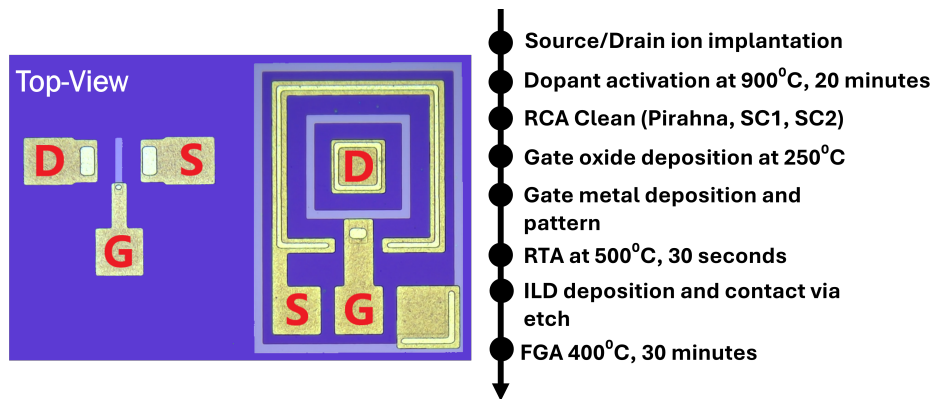
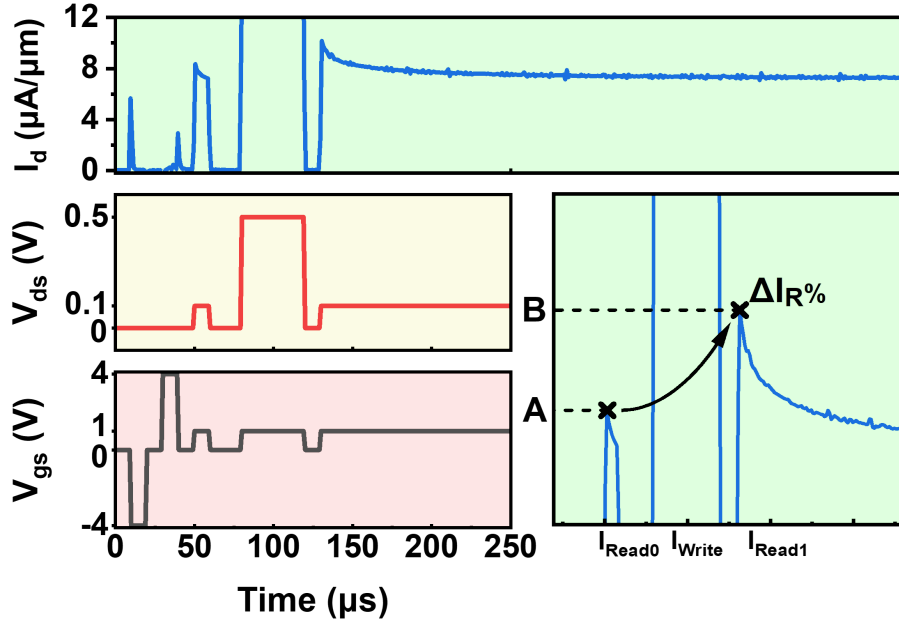


