Perspective

The future transistors

https://doi.org/10.1038/s41586-023-06145-x

Received: 19 August 2020

Accepted: 27 April 2023

Published online: 16 August 2023

Check for updates

Wei Cao[1], Huiming Bu[2], Maud Vinet[3], Min Cao[4], Shinichi Takagi[5], Sungwoo Hwang[6], Tahir Ghani[7] & Kaustav Banerjee[1]✉

The metal–oxide–semiconductor field-effect transistor (MOSFET), a core element of complementary metal–oxide–semiconductor (CMOS) technology, represents one of the most momentous inventions since the industrial revolution. Driven by the requirements for higher speed, energy efficiency and integration density of integrated-circuit products, in the past six decades the physical gate length of MOSFETs has been scaled to sub-20 nanometres. However, the downscaling of transistors while keeping the power consumption low is increasingly challenging, even for the state-of-the-art fin field-effect transistors. Here we present a comprehensive assessment of the existing and future CMOS technologies, and discuss the challenges and opportunities for the design of FETs with sub-10-nanometre gate length based on a hierarchical framework established for FET scaling. We focus our evaluation on identifying the most promising sub-10-nanometre-gate-length MOSFETs based on the knowledge derived from previous scaling efforts, as well as the research efforts needed to make the transistors relevant to future logic integrated-circuit products. We also detail our vision of beyond-MOSFET future transistors and potential innovation opportunities. We anticipate that innovations in transistor technologies will continue to have a central role in driving future materials, device physics and topology, heterogeneous vertical and lateral integration, and computing technologies.

The history of electronics is generally composed of three major electron devices—the vacuum tube[1], the bipolar junction transistor (BJT)[2] and the metal–oxide–semiconductor field-effect transistor (MOSFET), as illustrated in Fig. 1. The first two devices had important roles in advancing modern computing; however, it is the rise of the MOSFET, particularly complementary metal–oxide–semiconductor (CMOS) technology, during the past six decades[3–5] that really ignited the explosive development of information technology, which has been serving as the powerful engine of contemporary human civilization. The most appealing merit of the MOSFET for very-large-scale-integration (VLSI) applications is that the continuous scaling down of its physical size drives every important metric—cost, performance, energy consumption and so on—towards efficiency. Moreover, the 'field effect' nature and complementary circuit (containing both n-type and p-type MOSFETs) topology of CMOS enable ultralow leakage power, which makes them extremely favourable in low-power applications. Therefore, after the silicon (Si)/silicon dioxide (SiO_2) interface (trap) problem was solved, which was the main bottleneck in the early days[4], CMOS technology quickly overtook bipolar transistors in the ever-growing digital market, and paved the way towards the VLSI-based information technology era. Thus far, CMOS technology has served the digital VLSI industry for over half a century and has been scaled down to sub-10-nm technology nodes[6]. However, the road to scaling has not been smooth. Many challenges, primarily short-channel effects (SCEs), have been plaguing CMOS technology since the device size entered the 1-μm regime. In the historical CMOS scaling before the twenty-first century, efforts were focused on reducing the physical gate oxide thickness and engineering the source, drain and channel doping profile[7], whereas in the modern scaling scenario during the past two decades, novel materials and device architecture[8–14], such as strained channel, high-dielectric-constant (k) metal gate (HKMG), silicon-on-insulator (SOI) and fin field-effect transistors (FinFETs), have been introduced to suppress SCEs and other adverse effects. According to the latest International Roadmap for Devices and Systems (IRDS)[15], scaling at sub-5-nm technology nodes will stall at physical gate lengths of 14 nm and 12 nm for low-power (LP) and high-performance (HP) applications, respectively, which would be an undesirable situation. Therefore, the primary objective of this Perspective is to identify the most promising sub-5-nm logic devices and technologies as well as the required research efforts, thereby rendering the widely distributed research activities in this field to be more focused and more efficient.

