A 16nm pFET Floating Gate Transistor Enabling Programmable Analog Design in a FinFET Logic CMOS Process

Praveen Raj Ayyappan, Afolabi Ige, Pranav O. Mathews, Linhao Yang, Jennifer O. Hasler
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
{payyappan3,aige3,pmathews30,lyang319}@gatech.edu and jennifer.hasler@ece.gatech.edu

***Abstract*—This work presents the first demonstration of Floating Gate (FG) pFET devices integrated into crossbars and analog circuits in a 16nm FinFET logic CMOS process with no additional fabrication masks or steps. This FinFET FG demonstrates improved device-level performance, including the 1000× programming speed of the FG pFETs due to the higher impact-ionization efficiency compared to planar nodes, while lowering programming voltages due to the lower Si-HfO_2 energy barrier. This work presents the first demonstration of FinFET FG-based programmable analog standard cells and the first demonstration of programmable analog circuits and analog classifiers in a standard FinFET CMOS process, setting the foundation for synthesizing large-scale analog systems.**

I. FGs for Programmable Analog Design

This work demonstrates FinFET Floating-Gate (FG) pFETs that achieves 10× higher integration density and 1000× faster write speeds compared to similar devices fabricated in planar nodes [1] as well as demonstrating circuits with analog parameters programmed using these devices. FG transistors provide reliable Non-Volatile Memory (NVM) capabilities within standard logic CMOS process nodes [2] and facilitate programmable analog & mixed-signal computation, enabling Compute-In-Memory (CIM) approaches that perform Vector-Matrix Multiplication (VMM) with significantly low power consumption compared to traditional digital implementations [3]. The rapid write speeds permit the use of lower programming voltages for a FG pFET cell that requires no additional fabrication layers from the standard FinFET CMOS process.

This work demonstrates analog circuits using programmable FG FinFET pFETs setting bias currents (subthreshold or above threshold), and compensating for analog mismatches. FinFET FG programmability enables the customization of circuit specifications post-fabrication in a 16nm FinFET CMOS process, including eliminating threshold-voltage mismatch that limits efficient analog designs. These FG devices offer precise programming of bias currents spanning pA to μA levels (greater than 8 orders of magnitude).

This work demonstrates the first 16nm CMOS FinFET-based programmable analog standard cell library, and presents measured results for all the circuits implemented with it. Fine-grain FG device (Fig. 1) programming facilitates the abstraction of circuits that enables analog standard cells [1]. These standard cells are consistent with earlier analog standard-cell libraries, enabling high-level synthesis of large-scale analog systems [4]. The FG pFET device was the critical innovation to enable these cells that can be densely integrated into a crossbar array (e.g. CiM applications), integrated into programmable analog circuits such as Transconductance Amplifiers (TA), and combinations of these concepts, such as an analog classifier (e.g. one-layer XOR solution). These FG FinFETs enable the eventual demonstration of densely integrated, high-performance analog computing benchmark systems [5].

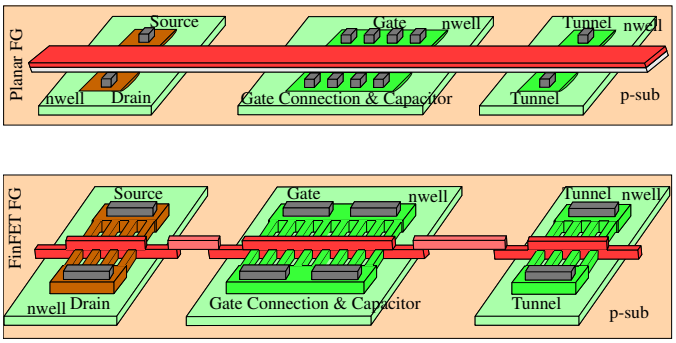


Fig. 1: A 3D physical view of a planar and FinFET FG device. FinFET FGs are constructed with aggressive DRC rules. Unlike planar FGs, FinFET FGs have poly running vertically and fins in horizontal direction. Additionally, the poly layer is also not allowed to change its layout orientation, which reduces the options for FG design in FinFET nodes.