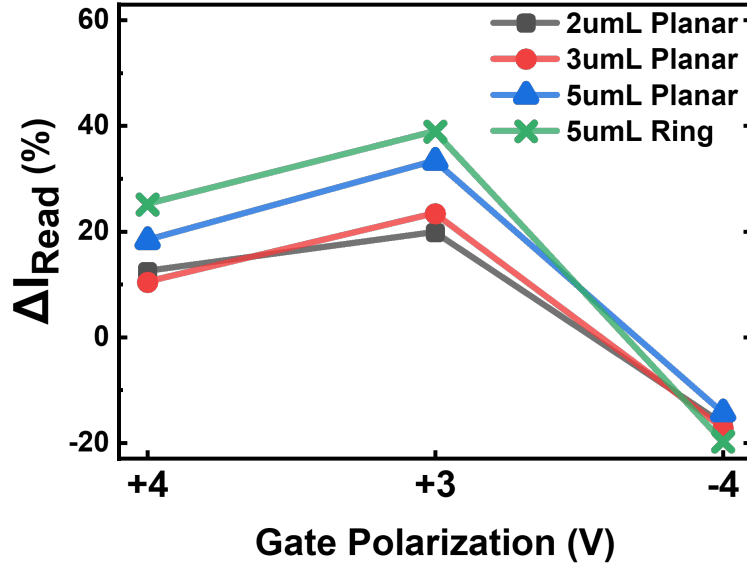
Appendix A Supplementary information



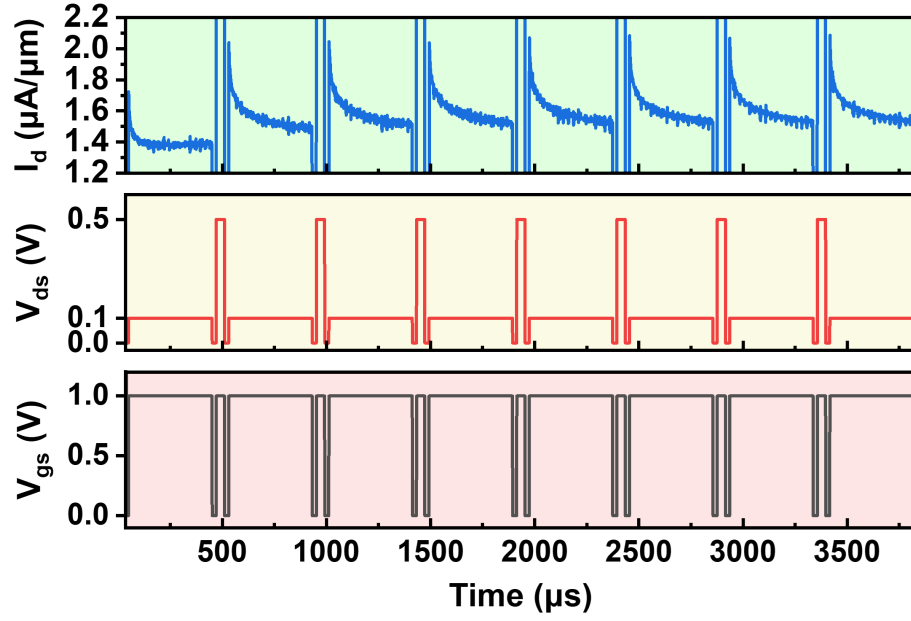
Supplementary Figure 1 | Device top-view and fabrication process. The image on the left side shows the top view of fabricated devices. Devices were made in two different geometries: planar (left) and ring (right) FETs. The flowchart on the right side illustrates the device's fabrication process flow.



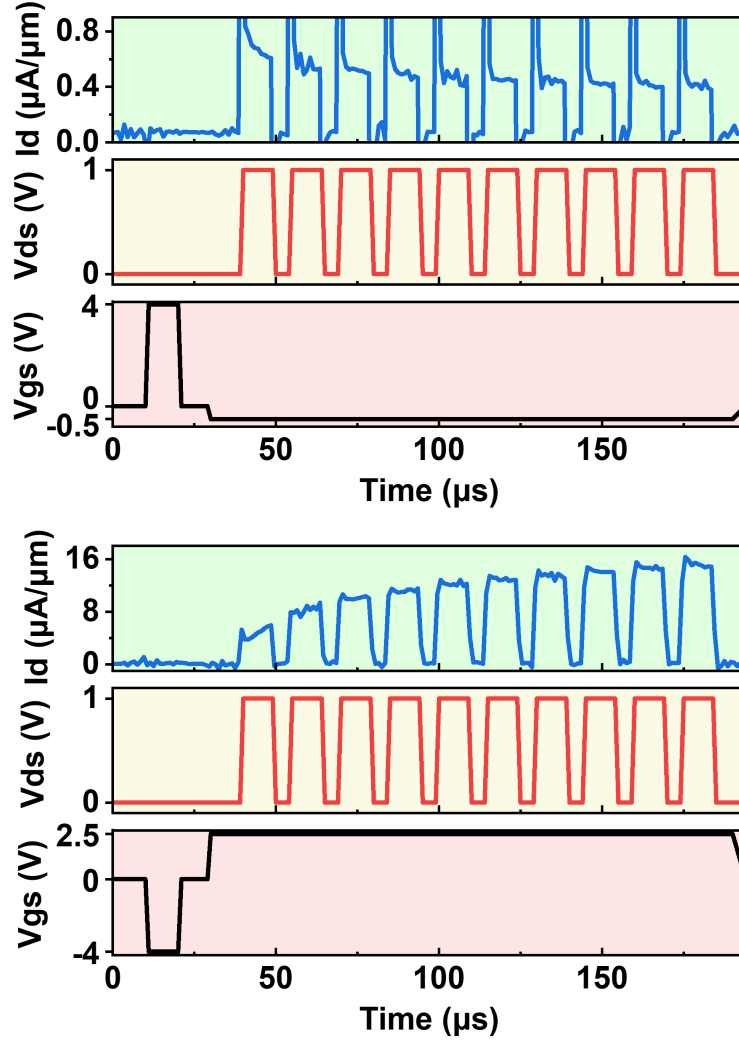
Supplementary Figure 2 | Short-term memory strength measurement. The device is given a -4V, 10 μs gate voltage pulse followed by a specific gate voltage for 10 μs to clear and set the desired gate polarization. Then we apply a series of drain voltage pulses of 0.1 V, 0.5 V, and 0.1 V as the read0, write1, and read1 pulses to test the effect of the drain write pulse. The calculated percentage change between the two read currents (as shown in the lower right sub-figure) is then plotted against different gate polarizations in the next supplementary figure.