FET fundamentals

The functionality of a typical MOSFET, depicted in the transmission electron microscope image in Fig. 2a, is analogous to that of a generic tap. In a tap, we control the flow of water using a mechanical knob (or gate), whereas in a MOSFET we control the flow of charge carriers (electrons or holes)—from source to drain through a channel, using an electric gate through 'field effect' or capacitive coupling. From a semiconductor energy-band perspective, gate bias is used to modulate the energy bands of the channel underneath the gate, as illustrated in Fig. 2b for an n-type MOSFET, thereby controlling its mobile charge-carrier population (n_{mob}). Efficient modulation of the channel potential (φ_{ch})

[1]Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA, USA. [2]Advanced Logic and Memory Technology, IBM Research, Albany, NY, USA. [3]Université Grenoble Alpes, CEA-LETI, Grenoble, France. [4]Pathfinding, Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan. [5]Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo, Japan. [6]Samsung Advanced Institute of Technology, Suwon-si, Korea. [7]Pathfinding and Technology Definition, Intel Corporation, Hillsboro, OR, USA. ✉e-mail: kaustav@ece.ucsb.edu

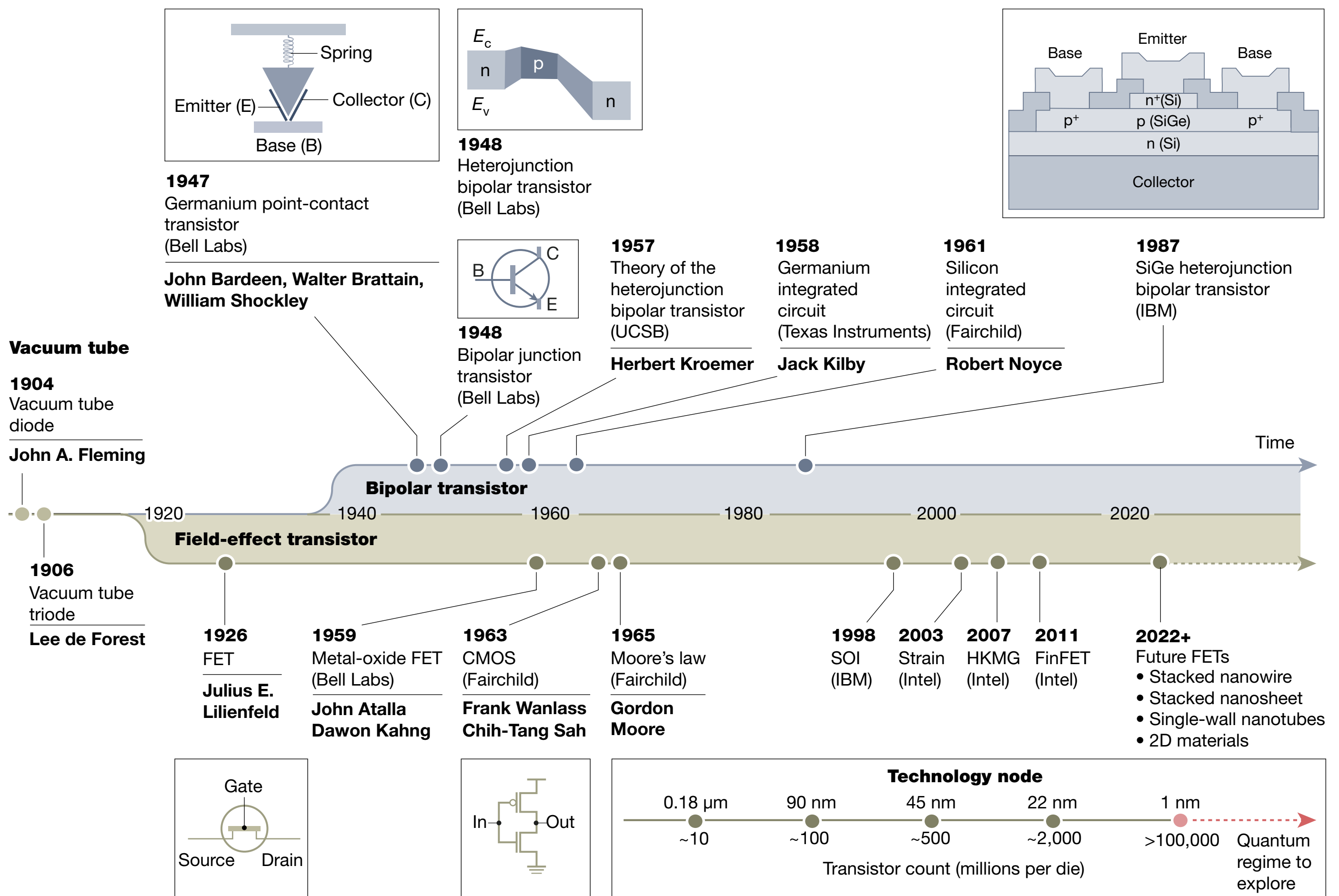


Fig. 1 | The history of transistor technology. All major transitions from vacuum tubes to BJTs, and eventually to MOSFETs, have been primarily driven by the need to reduce power consumption. Four major non-traditional FET scaling technologies, that is, SOI, strained channel, HKMG and FinFET are shown according to their commercialization time, corresponding to the 0.18-µm, 90-nm, 45-nm and 22-nm technology nodes, respectively. It is noted that beginning from the 22-nm node, the technology node becomes increasingly smaller than the FET physical dimension. E_c, conduction band minima; E_v, valence band maxima; IBM, International Business Machines; UCSB, University of California, Santa Barbara. Transistor count data are from https://en.wikipedia.org/wiki/Transistor_count.