II. Floating-Gates in FinFET Processes

Our FG device architecture incorporates a High-Voltage (HV) pFET and two MOS capacitors. One capacitor is the control gate coupling capacitor and the other capacitor facilitates Fowler-Nordheim (FN) tunneling for erase operations. The thicker gate oxide from a HV devices provides reliable long-term charge retention; standard low-voltage devices often exhibit charge leakage requiring refresh mechanisms [6]. Both MOS capacitors use minimum design rule dimensions (four fins each) resulting in the equal effective coupling from the control gate and tunneling line (Fig. 2a). A one-fin design, similar to those used in SRAMs, would decrease the tunneling

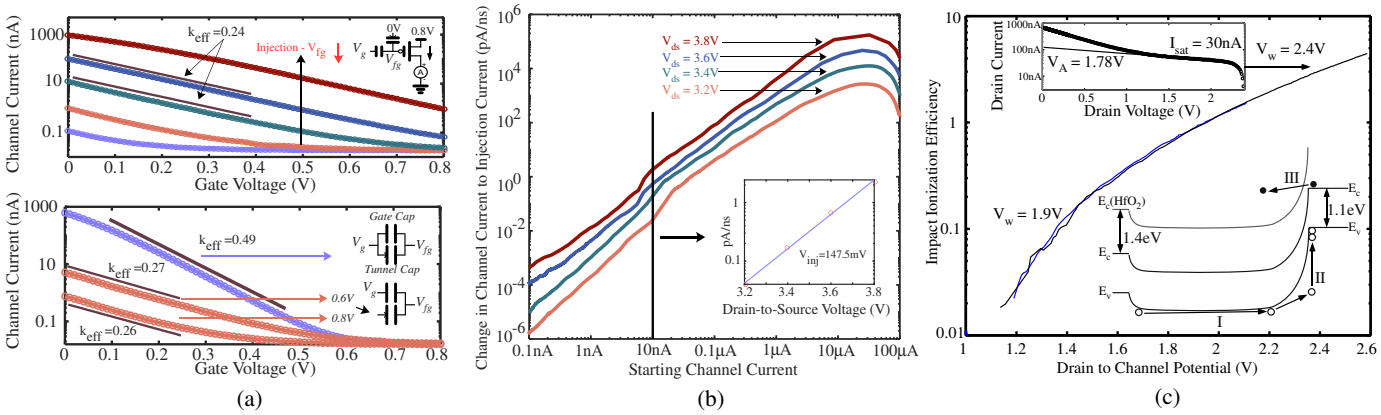


Fig. 2: FinFET FG Device Characterization. (a) Gate sweeps showing precise logarithmic current control via electron tunneling/injection. Parallel gate/tunnel capacitors double Vfg coupling. (b) Measured injection current for various drain to source voltage (Vinj) by pulsing drain at 100ns, demonstrating exponential dependence [2]. (c) Impact ionization current in an HV PFET via drain sweep, showing exponential increase and higher efficiency at lower drain potentials.

capacitor coupling. The control gate coupling is the ratio of this capacitance to the total capacitance at the FG node (C_T) that directly corresponds to the ratio of the number of fins allocated to each MOS capacitor. We estimate $C_T \approx 1-2fF$ from extraction, resulting in the maximum electron precision for a 1V FG voltage range of 13bits; FG programming accuracy has been shown to be 12-14bit accurate previously [7], and these measured 16nm FinFET FG measurements demonstrate similar precision as expected.

These FG FinFET pFETs are programmed using hot-electron injection, enabling high-resolution [2], compared with bi-directional FN tunneling, with its associated high mismatch, which limits the precision of nFET FG devices to 4bits in 14nm FinFET logic process [8], typical of EEPROM devices. Our FG precision is highly suitable for biasing analog circuits and constructing high-resolution crossbar arrays. The control gate coupling of these FG FinFET pFETs is built with area efficient fin-ratio based method, enabling good capacitor area matching, important for analog and digital circuits. Other capacitor couplings are possible, such as lateral slot contacts reported in the 16nm FinFET logic process [9], that typically require increased gate width and thus a larger cell area, as well as lower gate-coupling matching. Extending lateral coupling via slot contacts for the gate capacitor [9] to crossbar arrays would likely incur a significant area penalty.

