10	A7	A5
M12	720	126	126	
M11	720	126	76	
M10	126	76	76	
M9	76	76	76	
M8	76	76	76	
M7	76	76	76	
M6	76	76	76	
M5	76	42	42	
M4	42	35	35	
M3	42	35	32	
M2	30	26	24	
M1	34	32	30	
M0	28	22	20	24
ePR	/	/	/	96

Table. 3 BEOL metal design rule, shared between FS and BS BEOL. The ePR pitch equals to 2 HD CH (i.e. 4 M2 Pitch).

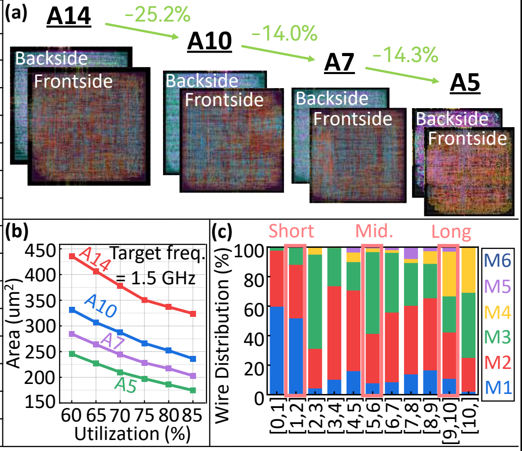

Fig. 8 Physical implementation and PP evaluation were carried out on 32-bit RISC-V cores based on A14-A5 FFET libraries @ Vdd = 0.7 V (each has > 50 cells), with utilization = 80% and target frequency = 1.5 GHz: (a) Dual-sided HP core layouts. (b) Area vs utilization. (c) BEOL layer distribution for nets in the critical paths. FS and BS metals at the same level (e.g. FS M1 & BS M1) are considered the same.

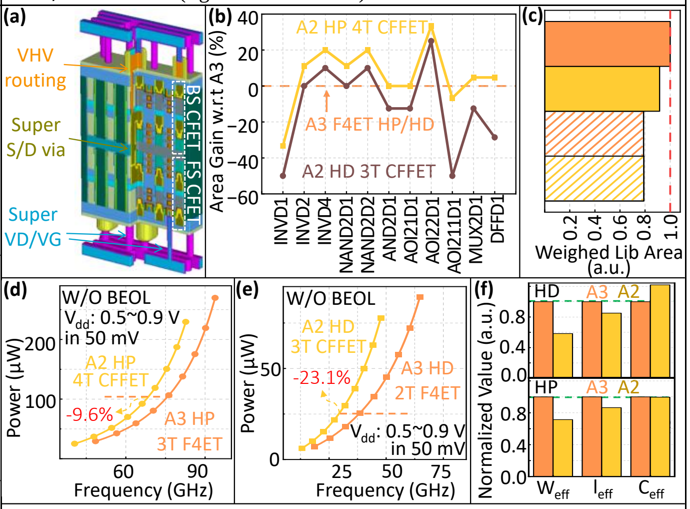

Fig. 10 (a) 3D schematic of the CFFET AOI211D2. (b) Standard cell area comparison between A3 F4ET and A2 CFFET. (c) Weighed library area (extracted from RISC-V cores) scaling from A3 to A2. Power-frequency curves of the 15-stage INVD2-based RO comparing A3 F4ET and A2 CFFET in (d) HP and (e) HD cells. (f) W_{eff}, I_{eff} and C_{eff} of the A3 F4ET and the A2 CFFET.

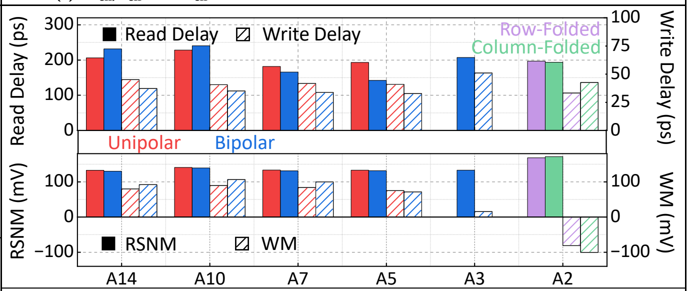

Fig. 12 FoMs of the worst-case bitcell in a 256×256 SRAM array. The I_{off} of SRAM devices is 2 pA. A2 CFFET shows reasonable performance and noise margin due to the halved BL/WL length (halved RC) in row-/column-folded designs respectively.

Reference

[1] J. Ryckaert et al., *VLSI 2018*. [2] M. Kobrinsky et al., *IEDM 2023*. [3] A. Pal et al., *IEDM 2024*. [4] A. Veloso et al., *IEDM 2023*. [5] H. Fukutome et al., *VLSI 2024*. [6] W. Hafez et al., *VLSI 2023*. [7] R. Chen et al., *IEDM 2022*. [8] H. Lu et al., *VLSI 2024*. [9] Y. Lu et al., *IEDM 2024*. [10] J. Lee et al., *IEDM 2024*. [11] W. Peng et al., *EDTM 2025*. [12] P. Weckx et al., *IEDM 2019*. [13] L. Liebmann et al., *IEDM 2021*. [14] H. Wu et al., *EDTM 2025*. [15] L. Clark et al., *Microelectronics Journal 2016*. [16] N. Paydavosi et al., *IEEE Access 2013*. [17] A. Farokhnejad et al., *IITC 2022*. [18] C. Porret et al., *IEDM 2022*. [19] D. Gall, *VLSI-TSA 2020*. [20] J. Park et al., *VLSI 2024*. [21] A. Vandooren et al., *IEDM 2024*. [22] S. Liao et al., *IEDM 2024*. [23] H. Kükner et al., *IEDM 2024*. [24] H. Lu et al., *DATE 2025*. [25] B. Chehab et al., *IITC 2021*. [26] J. Chang et al., *ISSCC 2017*. [27] S. Wu et al., *IEDM 2022*. [28] H. Liu et al., *TED 2023*.