through the gate voltage (V_g) is essentially an electrostatic problem, and can be understood by analysing the relevant capacitive couplings from all terminals to the channel where the centroid of mobile charges is located. These capacitances include the effective gate capacitance $C_{gox'}$ and quantum capacitance[16] C_Q, as well as parasitic capacitances from source/drain ($C_{s/d}$), substrate depletion (C_{dep}) and interface traps (C_{it}), as illustrated in Fig. 2c (see detailed explanation of these capacitances in the caption). The gate efficiency in modulating φ_{ch} can be derived as

$$\frac{\Delta\varphi_{ch}}{\Delta V_g} = \frac{C_{gox'}}{C_{gox'} + C_s + C_d + C_{dep} + C_{it} + C_Q} \quad (1)$$

C_Q, which is roughly proportional to n_{mob} (ref. 16), an exponential function (see equations in Supplementary Section 1) of φ_{ch} following the Fermi–Dirac distribution (see sketch in Fig. 2b), plays a unique role in determining the switching behaviour of FETs. When gate bias is below the threshold voltage (V_{th}), n_{mob} is minimal, thus C_Q is negligible with respect to $C_{gox'}$. Moreover, the parasitic capacitances do not change rapidly with V_g, thus V_g can tune φ_{ch} linearly according to equation (1). Therefore, the drain current (I_d, proportional to n_{mob}) changes exponentially (appears linear in the semi-log axis) with V_g in the subthreshold regime ($V_g < V_{th}$), as shown in Fig. 2d. The steepness of the I_d–V_g curve in this regime, usually quantified by a metric called subthreshold swing (SS), as shown in equation (2), is determined by the gate electrostatic efficiency (the first term) of a MOSFET, and the carrier transport mechanism (the second term) that equates to a thermionic emission limited constant minimum value of about 60 mV per decade of drain current span at room temperature for MOSFETs (see detailed derivation in Supplementary Section 2).

$$SS = \left(\frac{d[\log I_d]}{dV_g}\right)^{-1} = \frac{\Delta V_g}{\Delta\varphi_{ch}}\left(\frac{d[\log I_d]}{d\varphi_{ch}}\right)^{-1} \quad (2)$$

It is noted that n_{mob} and C_Q increase exponentially with V_g, and when C_Q becomes comparable to $C_{gox'}$, the linear φ_{ch} modulation gets decelerated. Eventually, C_Q becomes much larger than $C_{gox'}$ and any other capacitance components, causing the gate efficiency to approach zero, and any additional ΔV_g drops almost entirely across the gate oxide, instead of inside the channel. Therefore, the increase of n_{mob}, and hence I_d, begin to rely on $C_{gox'}\Delta V_g$, that is, in a linear manner (appears saturated in the semi-log axis), as illustrated in Fig. 2d. This transition of I_d from exponential increase to linear increase represents the physical essence of V_{th}.

In FET operation, the gate terminal is in charge of device switching, whereas the drain terminal is normally exploited to bias FETs in different operation modes after the device is turned ON. When the drain voltage (V_d) is small with respect to the overdrive voltage $V_{od} = V_g - V_{th}$, abundant mobile charges in the channel make FETs behave like resistors, which results in a linear I_d–V_d relation. When V_d increases above V_{od}, mobile