III. Floating-Gates Programming Mechanisms

For FN tunneling, a significant voltage (V_{tun}) is applied to the tunneling MOS capacitor to establish a high potential difference across the insulator. FN tunneling is observed to begin at V_{tun}= 4V, and the erase speed increases as V_{tun} is raised. In all experiments, applying V_{tun}= 5.5V for a duration of 1 to 5 seconds was sufficient to erase the devices, resulting in post-erase leakage currents below 100 pA. As FN tunneling is typically implemented as a global erasure mechanism, the required erase time per device is independent of the total number of FG devices being erased simultaneously.

Hot-electron injection involves the injection of hot electrons, generated within the pFET channel region under a high drain-to-channel potential, into the FG node. This process increases

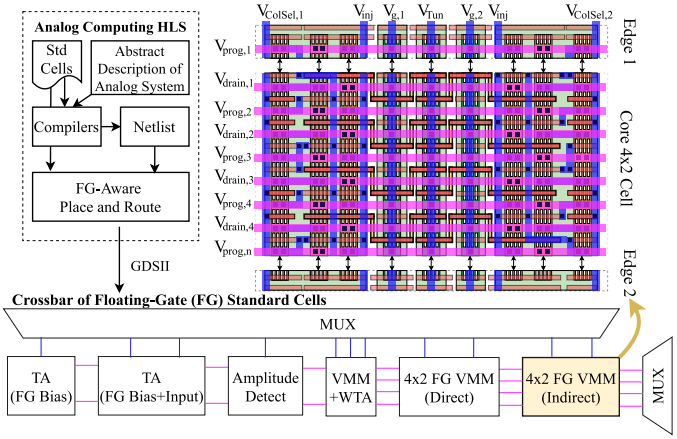


Fig. 3: Analog High Level Synthesis (HLS) enables abstract design. A crossbar of 16nm FG standard cells designed to work within an HLS framework was built.

the stored negative charge, thereby lowering the floating gate voltage (V_{fg}). For this specific process node, effective electron injection is achievable using programming voltages (V_{inj}) below 4V due to significantly higher injection rates (Fig. 2b). Compared to 65nm planar FG devices [1] operating under similar conditions, the injection current measured at V_{inj}= 3.8V is approximately 1000 times higher in our FinFET-based device, indicating a substantial increase in programming speed (Fig. 2b). This enhancement in programming speed is attributed to the higher impact ionization efficiency characteristic of this FinFET process, which becomes prominent at lower operating voltages relative to planar technologies (Fig. 2c).

IV. Programmable Analog Standard Cells

Programmable analog standard cell structure is set based on the dense FG crossbar. A core 4x2 FG array sets the standard cell pitch (2.86μm) that is used for all other analog standard cells. The eight FG devices in the 4x2 cell uses the minimum of eight polysilicon lines, bridging between poly with an intermediate layer to pass DRC (Fig. 3). This 4x2 core cell is the unit for creating larger crossbar arrays. Similarly to standard digital cell practices, dedicated boundary

