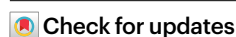


Monolithic 3D integration of atomic-layer-deposited oxide semiconductors on 200-mm silicon wafers

Received: 11 February 2026

Accepted: 12 August 2026

Published online: 15 September 2026



Chang Niu^{1,2,4}, Linjia Long^{1,2,4}, Luqi Zheng^{1,2}, Shuting Du^{1,2}, Jian-Yu Lin^{1,2}, Kisoo Nam^{1,2}, Zehao Lin^{1,2}, Chang Liu³, Juanjuan Lu³, Haiyan Wang³, Haitong Li^{1,2} & Peide D. Ye^{1,2}✉

Monolithic 3D (M3D) integration offers a pathway to overcome the scaling limits of conventional silicon complementary metal–oxide–semiconductor technology by extending dense vertical stacking of multifunctional logic and memory devices. Here we demonstrate wafer-scale M3D integration of three tiers of atomic-layer-deposited indium oxide (InO_x)-based devices (>100,000 fabricated), including ferroelectric, enhancement-mode and depletion-mode field-effect transistors, on 200-mm silicon wafers. We achieve threshold voltage standard deviations as low as 0.04 V, average electron mobilities of up to $91.6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and fully functional cross-tier circuits. A four-tier 3D computing-in-memory accelerator targeting large-language-model workloads is developed using a custom InO_x process design kit, delivering $1.4\times$ to $2.9\times$ speed-up and comparable energy–delay product improvements over 2D baselines. These results establish atomic-layer-deposited InO_x M3D integration as a scalable and complementary metal–oxide–semiconductor-compatible platform for next-generation artificial intelligence hardware and advanced electronics.

Heterogeneous integration of novel materials and devices is crucial for sustaining transistor scaling in the modern semiconductor industry and for advancing the complementary metal–oxide–semiconductor + X paradigm with expanded functionalities^{1–5}. This integration can mostly be achieved through wafer bonding or other transfer techniques; however, these methods are often limited by feature size, cost and success rate. An alternative approach is monolithic 3D (M3D) integration^{6–10}, which requires back-end-of-line (BEOL)-compatible processes^{11–15}. To preserve the functionality of underlying layers, the thermal budget must remain below 400 °C (refs. 16–19). This poses a challenge because the direct growth or monolithic integration of many novel

materials—such as III–V compounds or germanium—typically requires high-temperature epitaxial processes. Atomic-layer-deposited (ALD) amorphous oxide semiconductors^{20–24} offer a promising solution for M3D integration. Nanometre-thin indium oxide (InO_x) layers can be uniformly and conformally deposited on various substrates and complex structures using a low-temperature process ($\sim 225^\circ\text{C}$), making them fully BEOL-compatible. ALD InO_x -based thin-film transistors (TFTs) show excellent performance, including ultrahigh on-current²⁵, high bias stability²⁶, low off-current²⁷ and ultralow contact resistance^{28,29}.

In addition to their potential for logic applications, integrating ferroelectric materials with InO_x enables ferroelectric field-effect

¹Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. ²Birk Nanotechnology Center, Purdue University, West Lafayette, IN, USA. ³School of Materials Science and Engineering, Purdue University, West Lafayette, IN, USA. ⁴These authors contributed equally: Chang Niu, Linjia Long. ✉e-mail: yep@purdue.edu

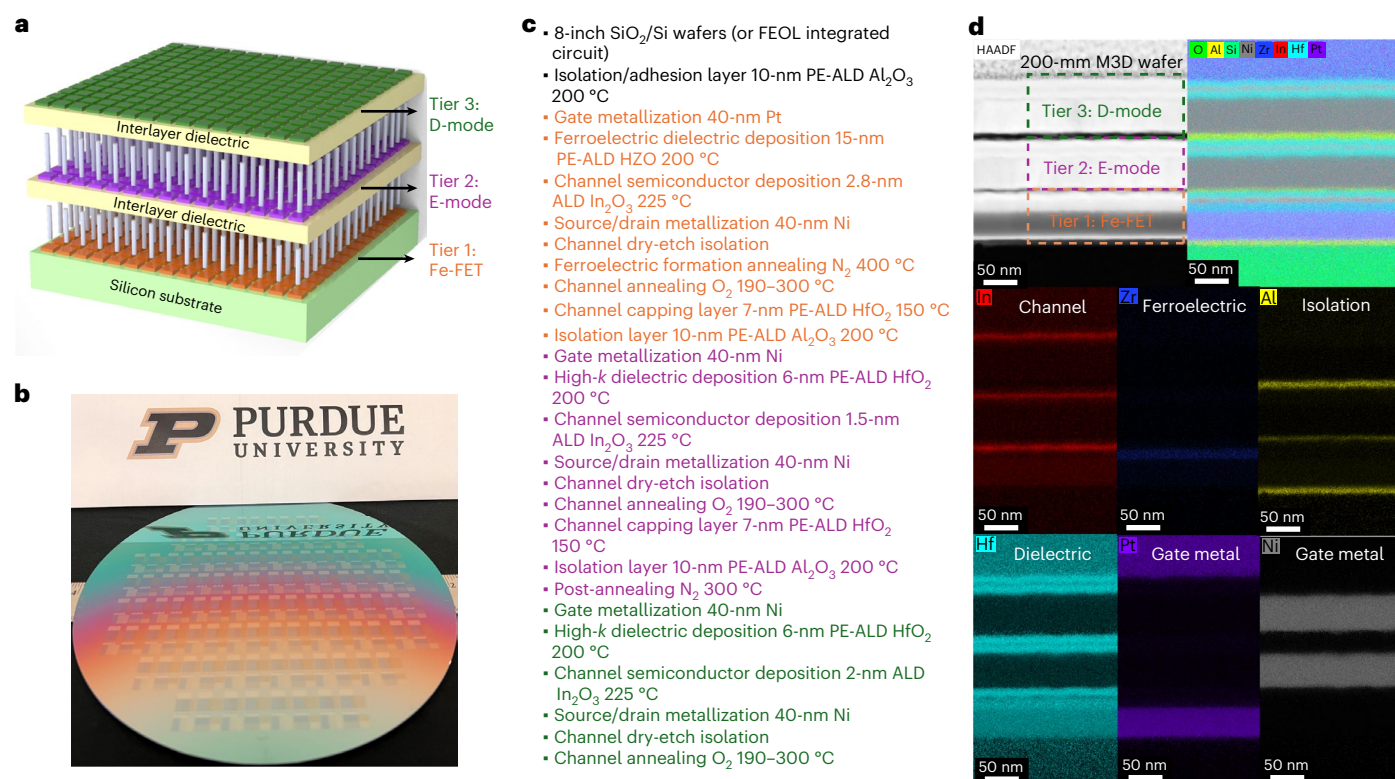


Fig. 1 | Multi-tier ALD InO_x integration on 200-mm wafers. a, Schematic illustration of the M3D integration of ALD InO_x -based devices, including ferroelectric, E-mode and D-mode FETs across three tiers. **b**, Optical image of a 200-mm silicon wafer with integrated multi-tier ALD InO_x devices. **c**, Fabrication process flow for the wafer-scale M3D integration, showing the sequential

stacking of three transistor tiers using BEOL-compatible processes. **d**, HAADF-STEM image and corresponding EDS elemental maps of the device cross-section, confirming the vertically stacked architecture and spatial distribution of the integrated materials.

transistors (Fe-FETs) to exhibit promising characteristics such as a wide memory window (MW), strong retention and high endurance³⁰. With the rising demand for computational power, emerging paradigms such as computing-in-memory and neuromorphic computing require compact integration of multiple functional transistors. M3D integration of vertically stacked, multi-tier architectures combining non-volatile memory technologies such as Fe-FETs with logic devices offers a highly promising route toward energy-efficient, high-performance computing systems. Furthermore, scaling up to 200-mm or 300-mm silicon wafers represents a critical milestone in evaluating this ALD oxide semiconductor technique as a viable option for future manufacturing, alongside other novel channel materials^{16,31–33}. For large-scale production, it is essential to evaluate device performance, yield and uniformity. Statistical analysis of wafer-scale transistors is crucial for assessing both material and process quality, bringing ALD oxide semiconductors one step closer to real-world applications.

In this work, we fabricated and systematically characterized 200-mm wafer-scale, monolithically integrated, multi-tier ALD InO_x -based Fe-FETs and FETs, achieving high uniformity, yield and device performance. Statistical analysis was conducted across a three-tier stack comprising Fe-FETs, enhancement-mode (E-mode) transistors and depletion-mode (D-mode) transistors. Furthermore, by interconnecting multifunctional devices across different tiers, we demonstrate multilevel ferroelectric non-volatile memory (MLFe-NVM), inverters and embedded dynamic random-access memory (eDRAM). This oxide-semiconductor-based M3D platform is further validated through a computing-in-memory (CIM) accelerator tailored for large language models (LLMs), demonstrating substantial gains in performance and energy efficiency. These results highlight ALD InO_x as a scalable and versatile

foundation for next-generation three-dimensional (3D) integrated semiconductor systems.

ALD oxide semiconductors on 200-mm wafers

The structure of M3D integrated multi-tier devices based on ALD InO_x is shown schematically in Fig. 1. By leveraging the vertical (out-of-plane) dimension, this architecture enables further semiconductor scaling and heterogeneous integration of diverse functionalities. A 200-mm silicon wafer incorporating three vertically stacked tiers of ALD InO_x devices was fabricated, comprising over 100,000 devices across 52 dies. The BEOL-compatible fabrication process, detailed in Methods, was used to sequentially integrate tier 1 Fe-FETs, tier 2 E-mode FETs and tier 3 D-mode FETs. Ferroelectric HfZrO_2 (HZO) was used as gate dielectric for Fe-FETs at tier 1. High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) imaging and energy-dispersive X-ray spectroscopy (EDS) elemental mappings at the gate metal region (Fig. 1d) clearly show the vertically stacked device layers, including three InO_x channels, ferroelectric HZO, Al_2O_3 isolation, and HfO_2 dielectric and capping layers. A line-scan along the out-of-plane direction (Supplementary Fig. 1) confirms the nanometre-scale uniformity of the individual layers and the presence of sharp, well-defined interfaces.