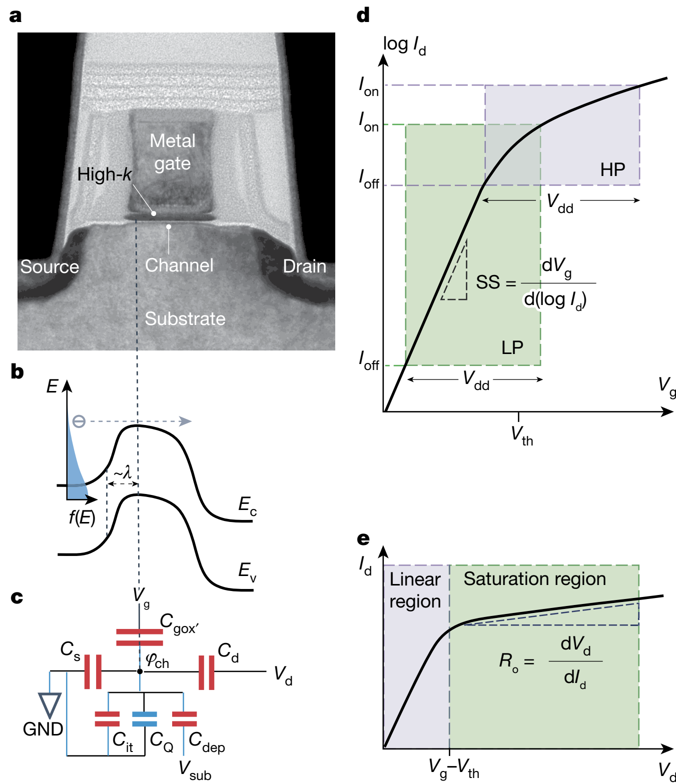


Fig. 2 | Fundamentals of FET physics and operation. a, Transmission electron microscope image of a 45-nm node planar n-MOSFET with HKMG. Credit: Intel Corp. **b**, Energy band diagram of an n-type FET. E_c, E_v and $f(E)$ are the conduction band minima (above which mobile charge carriers are electrons), the valence band maxima (below which mobile charge carriers are defined as holes) and the Fermi–Dirac distribution, respectively. λ is the natural length that determines the distance over which the potential changes from the source or drain to the channel. **c**, Schematic illustration of the capacitor network that determines the electrostatics in an FET with grounded source. $C_{s/d}$ are capacitances at source/channel and drain/channel junctions, respectively. In general, C_d is smaller than C_s owing to stronger depletion at the drain side. $C_{gox'}$ is the modified gate capacitance that accounts for the charge centroid shift from the channel surface owing to waveform spreading (can be described as an additional capacitance, C_{cent}, connected in series with the physical gate-oxide capacitance C_{gox}, which equivalently increases the gate oxide thickness). C_{dep} refers to capacitance induced by carrier depletion at the substrate surface. C_{it} is a conceptual capacitance induced by interface trap states ($= dQ_{it}/d\varphi_{ch}$), where Q_{it} is the interface trap charge density. C_Q is quantum capacitance ($= dQ_{mob}/d\varphi_{ch}$), where Q_{mob} ($= q \times n_{mob}$) is the mobile charge density and q is elementary charge. V_{sub} is the substrate bias, if any. **d**, Transfer characteristics, that is, I_d–V_g curve, and the definition of SS. The operation ranges of HP and LP applications are illustrated. **e**, Output characteristics, that is, I_d–V_d curve, and the definition of output resistance R_o.

charges become depleted at the drain side, leading to the pinch-off of the channel, and thereby to current saturation[17], as illustrated in Fig. 2e. In submicrometre-scale devices, large lateral electric-field-induced carrier velocity saturation can trigger an early current saturation before channel pinch-off occurs[17].

The most important metrics for FET in digital circuit performance are power consumption (P) and speed (or equivalently delay, τ)

$$P \propto C_{total} V_{dd}^2 f + I_{off} V_{dd} \tag{3}$$

$$\tau \propto C_{total} V_{dd} / I_{on} \tag{4}$$

where C_{total} is the total load capacitance, V_{dd} is the supply voltage, f is the operation frequency, I_{off} is the OFF current, which is normally specified for certain application, and I_{on} is the ON current, which is extracted by fixing I_{off} and V_{dd}, as illustrated in Fig. 2d.

FET scaling challenges

Although MOSFET scaling comes with great benefits[7], it raises many technological challenges.

Gate efficiency degradation

As FET gate length gets shorter, $C_{s/d}$ can become comparable to C_{gox}, that is, the source and drain begin to share the control over channel potential with the gate, thereby degrading SS, which is the primary manifestation of SCEs. Drain-induced barrier lowering (DIBL) is the other metric to quantify SCEs. As inferred from the name, DIBL lowers the channel potential barrier (with respect to the source) with increased drain bias, thereby lowering V_{th}, which leads to a non-ideal current saturation (quantified by output resistance R_o) in the output characteristics (Fig. 2e).