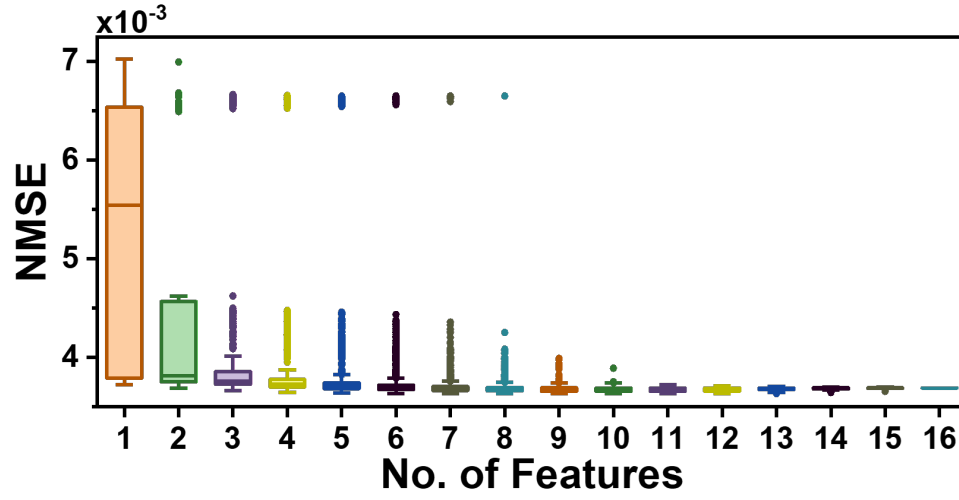
Supplementary Figure 3 | Short-term memory strength affected by long-term polarization. This plot shows the percentage change in drain read current against different gate polarizations. Four lines represent all four geometries of our FeFETs used. The drain current response shows that the device's gate polarization state directly controls the strength and nature of its short-term memory. When the device is polarized at +4 V or +3 V, the second drain read current increases compared to the baseline I_d read current. Polarizing the gate at -4 V and then holding the gate voltage at 2V during the drain voltage pulse causes the second I_d read current to decrease. This figure demonstrates how the device's long-term memory state modulates its short-term memory response.