The transfer and output characteristics of short-channel ($L_{\text{ch}} = 100$ nm) and long-channel ($L_{\text{ch}} = 1$ μm) ALD InO_x FETs are shown in Fig. 2. The devices exhibit excellent transistor behaviour, including negligible hysteresis, an on/off current ratio exceeding 10 orders of magnitude, a low subthreshold swing ($\text{SS} = 70$ mV per decade) and good saturation in long-channel devices. Short-channel devices demonstrate high drive current, reaching $1,000 \mu\text{A } \mu\text{m}^{-1}$ at $L_{\text{ch}} = 100$ nm. The channel length scaling behaviour (Fig. 2c,d) indicates consistent on-conductance and minor threshold voltage

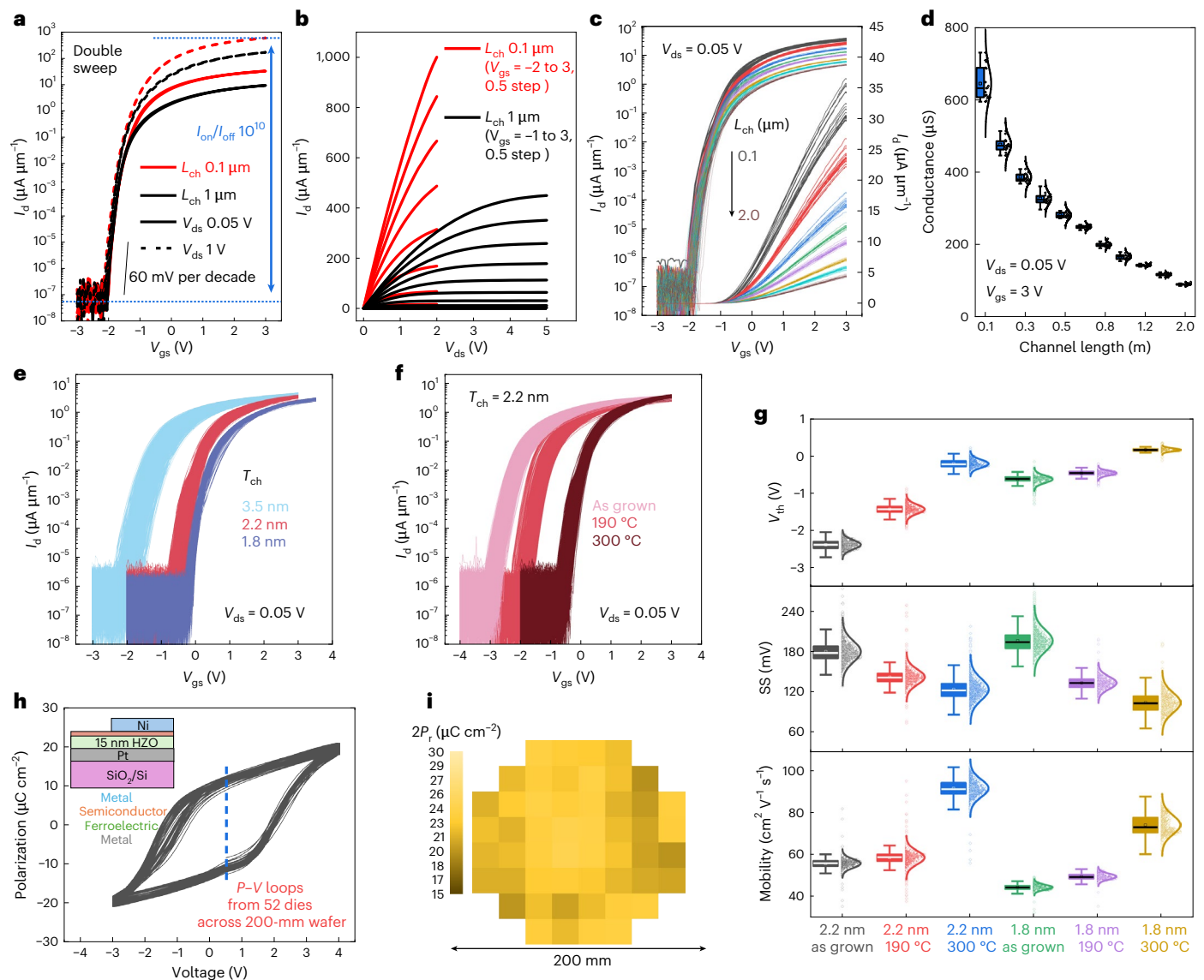


Fig. 2 | Process-dependent electrical performance of ALD InO_x on 200-mm wafers. **a, b, Transfer (**a**) and output (**b**) characteristics of InO_x FETs, showing high on-current ($1,000 \mu A \mu m^{-1}$) in short-channel devices and excellent saturation behaviour in long-channel devices. **c, d**, Channel length dependence of the transfer characteristics (**c**) and channel conductance (**d**), demonstrating the scalability of InO_x FETs. Sample sizes, $n = 19, 19, 21, 20, 19, 20, 18, 16, 20, 20$ and 18 , respectively. **e**, Statistical analysis of device characteristics for varying InO_x thicknesses across 200-mm wafers. **f**, Statistical analysis of device characteristics**

under different annealing temperatures. **g**, Extracted device parameters (threshold voltage, subthreshold swing and mobility) from 200-mm wafers. Sample sizes, $n = 1,601, 1,606, 1,596, 1,601, 1,600$ and 753 , respectively. **h**, P - V loop of metal-semiconductor-ferroelectric-metal capacitors on 200-mm wafers. **i**, Colour map of remanent polarization across a 200-mm wafer, confirming spatial uniformity. In **d, g** boxes are 25th–75th percentiles, centre lines are medians, whiskers are $1.5 \times$ interquartile range, symbols are means and the curves show Gaussian distributions.

(V_{th}) roll-off, with small device variation. To comprehensively evaluate device performance and uniformity across wafers, five 200-mm InO_x test wafers were fabricated, each incorporating variations in film thickness and gate dielectric configurations (as shown in Supplementary Fig. 2). A statistical analysis of over 700 devices per condition at $L_{ch} = 5 \mu m$ across 200-mm wafers is presented in Fig. 2e, highlighting the threshold voltage dependence on InO_x film thickness due to surface electron accumulation. Thicker films exhibit higher V_{th} variability (standard deviation $\sigma_{V_{th}} = 0.18 V$ for oxide channel thickness (T_{ch}) = $3.5 nm$), while thinner films improve uniformity ($\sigma_{V_{th}} = 0.04 V$ for $T_{ch} = 1.8 nm$), attributed to reduced thickness fluctuation across the wafer (Supplementary Fig. 3). Thermal treatment provides an effective tuning knob: oxygen annealing is shown to enhance device performance and shift V_{th} (Fig. 2f). Extracted parameters (V_{th} , SS and mobility) across different InO_x thicknesses and annealing conditions

are summarized in Fig. 2g. Devices annealed at 300 °C demonstrate improved performance, achieving an average field-effect mobility of $91.6 cm^2 V^{-1} s^{-1}$. The 200-mm wafer-scale ALD InO_x FETs exhibit excellent uniformity, as evidenced by the consistent output characteristics of 137 devices shown in Supplementary Fig. 4, demonstrating their suitability for large-scale fabrication in semiconductor technology. Building upon these optimizations, wafer-scale ALD InO_x Fe-FETs were fabricated by incorporating ferroelectric HZO directly with InO_x . The high thermal stability and appropriate interfacial strain provided by InO_x facilitate ferroelectric crystallization in HZO, enabling interfacial-layer-free Fe-FETs. Polarization–voltage (P - V) characterizations of the Fe-FET gate stacks, together with a remanent polarization (P_r) map extracted from metal-ferroelectric-semiconductor-metal (Pt-HZO- InO_x -Ni) capacitors across 52 dies, confirm uniform ferroelectric performance across the 200-mm wafer (Fig. 2h,i).

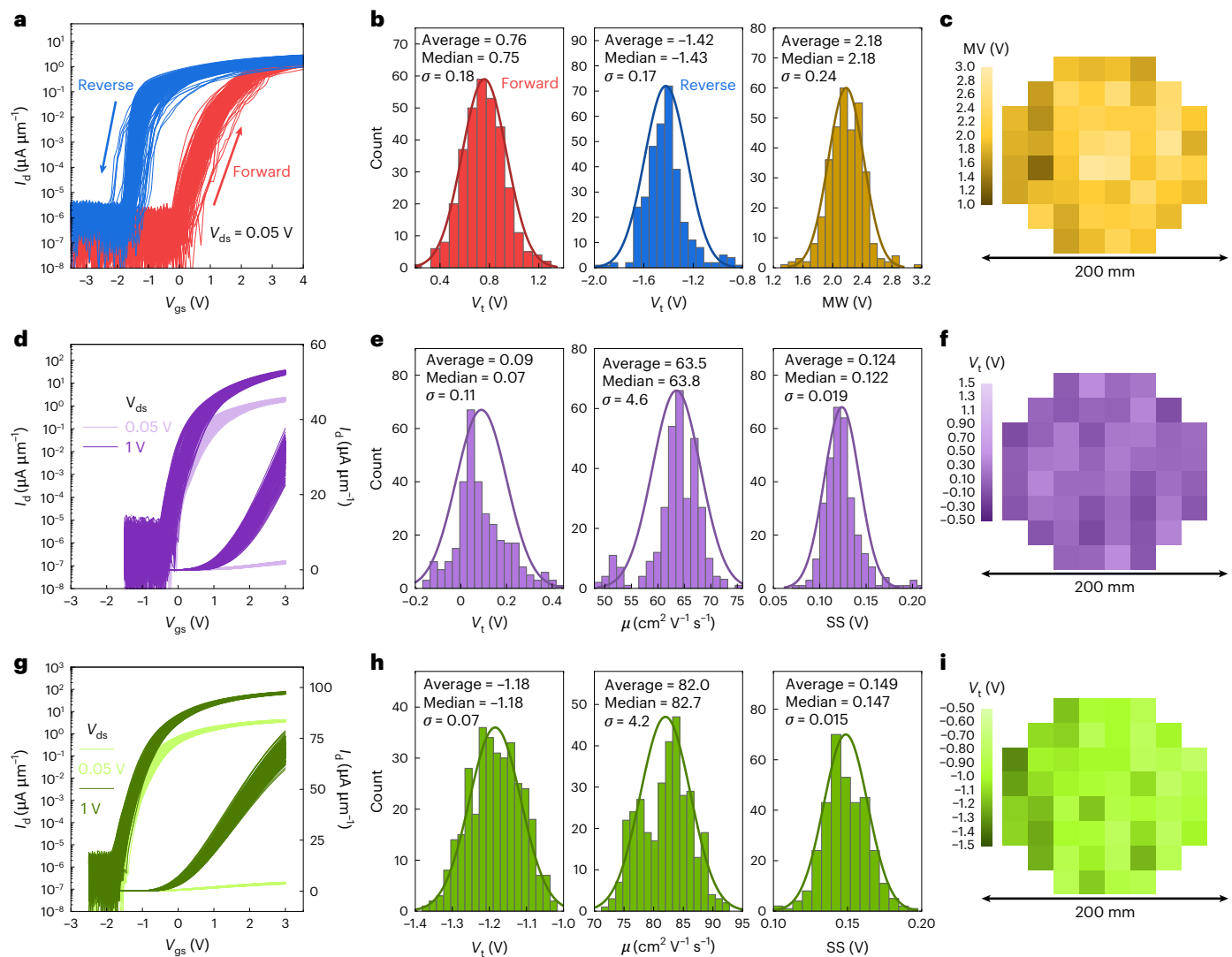


Fig. 3 | Device characterizations of multi-tier ALD InO_x transistors on a 200-mm wafer. **a–c**, Electrical performance of tier 1 Fe-FETs: transfer characteristics (**a**), statistical distributions of extracted parameters (**b**) and special colour map of threshold voltages across the wafer (**c**). **d–f**, Electrical performance of tier

2 E-mode FETs: transfer characteristics (**d**), extracted device statistics (**e**) and threshold voltage mapping across the wafer (**f**). **g–i**, Electrical performance of tier 3 D-mode FETs: transfer characteristics (**g**), extracted device statistics (**h**) and threshold voltage mapping across the wafer (**i**).