Parasitic resistance and capacitance

FET scaling requires the entire device length, which includes gate, contact and spacer lengths, termed as contacted gate pitch (CGP; Fig. 3a), to scale down. As a result, parasitic resistances, such as contact resistance $R_{contact}$, source/drain sheet resistance R_{sheet} and current-crowding-induced resistance R_{crowd} in the source/drain region, and parasitic capacitances, primarily composed of fringing and overlap capacitances (C_{frin}/C_{ov}) between the gate and the source/drain (contacts) (Fig. 3a), keep increasing, and hence begin to undermine the benefits of scaling.

Leakage currents

As scaling continues, the low-power merit of MOSFETs begins to be undermined by leakage currents, primarily from four sources (mechanisms are described in Fig. 3b): gate leakage that is composed of directing tunnelling (DT), Fowler–Nordheim (FN) tunnelling and hot-carrier (HC) injection; channel leakages comprising subthreshold (sub-V_{th}) that is enhanced by DIBL, and direct source-to-drain tunnelling (SDT) leakage; and junction leakages in the forms of reverse-biased diode leakage and gate-induced drain leakage (GIDL) in both of which band-to-band-tunnelling (BTBT) plays a key role. It is noted that as channel thickness reduces and V_{dd} becomes smaller, the drain depletion underneath the gate and, hence, the vertical (to channel surface) component of GIDL (shown in the sketch for GIDL) are suppressed, leaving only the lateral component along the channel direction.

Variability issues

Large-scale CMOS manufacturing, which is inherently imperfect, and simultaneous scaling unavoidably introduce variations in the channel and dielectric thicknesses, channel length, dopant density[18], gate material granularity[19] and so on. These variations are reflected in the device electrical characteristics, such as $I_{on/off}$ and V_{th}. As the device size (CGP) keeps decreasing, and the manufacturing process (such as lithography) gets increasingly complex, controlling and accounting for process variability has become a critical factor for any technology to be viable[20].

Reliability issues

In ultrascaled MOSFETs, both the vertical and the lateral electric fields get stronger, which inevitably stress the devices and degrade the reliability[21]. Moreover, the introduction of HKMG to replace the SiO_2/poly-Si gate stack introduces further device reliability issues. The degradation mechanisms of gate dielectric and channel/dielectric interface[22], such as time-dependent dielectric breakdown, bias temperature instability and HC injection, need re-examination.