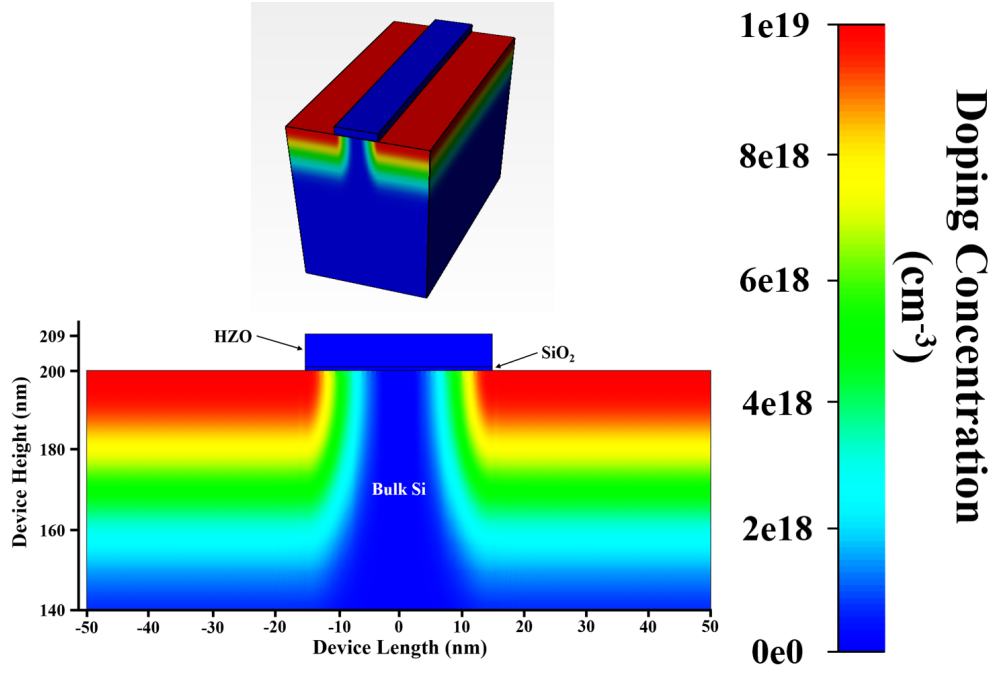
Supplementary Figure 4 | Multiple short-term write and read sequence. The device is gate polarized at +4 V. Then, the drain terminal inputs a baseline read voltage, followed by seven write and read voltage pulse sequences. The drain current response shows some initialization effect from the baseline read to the second read pulse. After that, the I_d write and read current responses are consistent. Based on this, we use the first three voltage pulses (read, write, read) as the pre-writing scheme to set up the device at the start of the reservoir computing electrical measurement.



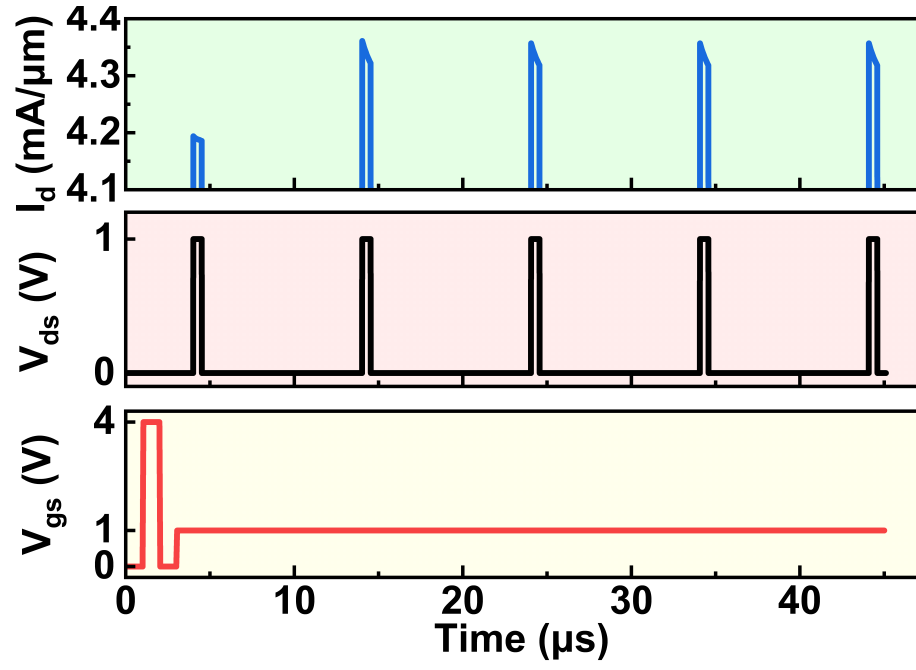
Supplementary Figure 5 | Verification of the non-quasi-static (NQS) origin of the observed short-term memory (STM). In the upper plot, for a positively polarized device, applying a small negative gate bias (-0.5 V) during the drain-pulse train places the channel in weak inversion and yields paired-pulse depression (PPD) in drain current under identical pulses. This is aligned with Figure 3b, where applying a weak inversion of $+2$ V gate bias for a negatively polarized device creates PPD. In the lower plot, for a negatively polarized state, holding V_{gs} at $+2.5$ V during the same V_{ds} pulse sequence drives strong inversion and produces paired-pulse facilitation (PPF). This is aligned with Figure 3a, where applying a strong inversion of $+1$ V gate bias for a positively polarized device creates PPF. These two plots show expected drain current short-term memory consistent with a non-quasi-static (NQS) channel-charge response.