With excellent device performance and low variability, along with the ability to tune threshold voltage and enhance device stability through annealing, film thickness control and surface capping, we are now positioned to implement vertical multi-tier integration of ALD InO_x-based M3D devices on 200-mm wafers. As shown in Fig. 3, The M3D integration begins with Fe-FETs in tier 1. The transfer characterization of over 300 Fe-FETs measured on the 200-mm multi-tier wafer under forward and reverse sweeps reveals a large MW and consistent performance. The devices exhibit a clear distinction between memory states ‘0’ and ‘1’, with no overlap in their distributions. Histograms of the threshold voltages for forward sweeps (average $V_{th} = 0.76$ V), reverse sweeps (average $V_{th} = -1.42$ V) and the MW (average MW = 2.18 V, standard deviation $\sigma = 0.24$ V) demonstrate tight distributions. The colour mapping of the MW extracted from 52 dies across the 200-mm wafer confirms excellent wafer-scale uniformity. The second-tier ALD InO_x E-mode FETs were fabricated directly on top of the tier 1 Fe-FETs using BEOL-compatible processes, separated by a 17-nm isolation dielectric layer. The transfer characteristics of over 300 devices measured at different drain voltages ($V_{ds} = 0.05$ V and 1 V), presented in both linear and logarithmic scales, are shown in Fig. 3d. The histograms in Fig. 3e reveal a tight threshold voltage distribution with minimal variation (average

$V_{th} = 0.09$ V, standard deviation $\sigma = 0.11$ V), high electron mobility (average $\mu = 63.5$ cm² V⁻¹ s⁻¹) and low SS (average SS = 0.124 V per decade). A V_{th} map across the 200-mm wafer confirms the excellent yield and uniform electrical characteristics of the tier 2 E-mode FETs. The third-tier ALD InO_x D-mode FETs were fabricated using a similar BEOL-compatible process, with a thicker InO_x layer to shift the threshold voltage. Electrical characterization and the corresponding statistical distributions of the tier 3 D-mode FETs across the 200-mm wafer are presented in Fig. 3g,i. The increased film thickness resulted in enhanced performance, including a higher electron mobility of 82.0 cm² V⁻¹ s⁻¹. Compared to tier 2 devices, the D-mode FETs also exhibited reduced variation, with an average $V_{th} = -1.18$ V and a standard deviation $\sigma = 0.07$ V. The devices fabricated on the top tiers exhibit no performance degradation compared with those on the lower tiers, indicating a robust and reliable integration process. Notably, the vertical stacking of multi-tier devices is not limited to three tiers; additional layers can be integrated through repeated processing steps. The thermal stability of the InO_x BEOL material platform is further supported by the non-degraded electrical performance of Fe-FETs, E-mode FETs, D-mode FETs and integrated circuits during capping and postfabrication annealing, including preserved/improved transfer characteristics, SS, V_t , μ and bias stability as shown in Extended

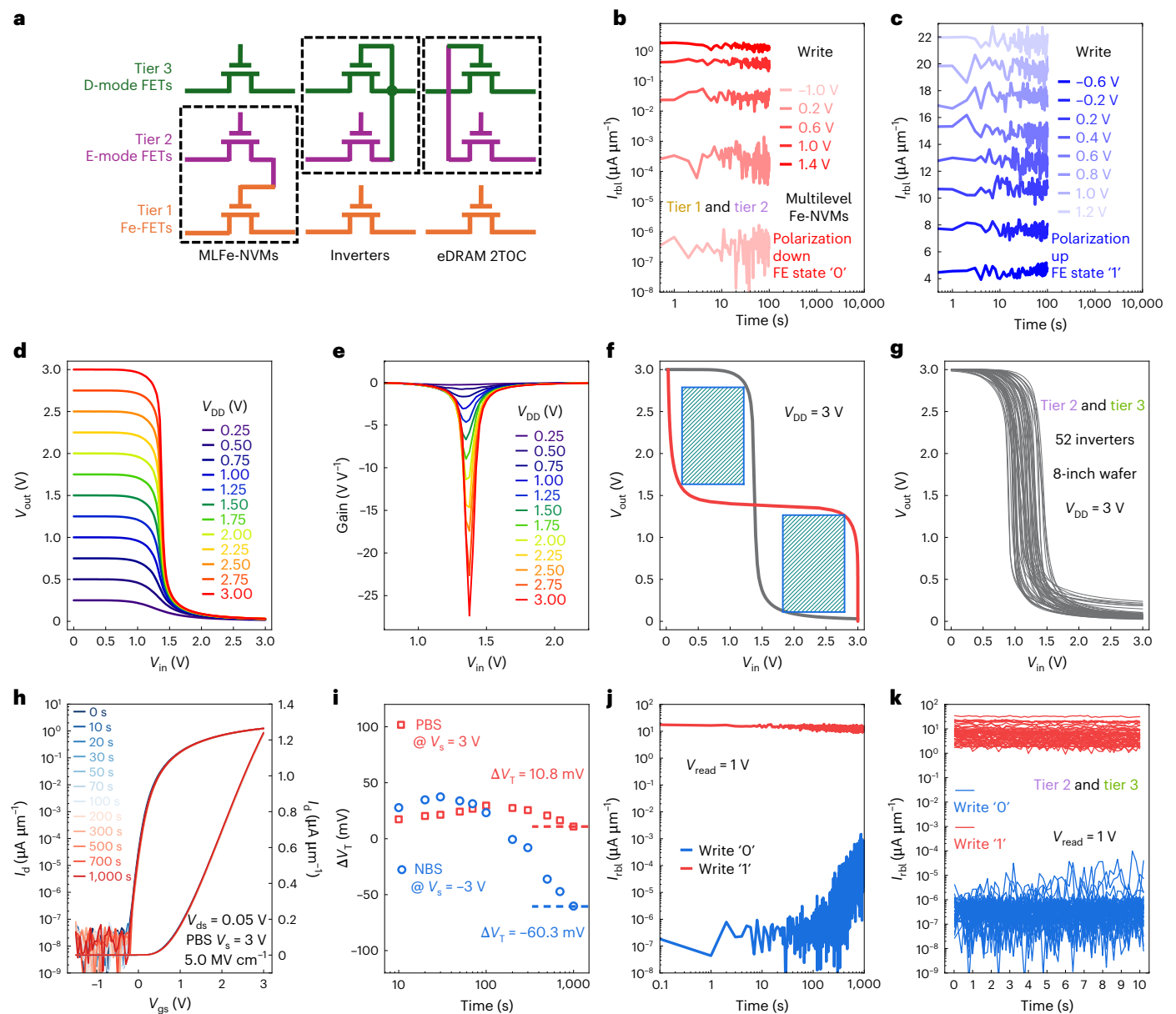


Fig. 4 | Cross-tier multifunctional vertically integrated logic and memory on a 200-mm wafer. **a**, Schematic illustration of cross-tier logic and memory integration, including Fe-NVMs, inverters and 2TOC eDRAM. **b, c**, Multilevel operation of a Fe-NVM cell under opposite ferroelectric polarization states (up (c) and down (b)). **d–g**, Electrical performance of cross-tier inverters: voltage transfer characteristics (d), voltage gain (e), butterfly curve showing the noise

margins as shaded regions (f) and uniform transfer characteristics of 52 inverters across the 200-mm wafer (g). **h–k**, Device performance of cross-tier 2TOC eDRAM cells: bias stability under positive (h) and negative (i) stress conditions, retention characteristics showing large on/off ratio (j) and consistent readout current levels across 52 cells measured over the 200-mm wafer (k).

Data Figs. 1–4 and Supplementary Figs. 5 and 6. In this work, we successfully demonstrated three-tier M3D integration on 200-mm wafers, confirming that lower-tier devices remain fully functional after the fabrication of upper tiers. This validates the scalability of the approach and its potential for further vertical expansion.

Benchmarking against recent BEOL-compatible channel platforms^{34–54} (Extended Data Table 1 and Extended Data Fig. 5) shows that our InO_x technology occupies a particularly favourable device-level operating regime. In addition to high mobility, the platform combines wafer-scale process compatibility, low-temperature integration and suitability for continued multi-tier stacking, distinguishing it from alternative oxide and 2D material approaches. These results indicate that InO_x provides not only competitive transistor performance, but also a more practical foundation for dense monolithic 3D integration.

Cross-tier integrated circuits and memory

Compared to traditional planar circuit architectures, M3D vertically stacked multi-tier logic and memory structures offer significant advantages in area efficiency and design flexibility, enabling integration of multiple functionalities within a compact footprint⁵. In our work, distinct device types are vertically integrated across three tiers: Fe-FETs in tier 1, E-mode FETs in tier 2 and D-mode FETs in tier 3, as illustrated in Fig. 4. This multi-tier configuration enables the realization of multifunctional logic and memory architecture. Specifically, E-mode FETs in tier 2 are used as access transistors for the tier 1 Fe-FETs to construct MLFe-NVM. As shown in Supplementary Fig. 7, multilevel operation is achieved by storing different voltages at the storage node (V_{SN}), facilitated by the ultralow leakage current of the wide-bandgap oxide semiconductors and the stable non-volatile polarization states of the

Fe-FETs. This enables the introduction of an additional dimension of data storage. Distinct readout current levels (I_{rd}) corresponding to different storage voltages, modulated by the applied write voltage at different ferroelectric polarization states (down and up), are demonstrated in Fig. 4b,c. These distinct current levels allow for expanded multilevel memory operation without overlapping between logic states. We attribute the retention fluctuation mainly to non-ideal parasitic capacitance and layout-dependent coupling in the current proof-of-concept routing scheme, rather than to an intrinsic limitation of the cross-tier structure. By combining n levels of voltage storage at the node and m polarization-induced states in Fe-FET, the MLFe-NVM architecture supports up to $n \times m$ storage levels. This multidimensional storage approach supports high-density, low-power memory with integrated functionality in advanced 3D electronic systems.

Furthermore, cross-tier inverters are demonstrated by integrating tier 2 E-mode FETs with tier 3 D-mode FETs. The voltage transfer characteristics and voltage gain of the zero-gate-to-source-voltage (V_{gs})-load cross-tier inverters are shown in Fig. 4d–g. These inverters exhibit full rail-to-rail swing (from supply voltage (V_{DD}) to 0) and a high voltage gain of 23 V V^{-1} at $V_{\text{DD}} = 3 \text{ V}$. The butterfly curve reveals a large noise margin of 65%, confirming robust inverter operation. The voltage transfer characteristics of 52 cross-tier inverters measured across 200-mm wafers demonstrate excellent electrical uniformity. The distributions of voltage gain and switching threshold (midpoint voltage) show low variation, with a standard deviation $\sigma = 0.15 \text{ V}$, as detailed in Supplementary Fig. 8. Using the same tier 2 and tier 3 devices, cross-tier two-transistor zero-capacitor (2TOC) eDRAM cells are implemented, in which the storage node (SN) is formed by connecting the tier 3 gate and the tier 2 drain. The key enablers for high-performance 2TOC eDRAMs are the ultralow off-state current of the access transistor and high bias stability under both positive and negative bias stress (PBS and NBS). The PBS results of the ALD InO_x multi-tier FETs exhibit measurement-limited off-currents and negligible V_{th} shift under stress ($V_{\text{S}} = 3 \text{ V}$, equivalent to 5.0 MV cm^{-1} for 1,000 s), even under challenging high-field conditions for high- k dielectrics. The extracted V_{th} shifts for PBS and NBS are only 10.8 mV and -60.3 mV , respectively, as plotted in Fig. 4h,i, demonstrating exceptional bias stability. Read currents corresponding to the '0' and '1' memory states at a read voltage of 1 V are shown for the 2TOC eDRAM in Fig. 4j,k. A large memory window with on/off current ratio exceeding eight orders of magnitude is achieved, highlighting the excellent performance of the cross-tier eDRAM. Furthermore, 52 cross-tier eDRAM cells measured across the 200-mm wafer exhibit consistent performance and high uniformity. Together, these results demonstrate fully functional cross-tier logic (inverters) and memory (MLFe-NVMs and 2TOC eDRAMs) integrated on 200-mm wafers, creating a significant milestone toward practical monolithic 3D integration based on ALD-grown InO_x semiconductors.