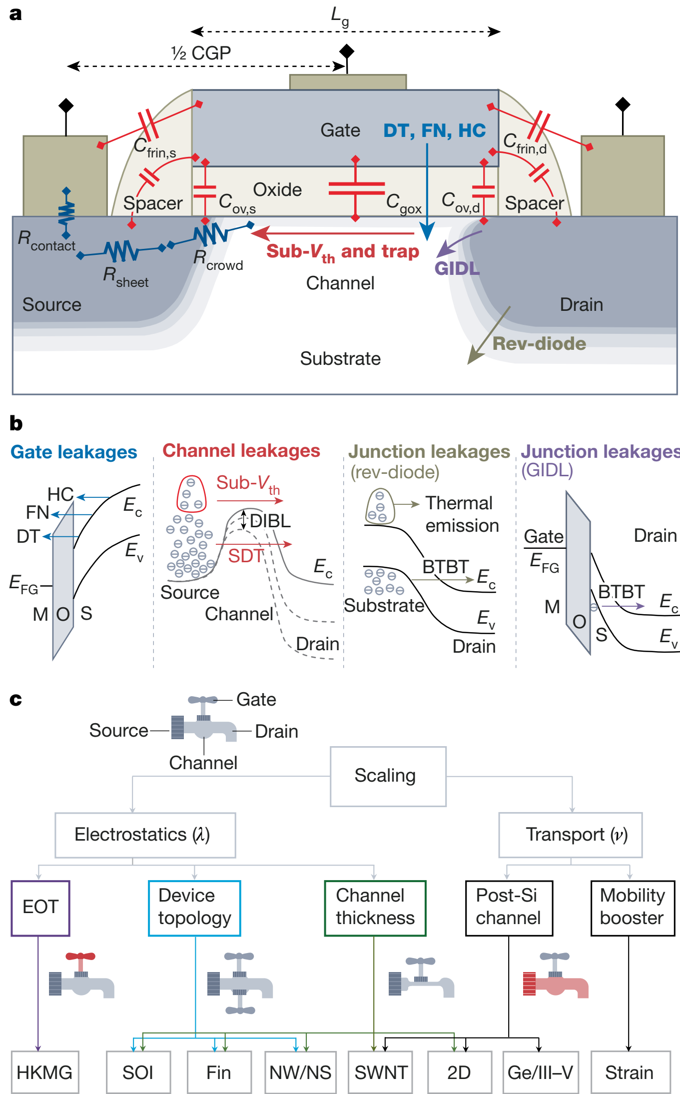


Fig. 3 | FET scaling. a, A typical FET structure, in which various issues during device scaling are illustrated. Rev-diode, reverse-biased diode leakage. **b**, Mechanisms of four major leakage paths in scaled MOSFETs. E_{FG} is the Fermi level of the gate electrode. **c**, A hierarchical illustration of FET scaling scenarios using a tap, whose operation mechanism serves as a good analogy to that of FETs, that is, the water/charge-carrier flow from source to drain through a channel is controlled by the knob/gate. Taps of different forms are employed to illustrate the general idea of different scaling scenarios. λ is the natural length (see equation (8)) and v is the carrier velocity (see equations (5) and (6)).

Modern CMOS scaling

Traditional MOSFET scaling[7], which relies on reducing the oxide thickness, along with source, drain and substrate doping-profile engineering[23], was eventually limited by large gate leakage through the gate oxide. Fortunately, modern MOSFET scaling has successfully addressed this issue, by employing HKMG, followed by many other novel technologies. In this section, the modern CMOS scaling history is reviewed within a hierarchical scaling framework (Fig. 3c).

Carrier transport engineering

Higher carrier velocity is desired to achieve better device performance at the same device size and supply voltage. However, if targeting the same device performance, higher carrier velocity alleviates the device-size scaling constraint, and/or enables lower operation voltages, and hence energy consumption.

Carrier velocity, either in the drift–diffusion limit υ_{DD}, (with various types of scattering involved) or in the ballistic limit $\upsilon_{ballistic}$ that determines the performance upper bound, is inversely proportional to transport effective mass (m^*_{tr})

$$\upsilon_{DD} = \xi_{tr}\mu_{DD} = \xi_{tr}\frac{q_0\tau_{MFT}}{m^*_{tr}} \tag{5}$$

$$\upsilon_{ballistic} = \frac{1}{\hbar}\frac{dE}{dk} = \frac{\hbar k}{m^*_{tr}} \tag{6}$$

where ξ_{tr} is the electric field along the transport direction, μ_{DD} is carrier mobility, q_0 is electron charge, τ_{MFT} is the mean free time (MFT) between two successive scattering events, $\hbar$ is the reduced Planck's constant, E is energy and k is the carrier wavevector. Therefore, low m^*_{tr} is desired to achieve high carrier velocity. There are mainly two pathways to achieve low m^*_{tr}.

One pathway is strain engineering (in production). Introducing strain into a Si lattice, thereby modifying the Si band structure, is one of these methods. The early efforts of introducing biaxial global strain by epitaxially growing a thin Si channel on top of a relaxed silicon–germanium (SiGe) virtual substrate[24] confronted two key challenges for process integration: (1) it could not provide the best strain configuration for both n-type and p-type transistors and (2) the SiGe layer induced a large number of defects in the strained Si. Subsequently, more integration-friendly local strain techniques were successfully developed, including the gate cap stressor (tensile strain)[25] and embedding SiGe in the recessed source/drain (compressive strain)[10] for enhancing electron and hole mobilities, respectively. These two techniques have been adopted in a complementary manner by the industry to achieve high-performance CMOS logic[10].

The other alternative (under research and development) is to replace the Si channel with high-mobility materials of either low m^*_{tr} or large τ_{MFT} (equation (5)). However, given the maturity of Si manufacturing technology, it is not practical to use any other materials as the primary substrate. Therefore, the first challenge for any potential new channel material is process integration with the Si substrate. Ge and III–V (specifically indium gallium arsenide (InGaAs)) materials are well known for their high hole and electron mobilities, respectively, and more importantly their lattice constants do not deviate much from that of Si (about 5.43 Å), which alleviates the process integration challenge to some extent. Therefore, these two materials are considered to be promising as post-Si channel materials[26–30]. The other challenge for any channel material is the lack of a high-quality gate oxide/insulator that can form a good interface with them, as in the Si/SiO_2 system. In this regard, Ge has been employed in the form of SiGe alloy with low Ge content for the channel[26]; thereby, the high-quality thermal SiO_2 could still be used. III–V materials cannot benefit from the well tuned Si/SiO_2 interface and, hence, a large density of interface states plagues the device performance and reliability of III–V channel MOSFETs[27]. The low bandgap of III–V materials requires V_{dd} to be small to avoid GIDL leakage currents, which limits V_{od} and hence the device performance. The larger dielectric constant of III–V materials, with respect to Si, is another unfavourable factor for scaling[30]. Moreover, the ultralow electron effective mass of III–V materials, while benefiting mobility, introduces a density of states (DOS) bottleneck[30,31] and large SDT leakage[32] starting from 20-nm gate length. The DOS for each quantized (vertical to the channel) level in any FET with a planar channel is