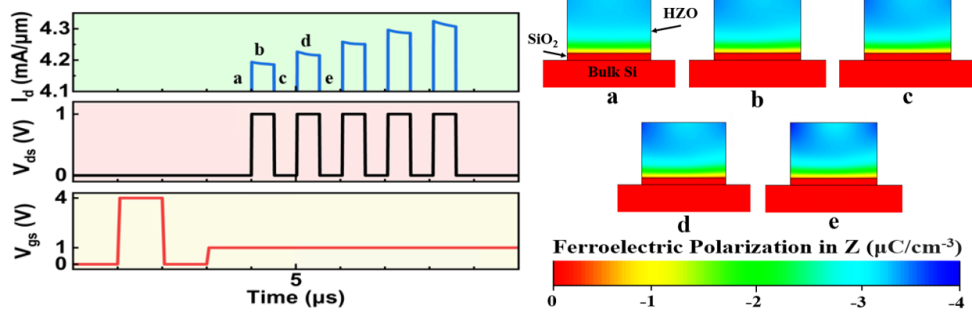
Supplementary Figure 6 | Change of NMSE due to Heterogeneous Reservoir. In our experiment, we have 16 different heterogeneous reservoirs with different gate polarization, gate voltage, voltage encoding, and device geometry. The box plot shows the change of NMSE with different combinations of reservoirs. We can see that after 11 features, all metrics shrank into a small range and are showing a consistent NMSE result.



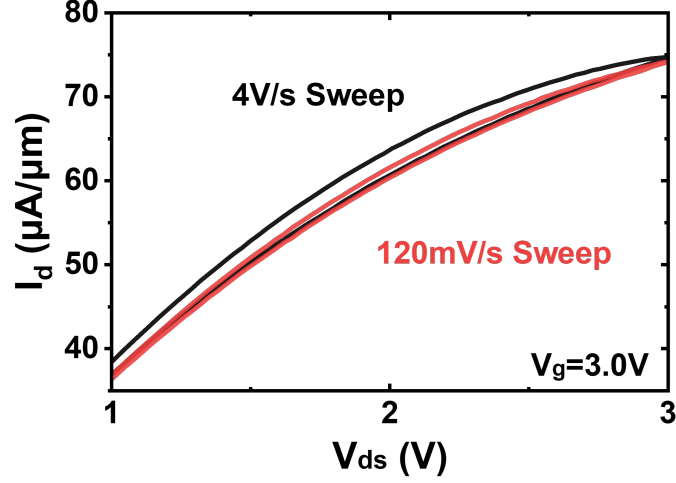
Supplementary Figure 7 | Geometry and doping profile of simulated device visualization. A visualization of the device setup, colormapped by doping concentration. The cross-section of the 3-dimensional view is shown with materials noted. The geometrical and doping parameters are noted in the Methods simulation section.



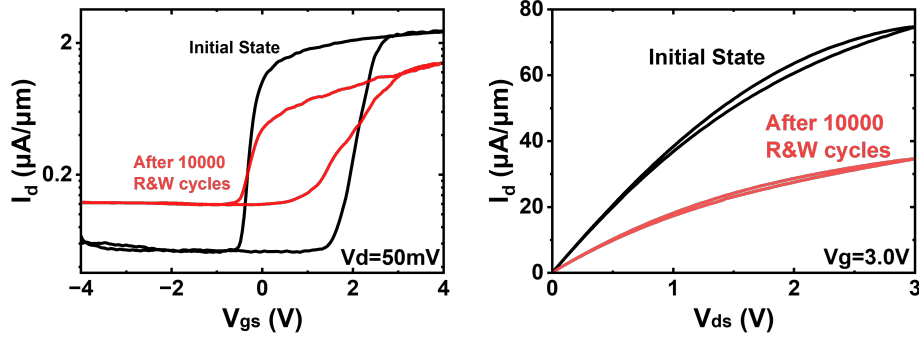
Supplementary Figure 8 | I_d vs time simulation results of long IPI. This simulation uses the same V_{gs} scheme as Fig 6. The drain pulse inter-pulse interval (IPI) changes from $0.5 \mu\text{s}$ to $9.5 \mu\text{s}$ but retains its +1 V amplitude. The resulting I_d exhibits no PPF after the first pulse.