Indium oxide multi-tier CIM accelerator

To explore the potential applications and advantages of our 3D multifunctional logic and memory platform, we present a case study featuring a real 3D digital CIM accelerator designed for LLM applications. This design utilizes a custom-developed oxide semiconductor process design kit (PDK) with n-type metal–oxide–semiconductor-only logic gates based on our platform (Extended Data Fig. 6). The 3D multi-tier architecture, illustrated in Fig. 5, consists of four vertically stacked tiers integrating analogue/mixed-signal circuitry, logic, and both non-volatile and volatile memory, all interconnected via interlayer vias (ILVs) to realize a true 3D logic-on-memory architecture. The 3D-integrated circuit (IC) design flow is depicted in Fig. 5b. The entire design is initially partitioned into multiple tiers according to system-level specifications, with each tier represented by a register-transfer level (RTL) digital-top description. The partitioned design is synthesized and hierarchically mapped onto logic gates constructed from our InO_x E-mode and D-mode FETs, generating netlists

for each tier. Subsequently, a tier-wise physical design methodology is adopted. Each tier's physical design takes as input both its own netlist and an abstract plus ILV location file from the preceding tier. A standard 2D physical design flow—including floor planning, analogue macro placement, standard cell placement, clock tree synthesis, routing and postroute optimization—is performed for each tier. An example of digital shift-and-add implementation with our PDK is shown in Supplementary Fig. 9. The final output for each tier consists of its GDS file along with an updated abstract and ILV location file, which are passed on to the next tier in the design sequence.

The InO_x multi-tier 3D CIM macro is shown in Fig. 5c. Tier 1 comprises a 4-Mb non-volatile ferroelectric memory bank utilizing a Fe-NVM array to store all static weights of the LLM model. The second tier integrates analogue/digital peripheral circuits supporting the Fe-NVM array in tier 1, including word line (WL) drivers for programming and sense amplifiers for data readout. Tier 2 also incorporates digital CIM logic, such as adder trees, shift-and-add logic, local controllers and register-based input/output buffers for enabling Fe-FET-based non-volatile computing-in-memory (nvCIM). Tier 3 integrates a 2-Mb 2-T gain-cell eDRAM macro, comprising both the 2-T array and corresponding peripheral circuits, to support dynamic computation in the LLM model. Benefiting from the ultralong retention time of the 2-T cell, the design achieves a refresh-free eDRAM macro implementation. Tier 4 is the top tier of the design, containing digital CIM logic interfacing with tier 3 eDRAM and a top-level cross-tier controller for orchestrating computation and data movement across all four tiers. The complete design consists of approximately 1.2 million logic gates and is implemented using a custom-developed oxide semiconductor PDK, scaled to the 40-nm technology node. The final merged 3D-IC GDS layout is illustrated in Fig. 5d.

As shown in Fig. 5e, the backbone operation of the Transformer-based LLM (Supplementary Fig. 10) can be categorized into static and dynamic classes based on the nature of matrix-vector multiplication (MVM). In the multihead attention mechanism, the Q , K and V projections are static because they originate from the pretrained backbone, making them well suited for non-volatile nvCIM in Fe-NVM. In contrast, the computation of the attention scores $Q \times K^T$ and $A' \times V$ is dynamic, rendering it more suitable for CIM in dynamic eDRAM. Leveraging the monolithic integration of Fe-NVM and eDRAM in the 3D CIM macro and scaling up our 3D CIM macros, we propose an end-to-end LLM accelerator design, as illustrated in Fig. 5f. This architecture integrates 3D CIM macros, a softmax unit, accumulators, a global buffer and a network-on-chip (NoC). The LLM accelerator is evaluated using our in-house framework⁵⁵, which performs comprehensive model-to-chip mapping by processing detailed model parameters, chip configuration settings and calibrated M3D technology libraries (Supplementary Fig. 11). The framework systematically compiles model layers into static and dynamic operations, and optimizes operation scheduling and dataflows through exploration of diverse mapping schemes, guided by the architectural characteristics of the 3D CIM macro. Performance metrics of the accelerator—including energy consumption, speed-up and energy–delay product (EDP) improvements—are evaluated for inference workloads across state-of-the-art LLM models such as BERT, GPT-2 and DeiT (as shown in Fig. 5g)^{56–59}. Compared to a conventional 2D design, the proposed M3D architecture achieves a $1.4\times$ to $2.9\times$ speed-up and comparable EDP benefits for GPT-2 and BERT-small models, respectively, demonstrating system-level performance enhancements enabled by the M3D design and technology platform. The 2D baseline is defined as a planar oxide semiconductor FET implementation under the same footprint constraint and NVM/eDRAM composition ratio, using the same workload-mapping strategy and model configurations. This formulation captures the impact of M3D integration in terms of effective resource density and data movement. Benchmarking against state-of-the-art systems (Extended Data Table 2), including

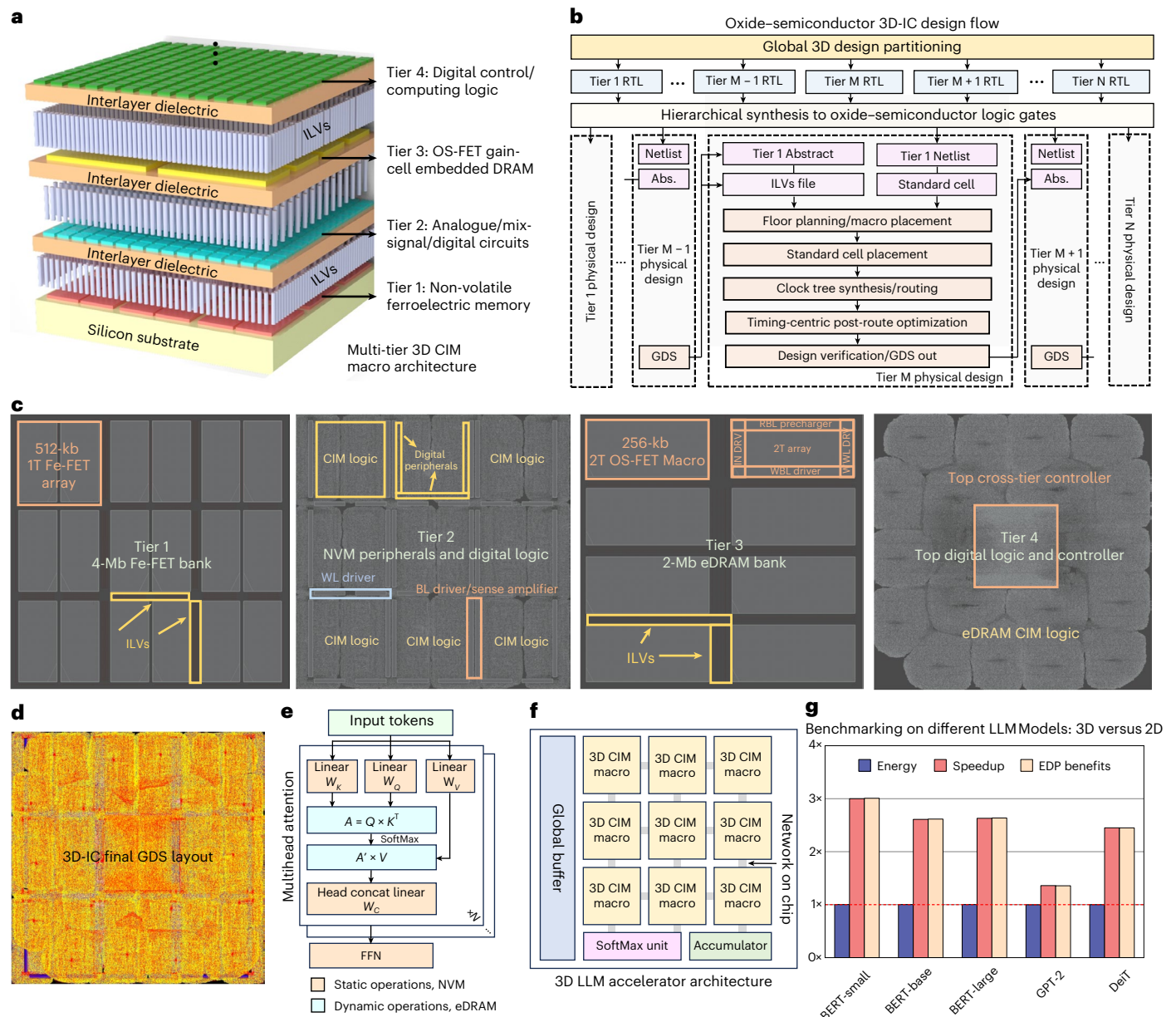


Fig. 5 | InO_x-based 3D multi-tier CIM LLM accelerator design and benchmarking. **a**, Schematic of the proposed multi-tier 3D CIM macro, featuring four tiers with scalability for additional stacking. **b**, Customized 3D-IC design flow from RTL to GDS using a PDK established for our InO_x-based oxide-semiconductor platform. **c**, Physical implementation of the multi-tier CIM macro, integrating non-volatile and dynamic memory, analogue peripherals, digital compute and control logic across vertical tiers. **d**, Final merged layout of the complete multi-tier 3D CIM macro. **e**, Partitioning of the Transformer model computation into static (non-volatile memory) and dynamic (eDRAM-based)

operations. **f**, Top 3D LLM accelerator architecture with multiple 3D CIM macros interconnected via on-chip networks. **g**, System-level benchmarking results showing energy, delay and energy–delay product comparisons between 3D and 2D architectures across various LLM models. A, attention score matrix; FFN, feed forward network; IN DRV, input driver; K^T, transposed key matrix; Q, query matrix; RBL, read bit line; V, value matrix; WBL, write bit line; W_c, concat layer linear matrix; W_k, projection matrix of key in attention; W_o, projection matrix of query in attention; W_v, projection matrix of value in attention; WWL DRV, write word-line driver.

2D-material-based processors⁶⁰, CNT-based processors⁴⁷ and recent 3D memory-centric architectures^{24,61}, highlights the key system-level advance of this work: the combination of enhanced integration capability and broader workload support.

Conclusions

M3D integration overcomes the limitations of traditional 2D architectures by enabling the vertical stacking of transistors with diverse functionalities on separate tiers, tightly interconnected to form compact, multifunctional systems. We demonstrated a wafer-scale M3D platform based on ALD-grown InO_x devices, achieving vertically

stacked and BEOL-compatible integration of ferroelectric, E-mode and D-mode FETs on 200-mm silicon wafers. The fabricated devices exhibit excellent electrical performance, high uniformity and robust cross-tier functionality, enabling the realization of multifunctional logic and memory systems in a compact 3D structure. By leveraging this platform, we implemented a full-stack CIM accelerator tailored for LLM workloads, showcasing the integration of non-volatile and volatile memory with digital and analogue logic across four vertically interconnected tiers. System-level evaluations highlight significant performance and energy-efficiency gains over conventional 2D architectures. Further scaling will require systematic design of the 3D stack,

tighter control of oxide semiconductor processing to tailor device characteristics, and careful management of parasitic coupling and heat transport, which cannot be inferred directly from 2D circuits. Larger hardware demonstrations are also needed to validate the projected CIM performance and energy efficiency.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-026-02276-0>.