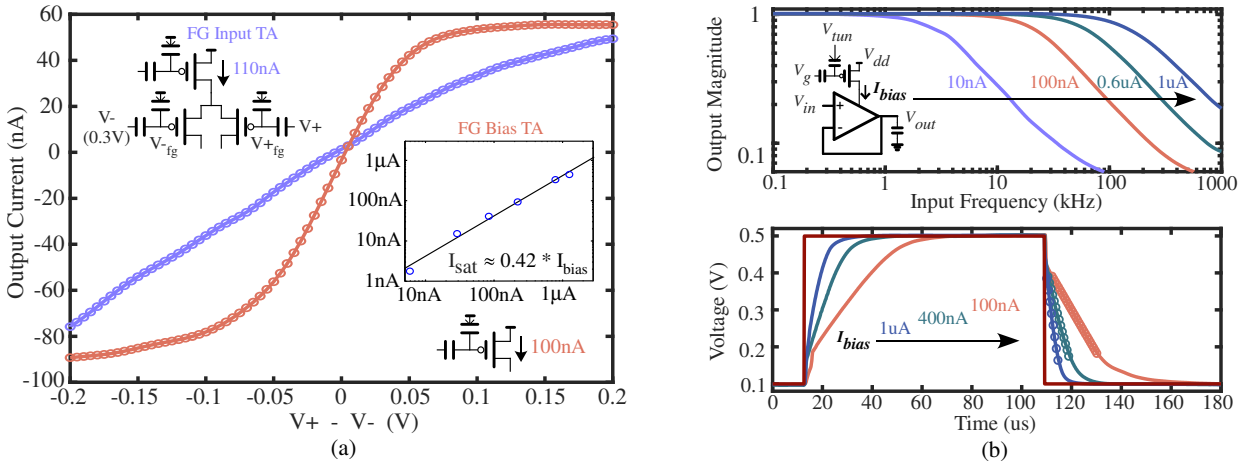


Fig. 4: Programmable Transconductance Amplifiers (TAs). (a) Open-loop output current comparison: FG TA programmed for similar input offset as non-FG input TA. FG input shows a wide linear range; saturation current of the TA correlates with programmed bias (subthreshold). (b) Programmable TA specifications: Gain Bandwidth Product (GBW) and Slew Rate (SR) are linearly proportional to programmed bias current (subthreshold), enabling speed-power trade-offs.

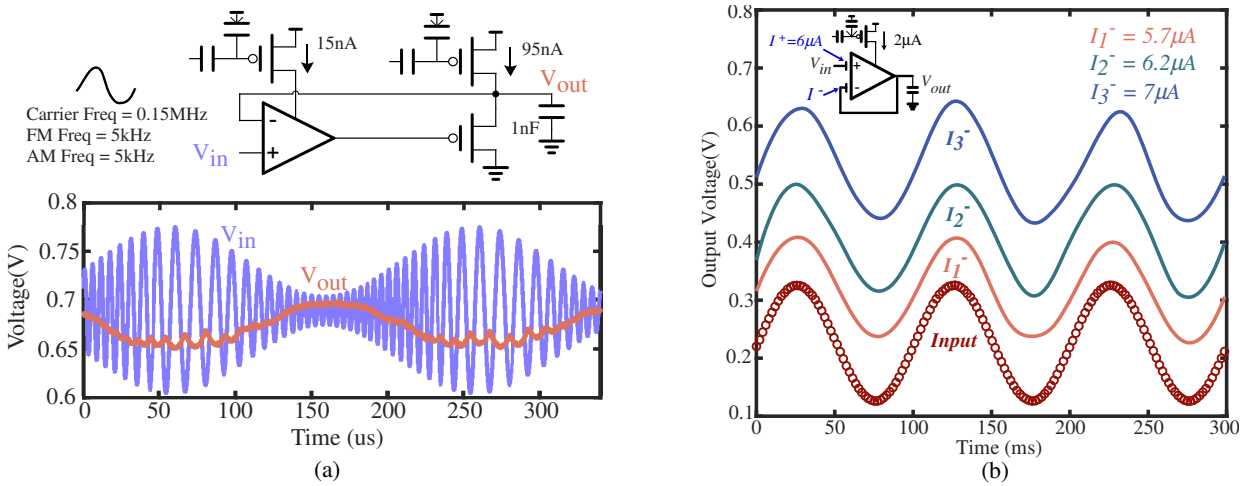


Fig. 5: Programmable TA Applications. (a) Amplitude detector circuit (TA + source-follower stage). Programmed TA bias sets a time constant larger than the source-follower stage for minimum detection, suitable for speech processing front-ends. (b) Voltage buffer using FG TA. Programmable input offset enables direct positive DC level shifting at the output.

cells must be incorporated in the peripheries of the FG array (Fig. 3). The programming methodology employs a shared infrastructure for accessing and programming FG devices across all integrated circuits, forming a single large crossbar (Fig. 3) where each FG device can be individually addressed, isolated, and programmed independently.