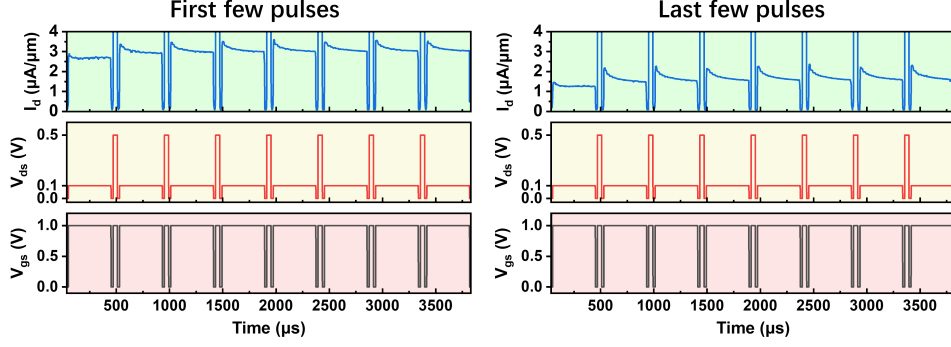
Supplementary Figure 9 | Evolution of ferroelectric polarization in the Z axis in gate polarization at +4 V. The pulsing scheme shown was the input for the GinestraTM simulations. An initial gate voltage pulse from 0 V to +4 V and back to 0 V precedes the final set at +1 V. 5 drain pulses are encoded between 0 V and +1 V with a duty cycle of 0.5 and a pulse width of 1 μ s. The highlighted region of the I_d subsection shows the locations (a-e) viewed using GinestraTM's microscope tool to gain an insight into ferroelectric evolution. Ferroelectric polarization in the Z axis is mapped in a cross-sectional view of the gate region. a) Taken at 3.75 μ s, after the gate setting operation but before the first drain pulse. b) Taken at 4.25 μ s, at the top of the first drain pulse. c) Taken at 4.75 μ s, between the first and second drain pulses. d) Taken at 5.25 μ s, at the top of the second drain pulse. e) Taken at 5.75 μ s, at 0 V after the second drain pulse. The evolution shows minimal change in polarization with time.



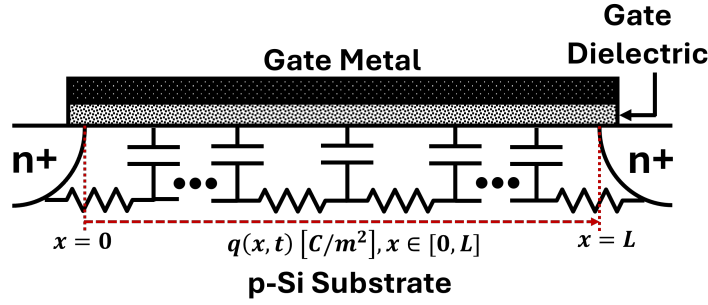
Supplementary Figure 10 | Sweep-rate dependence of FeFET output characteristics. Drain current density I_d as a function of drain voltage V_{ds} measured at a fixed gate bias $V_{gs} = 3.0$ V using two DC sweep rates (fast: 4 V/s, black; slow: 120 mV/s, red). The apparent hysteresis/opening between forward and reverse sweeps is more pronounced at the faster sweep rate, supporting a non-quasi-static (NQS) channel-charge contribution to the measured I_d - V_{ds} characteristics under these conditions.



Supplementary Figure 11 | DC characteristics endurance test comparison. Device underwent 10,000 cycles of read/write cycles as in Supplementary Figure 4. Transfer characteristics I_d - V_{gs} measured at $V_{ds} = 50$ mV (left) and output characteristics I_d - V_{ds} measured at $V_{gs} = 3.0$ V (right), comparing the initial state (black) and the state after 10,000 read/write cycles (red). The post-cycling curves show a clear reduction in absolute drain current, while the overall curve shapes and the ferroelectric hysteresis behavior (LTM signature) remain preserved, indicating functional endurance of the FeFET electrical characteristics under repeated cycling.