References

- Wong, S. et al. Monolithic 3D integrated circuits. In *International Symposium on VLSI Technology, Systems, and Applications* <https://doi.org/10.1109/VTSA.2007.378923> (IEEE, 2007).
- Wan, W. et al. A compute-in-memory chip based on resistive random-access memory. *Nature* **608**, 504–512 (2022).
- Zheng, L. & Li, H. CMOS+X technologies for neuro-vector-symbolic computing. In *Device Research Conference—Conference Digest, DRC* <https://doi.org/10.1109/DRC61706.2024.10605469> (IEEE, 2024).
- Xiang, C. et al. 3D integration enables ultralow-noise isolator-free lasers in silicon photonics. *Nature* **620**, 78–85 (2023).
- Datta, S. et al. Amorphous oxide semiconductors for monolithic 3D integrated circuits. In *IEEE Symposium on VLSI Technology and Circuits* <https://doi.org/10.1109/VLSITechnologyandCir46783.2024.10631324> (IEEE, 2024).
- Bishop, M. D., Wong, H. S. P., Mitra, S. & Shulaker, M. M. Monolithic 3-D integration. *IEEE Micro* **39**, 16–27 (2019).
- Jayachandran, D. et al. Three-dimensional integration of two-dimensional field-effect transistors. *Nature* **625**, 276–281 (2024).
- Ghosh, S. et al. Monolithic and heterogeneous three-dimensional integration of two-dimensional materials with high-density vias. *Nat. Electron.* **7**, 892–903 (2024).
- Kim, K. S. et al. Growth-based monolithic 3D integration of single-crystal 2D semiconductors. *Nature* **636**, 615–621 (2024).
- Lu, D. et al. Monolithic three-dimensional tier-by-tier integration via van der Waals lamination. *Nature* **630**, 340–345 (2024).
- Niu, C. et al. First demonstration of BEOL wafer-scale all-ALD channel CFETs using IGZO and Te for monolithic 3D integration. In *IEEE International Electron Devices Meeting (IEDM)* 1–4 <https://doi.org/10.1109/IEDM50854.2024.10873541> (IEEE, 2024).
- Sun, Y. et al. BEOL Three-dimensional stackable oxide semiconductor CMOS inverter with a high voltage gain of 233 at cryogenic temperatures. *Nano Lett* <https://doi.org/10.1021/acs.nanolett.4c04701> (2024).
- Datta, S. et al. Back-end-of-line compatible transistors for monolithic 3-D integration. *IEEE Micro* **39**, 8–15 (2019).
- Salahuddin, S., Ni, K. & Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **1**, 442–450 (2018).
- Li, S. et al. Nanometer-thin indium tin oxide for advanced high-performance electronics. *Nat. Mater.* **18**, 1091–1097 (2019).
- Zhu, J. et al. Low-thermal-budget synthesis of monolayer molybdenum disulfide for silicon back-end-of-line integration on a 200 mm platform. *Nat. Nanotechnol.* **18**, 456–463 (2023).
- Yuvaraja, S. et al. Three-dimensional integrated metal–oxide transistors. *Nat. Electron.* **7**, 768–776 (2024).
- Han, K., Kang, Y., Chen, X., Chen, Y. & Gong, X. Indium–Tin–oxide thin-film transistors with high field-effect mobility ($129.5\text{ cm}^2/\text{V}\cdot\text{s}$) and low thermal budget (150°C). *IEEE Electron Device Lett.* **44**, 1999–2002 (2023).
- Gong, X. et al. Reconfigurable 3D stacked oxide semiconductor devices for hybrid in-memory computing. Preprint at <https://doi.org/10.21203/rs.3.rs-5138544/v1> (2024).
- Si, M. et al. Why In_2O_3 can make 0.7 nm atomic layer thin transistors. *Nano Lett.* **21**, 500–506 (2021).
- Si, M. et al. Scaled indium oxide transistors fabricated using atomic layer deposition. *Nat. Electron.* **5**, 164–170 (2022).
- Li, Q. et al. First demonstration of sequential integration for stacked gate-all-around a-IGZO nanosheet transistors with record $I_d = 2.05\text{ mA}/\mu\text{m}$, $g_m = 1.13\text{ mS}/\mu\text{m}$ and ultralow $\text{SS} = 66\text{ mV}/\text{dec}$. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM45741.2023.10413891> (IEEE, 2023).
- Lee, J. et al. BEOL-compatible non-volatile capacitive synapse with ALD W-doped In_2O_3 Semiconductor Layer. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM50854.2024.10873481> (IEEE, 2024).
- Liu, S. et al. Edge continual training and inference with RRAM-gain cell memory integrated on Si CMOS. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM50854.2024.10873546> (IEEE, 2024).
- Lin, Z. et al. Nanometer-thick oxide semiconductor transistor with ultra-high drain current. *ACS Nano* **16**, 21536–21545 (2022).
- Zhang, Z. et al. Ultrahigh bias stability of ALD In_2O_3 FETs enabled by high temperature O_2 annealing. In *IEEE Symposium on VLSI Technology and Circuits* <https://doi.org/10.23919/VLSITechnologyandCir57934.2023.10185292> (IEEE, 2023).
- Matsuda, S. et al. 30-nm-channel-length c-axis aligned crystalline In-Ga-Zn-O transistors with low off-state leakage current and steep subthreshold characteristics. In *IEEE Symposium on VLSI Technology and Circuits*, T216–T217 <https://doi.org/10.1109/VLSIT.2015.7223680> (IEEE, 2015).
- Niu, C. et al. Record-low metal to semiconductor contact resistance in atomic-layer-deposited In_2O_3 TFTs reaching the quantum limit. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM45741.2023.10413708> (IEEE, 2023).
- Niu, C. et al. Surface accumulation induced negative Schottky barrier and ultralow contact resistance in atomic-layer-deposited In_2O_3 thin-film transistors. *IEEE Trans. Electron Devices* **71**, 3403–3410 (2024).
- Lin, Z. et al. High-performance BEOL-compatible atomic-layer-deposited In_2O_3 Fe-FETs enabled by channel length scaling down to 7 nm: achieving performance enhancement with large memory window of 2.2 V, long retention > 10 years and high endurance > 10^8 cycles. In *IEEE International Electron Devices Meeting (IEDM)* 17.4.1–17.4.4 <https://doi.org/10.1109/IEDM19574.2021.9720652> (IEEE, 2021).
- Huyghebaert, C. et al. 2D materials: roadmap to CMOS integration. In *IEEE International Electron Devices Meeting (IEDM)* 22.1.1–22.1.4 <https://doi.org/10.1109/IEDM.2018.8614679> (IEEE, 2018).
- Samavedam, S. B. et al. Future logic scaling: towards atomic channels and deconstructed chips. In *IEEE International Electron Devices Meeting (IEDM)* 1.1.1–1.1.10 <https://doi.org/10.1109/IEDM13553.2020.9372023> (IEEE, 2020).
- Shi, Y. et al. Superior electrostatic control in uniform monolayer MoS_2 scaled transistors via in-situ surface smoothing. In *IEEE International Electron Devices Meeting (IEDM)* 37.1.1–37.1.4 <https://doi.org/10.1109/IEDM19574.2021.9720676> (IEEE, 2021).
- Chiang, K. H. et al. Integration of 0.75 V V_{DD} oxide–semiconductor 1T1C memory with advanced logic for an ultra-low-power low-latency cache solution. In *IEEE Symposium on VLSI Technology and Circuits* <https://doi.org/10.23919/VLSITechnologyandCir65189.2025.11074854> (IEEE, 2025).
- Zhao, Y. et al. Revealing the partially coherent nature of transport in IGZO. Preprint at <https://doi.org/10.48550/arXiv.2512.21945> (2025).

36. Liu, A. et al. Selenium-alloyed tellurium oxide for amorphous p-channel transistors. *Nature* **629**, 798–802 (2024).
37. Kwon, J. et al. 200-mm-wafer-scale integration of polycrystalline molybdenum disulfide transistors. *Nat. Electron.* **7**, 356–364 (2024).
38. Schram, T. et al. Integration and electrical evaluation of WS₂ and MoS₂ FETs in a 300 mm pilot line. *Discov. Electron.* **3**, 15 (2026).
39. Dorow, C. J. et al. Exploring manufacturability of novel 2D channel materials: 300 mm wafer-scale 2D NMOS & PMOS using MoS₂, WS₂, & WSe₂. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM45741.2023.10413874> (IEEE, 2023).
40. Mitard, J. et al. sub-40 mV sigma V_{TH} IGZO nFETs in 300mm fab. *ECS Trans. Prime* **98**, 205–217 (2020).
41. Asselberghs, I. et al. Wafer-scale integration of double gated WS₂-transistors in 300mm Si CMOS fab. In *IEEE International Electron Devices Meeting (IEDM)* 40–42 <https://doi.org/10.1109/IEDM13553.2020.9371926> (IEEE, 2020).
42. Qin, B. et al. Two-dimensional indium selenide wafers for integrated electronics. *Science* **389**, 299–302 (2025).
43. Lee, J. et al. Monolithic 3D integration of dual-gated ALD oxide-channel non-volatile capacitive memory on 40nm Si CMOS for digital compute-in-memory. In *IEEE International Electron Devices Meeting (IEDM)* <https://doi.org/10.1109/IEDM50572.2025.11353811> (IEEE, 2025).
44. Kang, J. H. et al. Monolithic 3D integration of 2D materials-based electronics towards ultimate edge computing solutions. *Nat. Mater.* **22**, 1470–1477 (2023).
45. Pendurthi, R. et al. Monolithic three-dimensional integration of complementary two-dimensional field-effect transistors. *Nat. Nanotechnol.* **19**, 970–977 (2024).
46. Liu, C. et al. A full-featured 2D flash chip enabled by system integration. *Nature* **646**, 1081–1088 (2025).
47. Hills, G. et al. Modern microprocessor built from complementary carbon nanotube transistors. *Nature* **572**, 595–602 (2019).
48. Wang, Q. et al. Wafer-scale highly oriented monolayer MoS₂ with large domain sizes. *Nano Lett.* **20**, 7193–7199 (2020).
49. Li, T. et al. Epitaxial growth of wafer-scale molybdenum disulfide semiconductor single crystals on sapphire. *Nat. Nanotechnol.* **16**, 1201–1207 (2021).
50. Lin, Z. et al. Enhancement of In₂O₃ field-effect mobility up to 152 cm² V⁻¹ s⁻¹ using HZO-based higher-k linear dielectric. In *IEEE Symposium on VLSI Technology and Circuits* <https://doi.org/10.1109/VLSITechnologyandCir46783.2024.10631497> (IEEE, 2024).
51. Si, M. et al. Enhancement-mode atomic-layer-deposited In₂O₃ transistors with maximum drain current of 2.2 A/mm at drain voltage of 0.7 V by low-temperature annealing and stability in hydrogen environment. *IEEE Trans. Electron Devices* **68**, 1075–1080 (2021).
52. Li, S. et al. Nanometre-thin indium tin oxide for advanced high-performance electronics. *Nat. Mater.* **18**, 1091–1097 (2019).
53. Zhang, J. et al. Effects of gallium on electron transport and bias stability in ultrascaled amorphous InGaO transistors. *IEEE Trans. Electron Devices* **72**, 4156–4162 (2025).
54. Park, H. et al. Extreme EOT scaling in tungsten-doped In₂O₃ MOSFETs for enhanced stability and drive current. *IEEE Trans. Electron Devices* **72**, 7136–7144 (2025).
55. Du, S. et al. 3D-CIMlet: a chiplet co-design framework for heterogeneous in-memory acceleration of edge LLM inference and continual learning. In *IEEE Design Automation Conference* <https://doi.org/10.1109/DAC63849.2025.11133077> (IEEE, 2025).
56. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 6000–6010 (2017).
57. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* Vol. 1 (eds Burstein, J. et al.) 4186–4171 (Association for Computational Linguistics, 2019).
58. Radford, A. et al. Language models are unsupervised multitask learners. *OpenAI Blog* **1**, 8 (2019).
59. Touvron, H. et al. Training data-efficient image transformers & distillation through attention. In *Proc. 38th International Conference on Machine Learning* Vol. 139, 10347–10357 <https://proceedings.mlr.press/v139/touvron21a.html> (PMLR, 2021).
60. Ao, M. et al. A RISC-V 32-bit microprocessor based on two-dimensional semiconductors. *Nature* **640**, 654–661 (2025).
61. Yakubo, Y. et al. Crystalline oxide semiconductor-based 3D bank memory system for endpoint artificial intelligence with multiple neural networks facilitating context switching and power gating. In *IEEE International Solid-State Circuits Conference (ISSCC)* <https://doi.org/10.1109/ISSCC42615.2023.10067603> (IEEE, 2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2026