TAs are important programmable analog standard cells that are biased by FG devices to enable programmable currents both in the subthreshold and above threshold regimes. The fundamental topology is a nine-transistor TA with a pFET differential pair & a FG pFET current source setting the bias current. Employing a programmable FG pFET to set the bias current provides direct control over the amplifier's Gain BandWidth product (GBW) and slew rate, since both parameters scale proportionally with this programmed current level (Fig. 4b). Some differential pair inputs use FG pFETs to enable a capacitive divider at the gate to increase linearity compared to a standard input (Fig. 4a).

V. Circuits from Standard Cell Circuits

The standard cell components can be combined to develop larger circuits. Programmable TAs are often used as programmable amplifiers or current sources. For example, an amplitude detection circuit can be created with a programmable TA and an additional programmable source follower stage. This circuit can act as a maximum or minimum detector depending on the decay rate set by the FG bias current of the TA and the attack or rising rate set by the ratio of the source follower bias current and output load capacitance (Fig. 5a); the operation is entirely dependent on controllable FG programming. Another example uses a single FG TA in negative feedback as a level shifter or programmable voltage bias. By deliberately inducing an offset onto the differential pair, the output voltage will shift proportionally (Fig. 5b). When used with a dynamic input signal this acts as a level shifter, but when a DC bias is tied to V+, this circuit can be used as a programmable static voltage source.

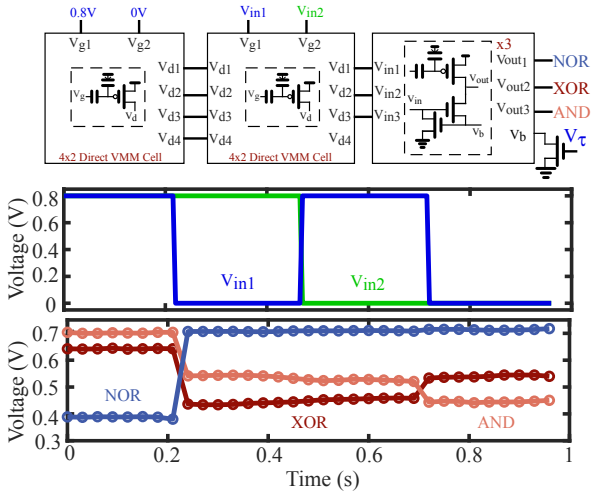


Fig. 6: Demonstration of a single layer universal classifier comprised of a couple 4x2 Direct VMM cells and a Winner-take-all (WTA) network which classifies the non-linear XOR problem. The WTA branch with higher input current inhibits the other stages through the shared node V_b and corresponds to a winning lower output voltage. The WTA speed is primarily limited by the time constant at the output node. In this implementation, outputs were directly connected to pads since functionality demonstration was prioritized over speed.

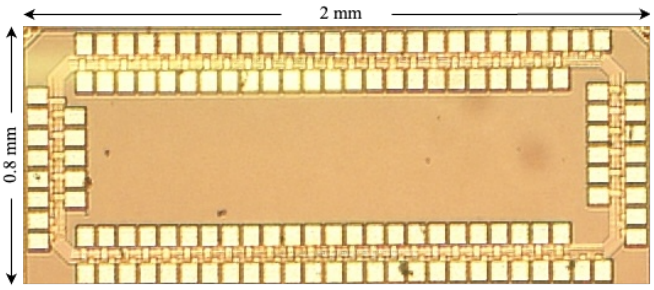
Efficient analog computing circuits show the opportunities of FG pFET devices, including signal classification and inference. When combined with another classic analog computing circuit, the Winner-Take-All (WTA), a standard VMM-WTA cell can be created that functions as a one-layer universal classifier. We demonstrate this computation by classifying the nonlinear XOR function (Fig. 6). The high density of our 16nm FG cells suggests that large analog classifiers can be built in relatively low area through FG-enabled VMM-WTA systems. CiM techniques utilize crossbars of NVM like FGs to efficiently perform VMM, a core operation in machine learning applications.