$$\mathrm{DOS} = \frac{g_s g_v\sqrt{m_x m_z}}{2\pi\hbar^2} \tag{7}$$

where $g_{s/v}$ is the spin/valley degeneracy, $m_{x/z}$ is the effective mass along the channel length/width direction. In summary, high performance of

III–V FETs heavily relies on the ultralow effective mass of III–V materials, which, however, prevents their scaling down.

It is worth mentioning that recently, III-nitride semiconductors, in particular, gallium nitride (GaN), have started gaining interest for CMOS application[33], because of their intrinsic material property merits such as higher (with respect to Si) electron mobility, bandgap and breakdown voltage. However, numerous technical challenges still remain, including large contact resistance, severely imbalanced n-type and p-type device performances, and process integration on Si wafer. Simply put, substantial efforts are needed to exploit the advantages of GaN for CMOS.

Improving device electrostatics

The scalability of FETs[34] can be quantified by a feature length called the natural length *λ*, which essentially captures the steepness of the potential variation from the source or drain to the channel. The first-order approximation (refer to ref. 35 for a more rigorous but complex form) can be written as

$$\lambda = \sqrt{\alpha T_{\mathrm{gox}} T_{\mathrm{ch}} \frac{\varepsilon_{\mathrm{ch}}}{\varepsilon_{\mathrm{gox}}}} \quad (8)$$

where $\varepsilon_{\mathrm{gox/ch}}$ and $T_{\mathrm{gox/ch}}$ are the permittivity ($= k \times \varepsilon_0$, where ε_0 is permittivity of free space) and thickness of gate dielectric/channel, respectively. α is a geometric factor that captures the gate topology, and decreases with an increasing number of gates in an FET. For good electrostatics, channel length $L_{\mathrm{ch}} > (3\text{–}5)\lambda$ is required. According to equation (8), scalability can be improved along three main directions—reducing the equivalent oxide thickness (EOT; that is, reducing the $T_{\mathrm{gox}}/\varepsilon_{\mathrm{gox}}$ ratio), employing multi-gate device topology (that is, reducing α) and reducing the channel thickness (T_{ch}).

The first direction, EOT reduction, is in production. To avoid large gate leakage while continuing scaling, a high-*k* dielectric was employed[11] to replace SiO_2. A high-*k* dielectric can deliver an equivalent C_{gox} (per unit area) that SiO_2 can provide at only a very small thickness, which is called EOT.

$$\mathrm{EOT} = T_{\mathrm{HK}} \frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{HK}}} \quad (9)$$

$$C_{\mathrm{gox,HK}} = \frac{\varepsilon_{\mathrm{HK}}}{T_{\mathrm{HK}}} = \frac{\varepsilon_{\mathrm{SiO_2}}}{\mathrm{EOT}} \quad (10)$$