Methods

ALD and device fabrication

The full fabrication process flow is shown in Fig. 1c. The device fabrication process began with 200-mm silicon wafers with 1.5- μm SiO_2 , followed by a standard cleaning process, including ultrasonic rinsing with toluene, acetone and isopropyl alcohol to remove possible organic particles and contaminants. A 10-nm layer of Al_2O_3 is deposited using plasma-enhanced atomic layer deposition (PE-ALD) at 200 °C. A standard lift-off process was then applied for the 40-nm nickel or platinum bottom gate for FETs and Fe-FETs. Next, a 6-nm HfO_2 dielectric or 15-nm HZO ferroelectric layer was deposited by PE-ALD at 200 °C, using $[(\text{CH}_3)_2\text{N}]_4\text{Hf}$ (TDMAHf) and $[(\text{CH}_3)_2\text{N}]_4\text{Zr}$ (TDMAZr) as hafnium and zirconium precursors. Subsequently, 1–3 nm In_2O_3 was deposited by ALD at 225 °C using $(\text{CH}_3)_3\text{In}$ (TMIn) and H_2O as indium and oxygen precursors. The TMIn precursor was heated to 60 °C to provide sufficient vapour pressure. Film thickness was accurately controlled by the number of ALD cycles. Channel isolation was done by dry etching the In_2O_3 layer using argon plasma. A 40-nm nickel layer was deposited as source/drain contacts. The devices were then annealed at various temperatures (190–300 °C) in an O_2 atmosphere. For ferroelectric layer activation, a 400 °C N_2 annealing was conducted. A 7-nm HfO_2 capping layer was deposited by PE-ALD at 150 °C, followed by a 10-nm PE-ALD Al_2O_3 insulating layer at 200 °C. Multi-tier structures were realized by repeating the fabrication steps. Three-tier oxide semiconductor Fe-FETs and FETs were successfully demonstrated on 200-mm silicon wafers. Notably, the entire fabrication process is compatible with BEOL integration, with a maximum thermal budget of 400 °C. In the present M3D device demonstration, the layout has not yet been optimized to minimize parasitic capacitance and coupling, both of which can be further reduced in future designs.

Wafer-scale electrical measurements and statistical methods

The fabricated multi-tier 200-mm wafer was characterized using a FormFactor PAV200 200-mm semiautomated probe station at 25 °C in a N_2 environment at ambient pressure. The electrical characterization of transistors, inverters and memories was measured with a Keysight B1500A Semiconductor Device Parameter Analyzer. The polarization versus voltage loops were measured using Radiant Premier II Ferroelectric Tester. Unless otherwise stated, the characterized devices had a channel length of 5 μm .

The subthreshold swing was extracted from the inverse slope of the transfer characteristic around a normalized drain current of $10^{-4} \mu\text{A} \mu\text{m}^{-1}$. The threshold voltage was extracted using the constant-current method at a normalized drain current of $10^{-3} \mu\text{A} \mu\text{m}^{-1}$. The field-effect mobility was extracted in the linear regime using the maximum-transconductance method. All device parameters were extracted from transfer characteristics measured at $V_{\text{ds}} = 0.05 \text{ V}$. Threshold voltage shifts in the bias stability tests were determined using the linear-extrapolation method.

Box plots show the 25th–75th percentiles, with centre lines indicating the medians and centre symbols indicating the means. Whiskers extend to the most extreme data points within $1.5 \times$ the interquartile range, where the interquartile range is the difference between the 75th and 25th percentiles. For the transfer characteristic overlays in Fig. 2, devices classified as outliers were excluded from the displayed curves to facilitate visualization of process-dependent device behaviour. Gaussian curves were generated using the sample means and standard deviations. Exact sample sizes are provided in the corresponding figure legends.

High-resolution STEM

TEM, selected-area electron diffraction, EDS elemental mappings and HAADF-STEM analysis were performed with an FEI Talos F200X operated at 200 kV. Samples were prepared by focused ion beam lift-out using a Helios G4 UX dual-beam system.

Oxide semiconductor PDK

A custom oxide semiconductor PDK was developed based on the measured electrical characteristics of the InO_x enhancement-mode and depletion-mode transistors. Parameterized Verilog-A transistor models were created for device- and circuit-level simulation in Cadence Virtuoso. An n-type metal–oxide–semiconductor-only standard-cell library was constructed using enhancement-mode transistors as pull-down switching devices and depletion-mode transistors as active loads. The logic functions and transient responses of the standard cells were verified through transistor-level simulations. The oxide semiconductor PDK includes Liberty-format (LIB) files describing the logic, timing and power characteristics of the standard cells, library exchange format (LEF) files containing their physical abstracts, technology files (TechLEF) defining the routing layers and ILV design rules, and GDS files containing the complete standard-cell layouts.

Circuit and physical design flow

The memory arrays and analogue or mixed-signal peripheral circuits were designed and simulated in Cadence Virtuoso using the custom Verilog-A transistor models. The digital computing and control modules were described in Verilog HDL and functionally verified using Questa. Logic synthesis was performed using Cadence Genus, where the RTL designs were hierarchically mapped to the custom InO_x standard-cell library. The synthesized netlists were then implemented in Cadence Innovus using a tier-by-tier 3D physical design flow. For each tier, the physical design process included floor-planning, analogue macro placement, standard-cell placement, clock tree synthesis, routing and post-route optimization. In addition to its own synthesized netlist, each tier used the physical abstract and ILV location file generated from the preceding tier as design inputs. After physical implementation, an updated abstract file and ILV location file were exported together with the GDS layout and passed to the subsequent tier. This sequential flow maintained the alignment of the ILVs and ensured consistent cross-tier connectivity. Finally, the layouts of all tiers were vertically aligned and merged to generate the complete 3D-IC GDS database.

System simulation and benchmarking

System-level evaluation was performed using our model-to-chip mapping framework. The framework takes the model topology and parameters, chip configuration, and calibrated InO_x technology libraries as inputs. Circuit-level characteristics were obtained from simulations of the memory arrays, peripheral circuits and standard cells using the custom oxide semiconductor PDK. For each workload, model-layer computations were partitioned into static and dynamic matrix-vector multiplication operations. These operations were mapped to the Fe-FET- and eDRAM-based CIM resources, respectively, within the 3D macro. Network-on-chip traffic patterns were simulated using the router microarchitecture and trace-driven traffic injection, and the associated communication costs were evaluated based on physical-design considerations. By integrating the technology-library parameters with layer-wise execution traces, the end-to-end latency, energy consumption and energy–delay product were estimated. Each inference workload was evaluated using identical model configurations for the 3D architecture and its 2D baseline. The 2D baseline was a planar oxide semiconductor implementation constrained to the same chip footprint, Fe-FET/eDRAM composition ratio and workload-mapping strategy as the 3D design.

Data availability

Data are available from the corresponding author upon request. Source data are provided with this paper.

Acknowledgements

We acknowledge Cadence for Innovus 3D-IC license support.

Author contributions

P.D.Y. supervised the project. C.N. designed the experiments. C.N., L.L., Z.L. and J.-Y.L. fabricated the devices. L.Z., S.D. and H.L. conducted the circuit-level simulation. C.L., J.L. and H.W. carried out the TEM/STEM measurements and image analysis. C.N., L.L., J.-Y.L. and K.N. performed the electrical measurements and analysed the data. P.D.Y. and C.N. wrote the manuscript and all the authors commented on it.

Funding

P.D.Y. and H.L. were supported in part by the US National Science Foundation under award number 2425498 with industry partners as specified in the Future of Semiconductors (FuSe2) programme. P.D.Y. was also supported by Semiconductor Research Corporation Global Research Program and Samsung Electronics, Inc. H.W. was supported by the US National Science Foundation (DMR-2016453 and DMREF-2323752).

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41565-026-02276-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41565-026-02276-0>.

Correspondence and requests for materials should be addressed to Peide D. Ye.

Peer review information *Nature Nanotechnology* thanks Lin Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Extended Data Table 1 | Device-level benchmark of BEOL-compatible channel platforms for monolithic 3D integration