Supplementary Figure 12 | STM pulse characteristics endurance test comparison. Device underwent 10,000 cycles of read/write cycles as in Supplementary Figure 4. Drain-current transient response I_d (top) under the same drain-pulse sequence V_{ds} (middle) and gate-bias scheme V_{gs} (bottom), shown for the initial device state and after 10,000 read/write cycles. Although the absolute current level decreases by approximately a factor of two after cycling, the transient response maintains the same qualitative pulse-to-pulse evolution and fading-memory behavior, consistent with the persistence of overlap-capacitance-excited NQS channel-charge dynamics (STM) after endurance stress.



Supplementary Figure 13 | Equivalent RC network representing the Si MOSFET channel. Non-Quasi Static (NQS) model of the transistor is useful for explaining transient behavior, specially in rapid switching mode. The central idea for this model is the hypothesis that any transient voltage change at any terminal of a transistor do not result in an instantaneous change in the channel charge ($q(x, t)$) that gives rise to a transient component of the drain current. If the subsequent input pulse comes within the decay time of the transient current, the total current will change momentarily giving rise to short-term memory described in this paper.

Supplementary Note: Mathematical formalism of the NQS effect

Scope.

This note summarizes a compact derivation of the non-quasi-static (NQS) response in MOSFETs under drain-voltage pulsing when a designed gate-source/drain overlap capacitance provides the input coupling. We follow the classical continuity-based transient treatment and its compact reduction to the “charge-deficit” model used in BSIM3/BSIM4, emphasizing quantities that appear in our experiments and simulations. [32, 37–39]

S1. Charge-based baseline and continuity

Let $q(x, t)$ [C/m²] denote the inversion charge density along the channel ($x \in [0, L]$), referenced to the operating point. Conservation of charge gives

$$\frac{\partial q(x, t)}{\partial t} + \frac{\partial i(x, t)}{\partial x} = 0, \quad (\text{A1})$$

where $i(x, t)$ is the lateral channel current density (A/m). Linearizing the constitutive relation about the bias point yields

$$i(x, t) \approx \sigma_{\text{ch}} \frac{\partial \psi(x, t)}{\partial x}, \quad q(x, t) \approx C'_{\text{ch}} \psi(x, t), \quad (\text{A2})$$

with sheet conductivity σ_{ch} and channel charge susceptibility (capacitance per unit area) C'_{ch} that are bias dependent. Eliminating ψ gives a diffusion-type PDE for the incremental charge:

$$\frac{\partial q(x, t)}{\partial t} = D_{\text{ch}} \frac{\partial^2 q(x, t)}{\partial x^2}, \quad D_{\text{ch}} \equiv \frac{\sigma_{\text{ch}}}{C'_{\text{ch}}}. \quad (\text{A3})$$

Equation (A3) is the continuum origin of NQS: the channel requires a finite time to redistribute charge after a fast perturbation. [32]

S2. Boundary excitation via overlap capacitance

A drain step $V_D(t) = V_{D0} + \Delta V_D u(t)$ couples into the gate/channel edge near $x = L$ through the gate-drain overlap capacitance $C_{gd,ov}$. The injected displacement current is

$$i_{\text{ov}}(t) = C_{gd,ov} \frac{dV_D}{dt}, \quad (\text{A4})$$

which imposes a boundary condition at $x = L$ equivalent to a localized charge injection/extraction that sets the initial condition $q(x, 0^+)$ for Eq. (A3). For compactness, we collect the distributed dynamics into an internal state $Q_{\text{def}}(t)$ (the charge-deficit variable).