VI. Conclusion

This work demonstrates a high bit precision in FinFET FG pFET device in a logic CMOS process with no additional fabrication masks or layers. The FG devices scaled exceptionally well in FinFET nodes, achieving 10x higher density (Fig. 7b) and 1000x faster programming speeds compared to 65nm FGs. This work demonstrates the viability of using FinFET FG pFETs for programmable dense analog design components (e.g. TAs) and building FG crossbars (e.g. CiM VMM) with programmable analog circuitry in an analog classifier. This work demonstrates the first 16nm CMOS FinFET-based programmable analog standard cell library, paving the way for demonstrating large-scale, high-performance analog systems [5] using established synthesis frameworks [4].

References

[1] P. Mathews, P. R. Ayyappan, A. Ige, S. Bhattacharyya, L. Yang, and J. Hasler, "A 65nm and 130nm cmos programmable analog standard cell library for scalable system synthesis," in *2024 IEEE Custom Integrated Circuits Conference (CICC)*, 2024, pp. 1–2.



(a)

		This work	CICC 24 [1]	TCAS 24 [10]
Process Node		16 nm	65 nm	130 nm
Programmable Biasing?		Yes	Yes	Yes
Voltage (V)	Supply	0.8	1.8	1.6
	Injection	3	3.25	4
	Tunneling	4.5	6.25	8
FG TA	Input range (mV)	50 – 400	85 – 850	–
	Bandwidth (kHz)	0.1 – 200	1.8 – 14	–
	Power (W)	8 – 800 n	1.1 – 7.6 μ	–
Cell Pitch (µm)		2.86	6.50	6.50
Area (μm^2)	Direct 4x2 VMM	12.34	119.11	130.8
	Indirect 4x2 VMM	17.82	136.8	–
	2 – TA (FG bias)	13.16	116.2	116.5
	2 – FG input TA	26.07	175.4	182.6
	Amp Detect	12.72	–	–
	VMMWTA	46.75	288.46	–

(b)

Fig. 7: Die photo and performance comparison. (a) Die photograph of the implemented chip. (b) Comparison of key analog processing parameters with prior work.

[2] J. Hasler, S. Kim, and F. Adil, "Scaling floating-gate devices predicting behavior for programmable and configurable circuits and systems," *JPLEA*, vol. 6, no. 3, p. 13, Jul. 2016.

[3] J. Hasler and A. Basu, "Historical perspective and opportunity for computing in memory using floating-gate and resistive non-volatile computing including neuromorphic computing," *Neuromorphic Computing and Engineering*, vol. 5, no. 1, p. 012001, Jan. 2025.

[4] A. Ige and J. Hasler, "Analog system synthesis for fpaas and custom analog ic design," in *DATE*, 2025.

[5] J. Hasler and C. Hao, "Programmable analog system benchmarks leading to efficient analog computation synthesis," *ACM Transactions on Reconfig Technology and Systems*, vol. 17, no. 1, p. 1–25, Jan. 2024.

[6] T.-M. Wang, W.-Y. Chien, C.-L. Hsu, C. J. Lin, and Y.-C. King, "P-channel differential multiple-time programmable memory cells by laterally coupled floating metal gate fin field-effect transistors," *Japanese Journal of Applied Physics*, vol. 57, no. 4S, p. 04FE14, Mar. 2018.

[7] S. Kim, J. Hasler, and S. George, "Integrated floating-gate programming environment for system-level ics," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 6, pp. 2244–2252, 2016.

[8] L. Wang, W. Li, Z. Zhou, H. Gao, Z. Li, W. Ye, H. Hu, J. Liu, J. Yue, J. Yang, Q. Luo, C. Dou, Q. Liu, and M. Liu, "34.9 a flash-sram-adc-fused plastic computing-in-memory macro for learning in neural networks in a standard 14nm finfet process," in *2024 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 67, 2024, pp. 582–584.

[9] W.-Y. Chien, T.-M. Wang, Y.-D. Chih, J. Chang, C. J. Lin, and Y.-C. King, "Twin mode nv logic gates for high speed computing system on 16nm finfet cmos logic process," in *IEEE IEDM*, 2017, pp. 12.1.1–12.1.4.

[10] J. Hasler, P. R. Ayyappan, A. Ige, and P. Mathews, "A 130nm cmos programmable analog standard cell library," *IEEE Transactions on Circuits and Systems I*, vol. 71, no. 6, pp. 2497–2510, 2024.