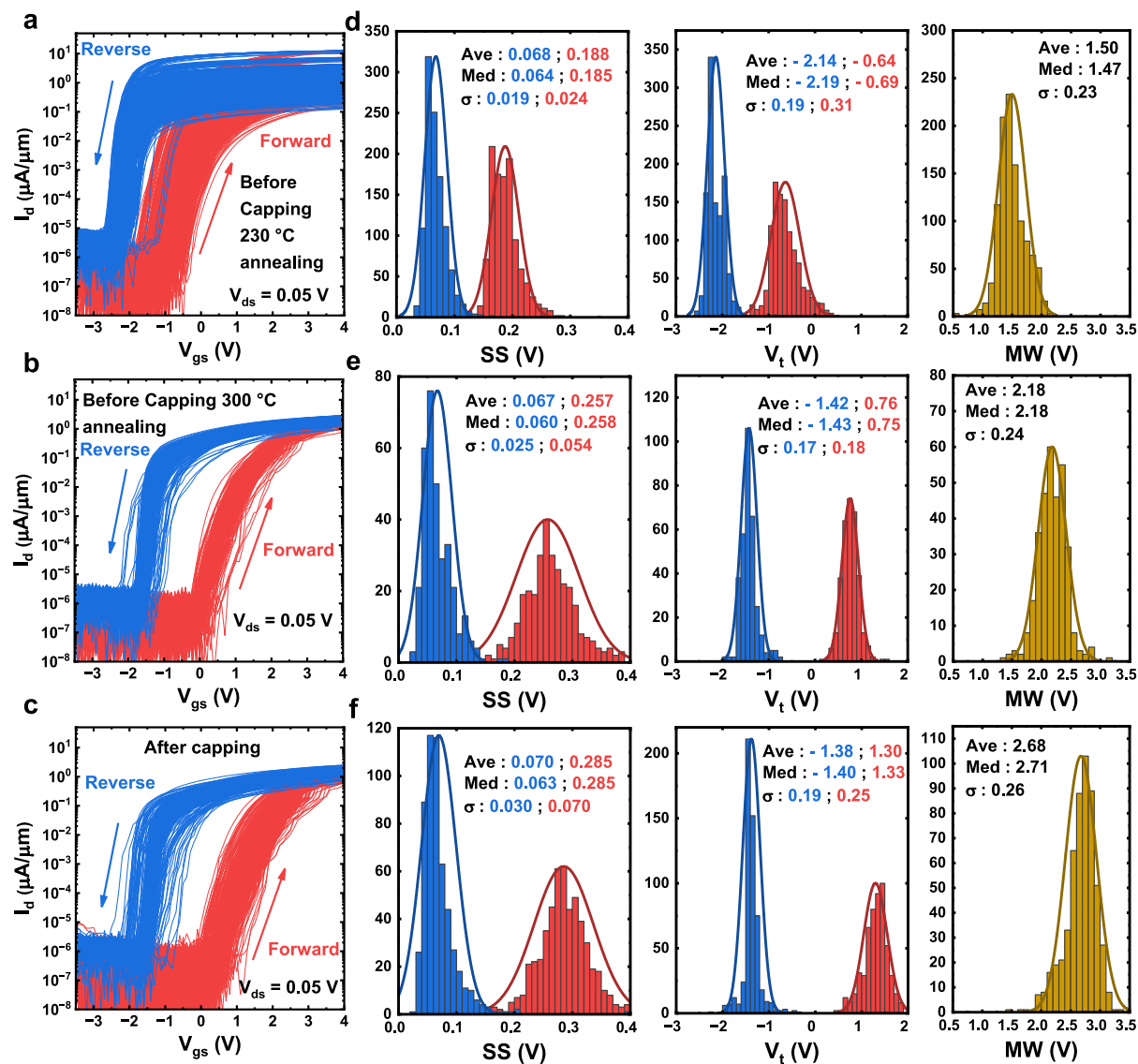
Material/Work	Method	Wafer Size	Monolithic Integration	Continuous Stacking	Max Process T	BEOL	Mobility (cm ² V ⁻¹ s ⁻¹)	Variability σV_t	System Evidence	Ref.
InO _x	ALD	200 mm	Yes	Yes (3-Tier)	400 °C	Yes	91.6	40.8 mV	Yes	This work (2026)
Si/OS	–	Fab	Yes	No	–	Yes	-	–	Yes	34 (2025)
IGZO	PVD	300 mm	No	No	< 400 °C	Yes	13	–	No	35 (2025)
MoS ₂	MOCVD	200 mm	Yes	No	300 °C	Yes	26.3	< 1 V *	No	16 (2023)
MoS ₂	MOCVD	200 mm	No	No	Transfer	Yes	21	< 1 V *	No	37 (2024)
WS ₂	MOCVD	300 mm	No	No	Transfer	Yes	3	1.1 V	No	38 (2026)
WS ₂	MOCVD	300 mm	No	No	–	–	-	–	No	39 (2023)
WSe ₂	MOCVD	300 mm	No	No	–	–	30	–	No	39 (2023)
MoS ₂	MOCVD	300 mm	No	No	–	–	7	–	No	39 (2023)
IGZO	PVD	300 mm	No	No	< 400 °C	Yes	-	40 mV	No	40 (2020)
WS ₂	PE-ALD/MOCVD	300 mm	No	No	300/Transfer	Yes	0.5	–	No	41 (2020)
Se-alloyed Te–TeO _x	Thermal Evaporation	100 mm	No	No	< 400 °C	Yes	15	–	No	36 (2024)
InSe	SLS Growth	50 mm	No	No	550 °C	No	287	< 0.2 V *	No	42 (2025)
MoS ₂ /WSe ₂	MOCVD	50 mm	Yes	Yes	Transfer	Yes	–	< 1 V *	No	7 (2024)
In ₂ O ₃	Sputter	Chip	Yes	Yes	Low	Yes	15	–	No	17 (2024)
Te–TeO _x /IGZO	ALD	Chip	Yes	Yes	225 °C	Yes	–	–	No	11 (2024)
Si/IWO	ALD	Chip	Yes	No	450 °C	Yes	–	–	Yes	43 (2025)
2D Materials	MOCVD	Chip	Yes	Yes	Transfer	Yes	10.4	–	Yes	44 (2023)
2D Materials	CVD	50 mm	Yes	Yes	Transfer	Yes	42	–	No	10 (2024)
2D Materials	MOCVD	50 mm	Yes	Yes	Transfer	Yes	3.11	< 1 V *	No	45 (2024)
2D Materials	CVD	Chip	Yes	Yes	385 °C	Yes	56.1	–	No	9 (2024)
Si/MoS ₂	CVD	Chip	Yes	No	Transfer	Yes	–	–	Yes	46 (2025)
CNT	-	150 mm	No	No	325 °C	Yes	–	–	Yes	47 (2019)

Comparison of mobility, wafer-scale process compatibility, process temperature, and suitability for continued multi-tier stacking. *Variability was approximated from the information provided in the paper^{79,11,16,17,34–47}.

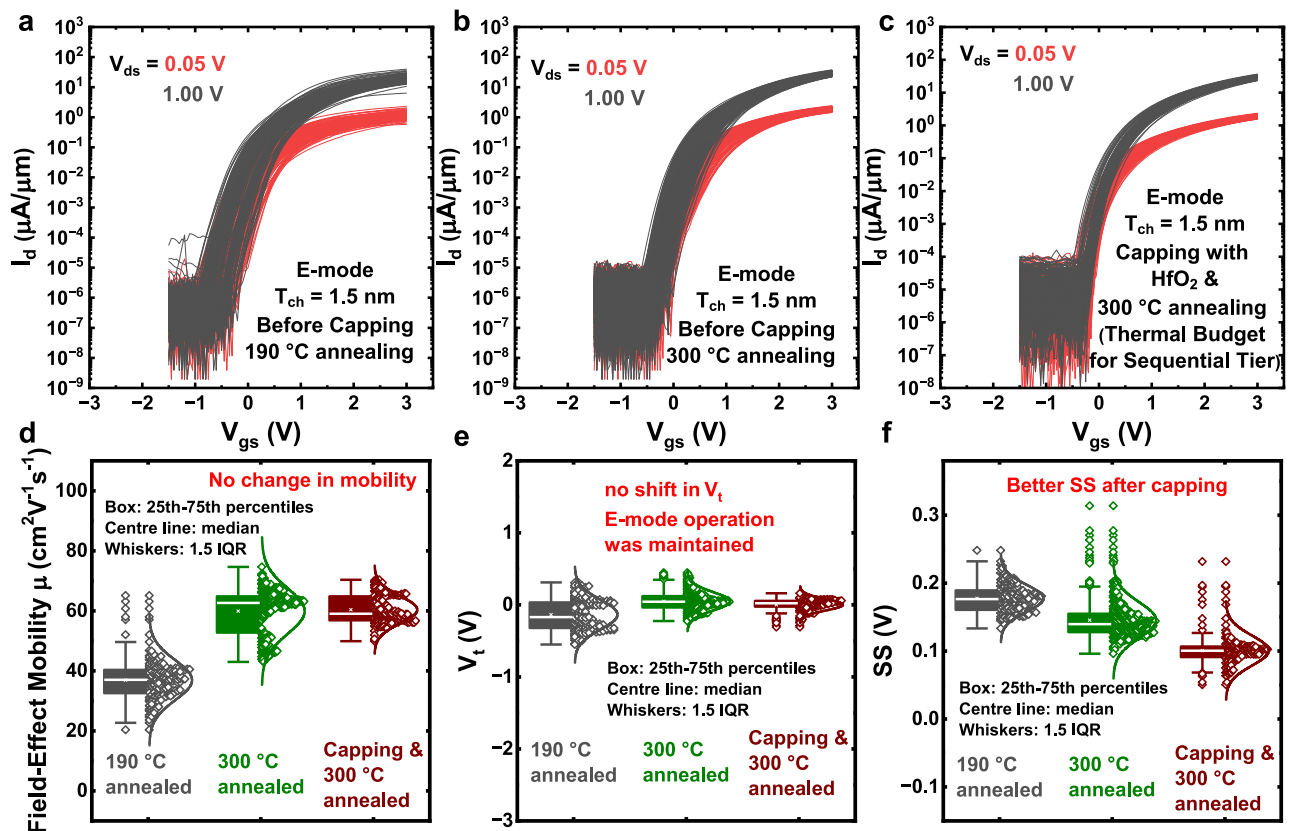
Extended Data Table 2 | System-level benchmark of Multi-Tier ALD Oxide Semiconductor 3D Integrated Circuits

Metric	Nature'25 ⁶⁰	Nature'19 ⁴⁷	ISSCC'23 ⁶¹	IEDM'24 ²⁴	This work
Material	MoS ₂	CNT	OS + Si	ITO + Si	InO _x
2D/3D	2D	2D	3D	3D	3D
Non-Silicon Logic	Yes	Yes	No	No	Yes
NVM Integration	No	No	No	Yes	Yes
Memory Type	Register file	Register file	eDRAM	eDRAM +RRAM	eDRAM +FeFET
Application	RISC-V CPU	RISC-V CPU	3D Memory SoC	LLM Accelerator	LLM Accelerator
Compute Density (OPS/mm ²)	13.9 ^{*1}	209.0 ^{*2}	NR	1.8×10 ¹⁰	2.1×10 ¹⁴
Energy Efficiency (OPS/W)	4.65×10 ⁵ @ 1 kHz ^{*1}	1.03×10 ⁷ @ 10 kHz ^{*2}	4.44×10 ¹² @ 400 kHz	9.91×10 ¹⁰ @ 100 MHz	1.49×10 ⁹ @ 100 MHz

^{*1} Estimated from reported data ^{*2} Assume one arithmetic operation per cycle (NR= Not Report)