S3. Modal solution and single-pole reduction

Solving Eq. (A3) with standard mixed boundary conditions gives a modal expansion

$$q(x, t) = \sum_{n \geq 1} A_n \phi_n(x) e^{-t/\tau_n}, \quad \tau_n \sim \frac{L^2}{(\pi n)^2 D_{\text{ch}}}. \quad (\text{A5})$$

The lowest pole τ_1 dominates most device-level transients. Compact models therefore replace the distributed channel by a single-pole surrogate with an internal state:

$$\frac{dQ_{\text{def}}}{dt} = -\frac{Q_{\text{def}}}{\tau} + S(t), \quad I_k(t) = I_k^{\text{DC}} + p_k \frac{dQ_{\text{ch}}}{dt}, \quad (\text{A6})$$

where $S(t)$ encodes the excitation (e.g., Eq. (A4)), Q_{ch} is the total channel charge, and p_k are charge-conserving partition factors for terminals $k \in \{G, D, S\}$. [32, 37, 39]

S4. The NQS time constant and its bias dependence

In BSIM-style charge-deficit formulations, the dominant NQS time constant is written

$$\tau = R_{ii} W_{\text{eff}} L_{\text{eff}} C_{\text{ox,eff}}, \quad (\text{A7})$$

with R_{ii} the intrinsic input resistance that aggregates drift and diffusion contributions; $C_{\text{ox,eff}}$ is the effective oxide capacitance per unit area. [37, 38] Because the diffusion component of R_{ii} increases near threshold,

$\tau(V_G)$ is larger in weak inversion (near/subthreshold) and smaller in strong inversion,

which is the established NQS bias dependence. [37, 38]

S5. Small-signal NQS (frequency domain)

Linearizing Eq. (A6) gives first-order frequency responses; to leading order,

$$C_{gd}(\omega) \approx \frac{C_{gd,0}}{1 + j\omega\tau}, \quad g_m(\omega) \approx \frac{g_{m,0}}{1 + j\omega\tau_g}, \quad (\text{A8})$$

capturing the roll-off of the effective drain-gate coupling and transadmittance with $j\omega\tau$. [37, 39] These forms are sufficient to fit the observed drain-step transients and AC corners.

S6. Drain-step transient and paired-pulse metrics

For a single drain step, the terminal currents contain (i) a prompt displacement spike due to $C_{gd,\text{ov}}$ and C_{db} , and (ii) a conduction relaxation with time constant τ set by Eq. (A7). For a paired-pulse protocol with inter-pulse interval T (IPI), the second pulse samples the residual channel-charge perturbation left by the first. A

convenient scalar metric is the paired-pulse facilitation/depression (PPF/PPD) ratio, which—under broad conditions—follows

$$\text{PPF/PPD}(T) \propto e^{-T/\tau}, \quad (\text{A9})$$

where the sign depends on the operating point and excitation polarity (see S7). Equation (A9) underpins our one-parameter fits of PPF vs. IPI.

S7. Operating-regime dependence (strong vs. weak inversion)

Under a fixed external V_G , ferroelectric polarization selects the local regime at the drain edge. Positive polarization lowers the effective threshold $\Rightarrow$ strong inversion (smaller R_{ii} , shorter τ); negative polarization raises it $\Rightarrow$ weak inversion (larger R_{ii} , longer τ). [37, 38] In our biasing, the strong-inversion case exhibits paired-pulse facilitation, while the weak-inversion case exhibits depression. We interpret this polarity as consistent with NQS dynamics: the sign and magnitude of the residual channel-charge perturbation near the drain at time T (just before the second pulse) depend on regime and partitioning; the shorter (longer) τ in strong (weak) inversion yields distinct residuals and hence PPF/PPD. [32, 39] We emphasize that the polarity itself is operating-point specific rather than universal.

S8. Practical extraction recipe (time domain)

From a measured or simulated PPF-vs-IPI dataset $\{T_i, \text{PPF}(T_i)\}$:

1. Subtract any baseline offset and fit $\text{PPF}(T)$ to $A e^{-T/\tau}$ to obtain τ and its confidence interval.
2. Repeat for two polarization states (strong vs. weak inversion); τ should increase as the operating point moves toward weak inversion.
3. Optionally, vary V_G in small steps around threshold to map $\tau(V_G)$ and corroborate the NQS bias trend.

S9. Summary

The NQS formalism follows directly from charge conservation and a distributed RC description of the inversion channel. Its compact reduction yields a single-pole time constant τ [Eq. (A7)] that is predictably bias dependent. In our overlap-engineered devices, drain-pulse experiments probe this τ via paired-pulse transients; the observed PPF/PPD and their timescales are consistent with the overlap-driven NQS mechanism. [32, 37–39]