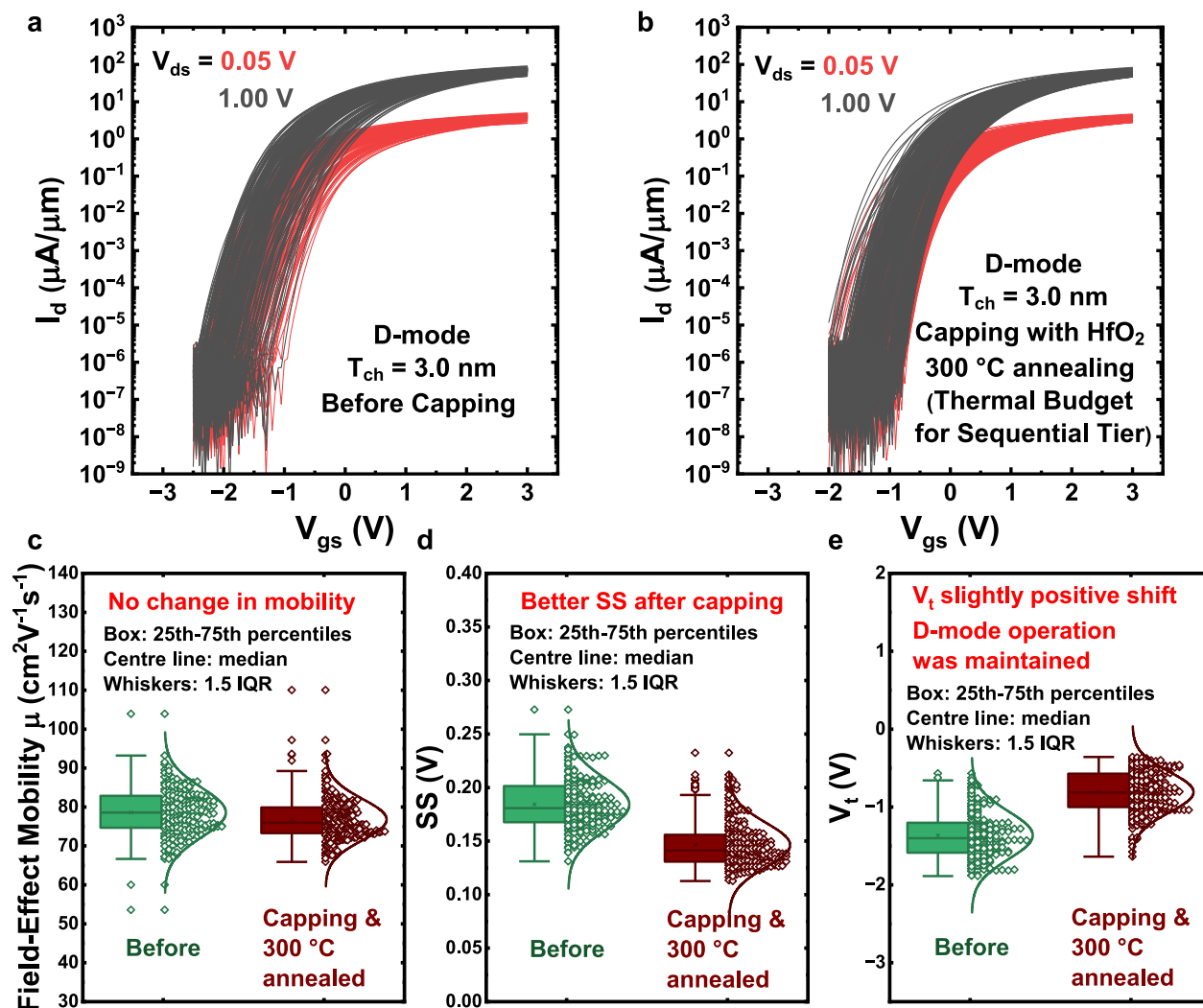


Extended Data Fig. 1 | Process-dependent electrical performance of Fe-FETs. Transfer characteristics of InO_x Fe-FETs measured after 230 °C annealing before capping (a), after 300 °C annealing before capping (b), and after capping followed by 300 °C annealing (c). **d-f.** The corresponding statistical distributions of SS, V_t , and MW, showing the thermal stability.



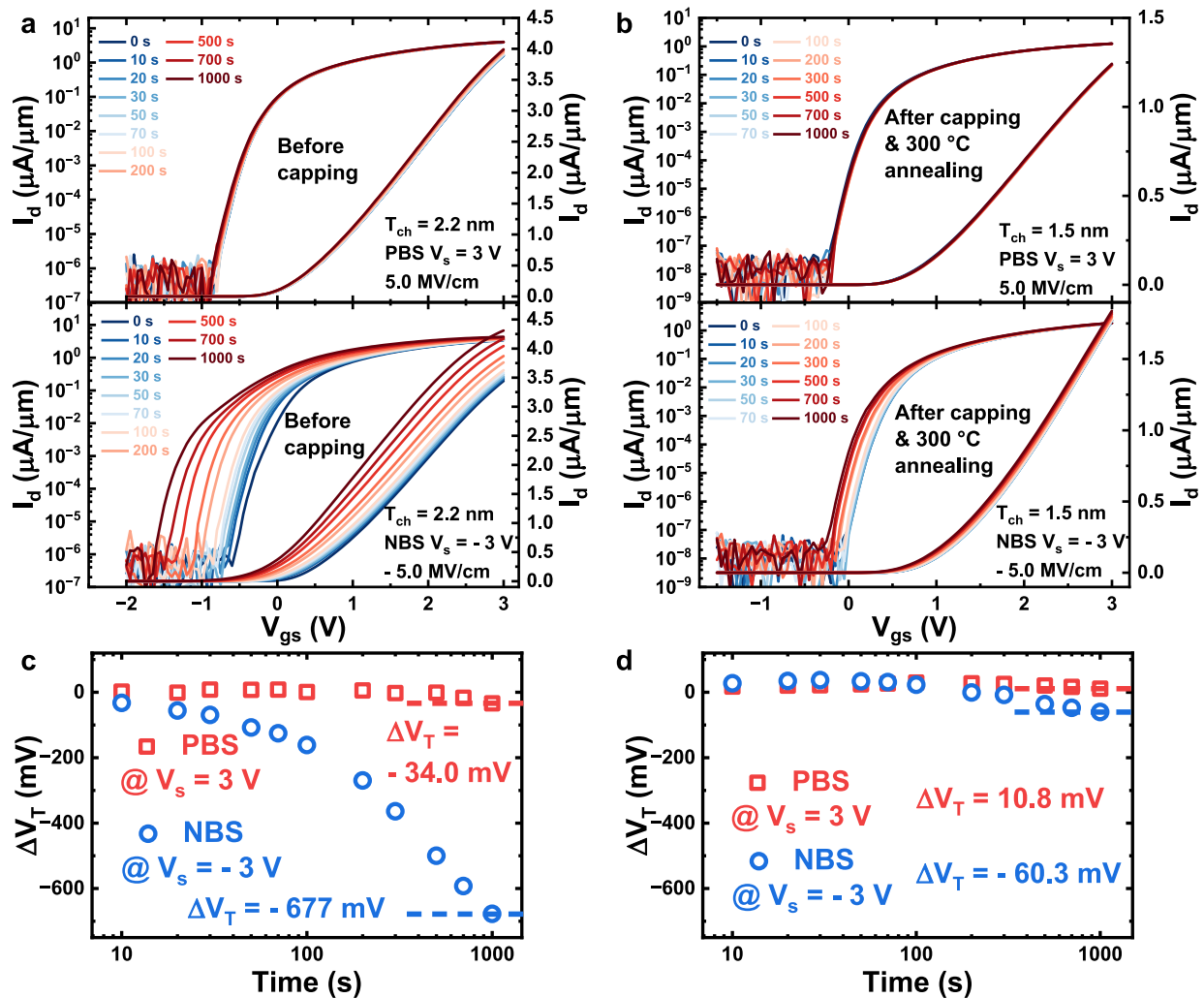
Extended Data Fig. 2 | Process-dependent electrical performance of E-mode FETs. Transfer characteristics of 1.5-nm-thick InO_x E-mode FETs measured after 190 °C annealing before capping (a), after 300 °C annealing before capping (b), and after capping followed by 300 °C annealing (c). The extracted μ (d),

V_t (e), and SS (f) indicate enhanced mobility, negligible threshold-voltage shift and improved electrostatics in the capped and post-annealed devices, further confirming the thermal stability of the InO_x material platform. The sample sizes were $n = 189, 489$, and 192 , respectively. Centre symbols indicate mean values.



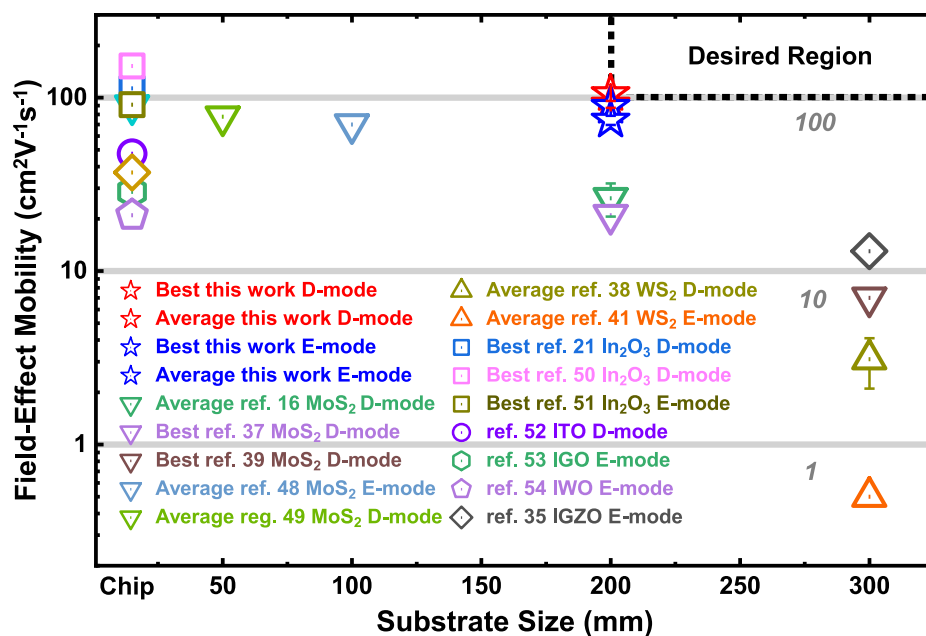
Extended Data Fig. 3 | Process-dependent electrical performance of D-mode FETs. Transfer characteristics of 3.0-nm-thick InO_x D-mode FETs measured after $300\text{ }^\circ\text{C}$ annealing before capping (a), and after capping followed by $300\text{ }^\circ\text{C}$ annealing (b). The extracted μ (c), SS (d), and V_t (e) collectively indicate

preserved mobility, minimal threshold-voltage shift and improved electrostatic control in the capped and post-annealed devices, further confirming the thermal robustness of the InO_x material system. The sample sizes were $n = 193$ and 292 , respectively. Centre symbols indicate mean values.

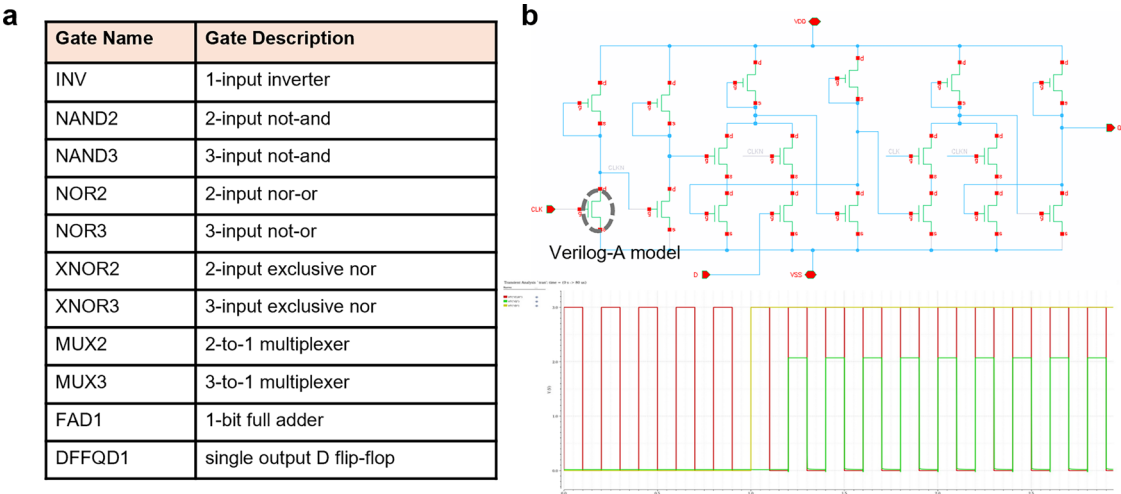


Extended Data Fig. 4 | Process-dependent bias stability of InO_x FETs. Positive- and negative-bias stability tests performed for up to 1000 s before capping (a) and after capping followed by 300°C annealing (b). The summarized

threshold-voltage shifts before capping (c) and after capping followed by 300°C annealing (d) show a substantial improvement in bias stability after capping, indicating that surface protection effectively suppresses reliability degradation.



Extended Data Fig. 5 | Device-level benchmark of BEOL-compatible channel materials. Comparison of representative BEOL-compatible channel materials by field-effect mobility and processed substrate size, showing that our platform combines high electron mobility with fab-scale substrate processing.



Extended Data Fig. 6 | ALD InO_x Process Design Kit (PDK). **a**, Compact OS PDK standard cell library consisting of 11 logic gates. **b**, D flip-flop gate circuit simulation with Verilog-A OSFET model and simulation waveform.