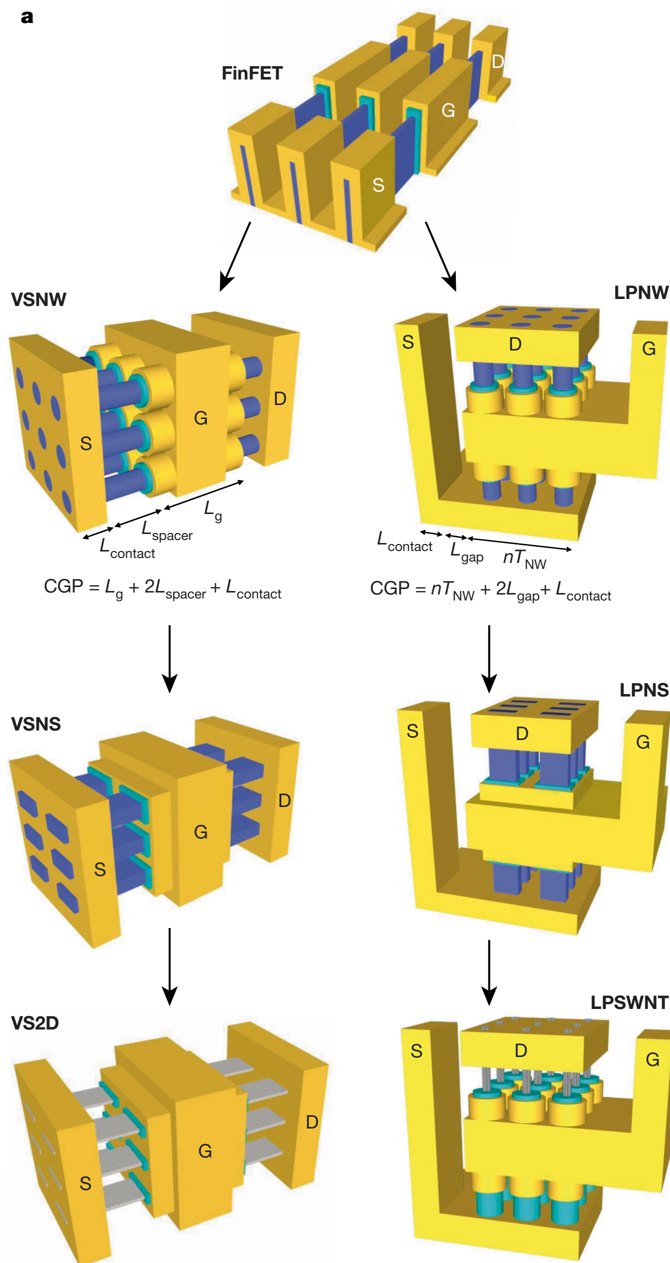
where T_{HK} is the physical thickness of the high-*k* dielectric, and $\varepsilon_{\mathrm{HK}}$ and $\varepsilon_{\mathrm{SiO_2}}$ are the dielectric permittivities of the high-*k* dielectric and SiO_2, respectively.

Although a high-*k* dielectric is appealing in terms of device scalability and performance, there are many challenges[11,13,14,17] to address, including the difficulty in obtaining a high *k* and a high bandgap simultaneously, high-*k* dielectric/poly-Si interface thermal-instability-induced Fermi-level pinning and hence large V_{th}, and the channel-surface optical phonon scattering induced mobility degradation. Through a decade of research, the semiconductor industry converged on hafnium oxide (HfO_2), which provides a reasonably high *k* of about 20 and a high bandgap of about 5.7 eV. A metal gate was employed to replace the poly-Si gate, which not only avoids the enduring gate depletion issue and the Fermi-level pinning effect for the poly-Si gate but also screens out the surface optical phonon scattering and thereby significantly improves carrier mobility. The high-*k* thin film and a high-*k* dielectric/Si channel interface quality have been greatly improved by employing the reaction self-limiting atomic-layer-deposition technique, and a SiO_2 interfacial layer between the high-*k* dielectric and the Si channel. Aided by these efforts as well as the aggressive gate last process (with respect to the traditional gate first process), Intel delivered the revolutionary first-generation HKMG based chips in 2007[11]. Further improvement directions[36,37] include introducing new high-*k* dielectrics with a higher *k* (such as lanthanum oxide) and/or reducing the thickness of the interfacial layer.

The second direction, novel device topology, is in production and under development. The introduction of SOI technology, in which a buried SiO_2 layer is employed to decouple the channel and substrate, not only eliminates any substrate leakage[8] and latch-up effects but also brings immunity to irradiation-induced failure from high-energy particles, which is crucial for outer-space and high-altitude electronics. However, it also raised great concerns such as the high cost of SOI wafers and the self-heating effect[9] caused by the much lower thermal conductivity of SiO_2, with respect to Si, thus severely limiting the SOI market.

The introduction of non-planar or three-dimensional (3D) transistors[38,39], particularly FinFETs[40,41] (Fig. 4a) revolutionized FETs and accelerated the shrinkage of the SOI market. The tri-gate FinFETs are manufacture friendly and can improve the current drive capability by increasing the fin height, and/or decreasing the fin pitch. A side effect of the FinFET structure in circuit design is the quantized fin height that narrows the design space in terms of allowed device widths[42]. It is worthwhile mentioning that FinFETs can also be made on SOI substrate, which has the advantage of minimized substrate leakage such as subfin leakage[43], while suffering the penalty of increased cost and the self-heating effect. Thus far, commercial FinFETs have evolved to the 5-nm technology node and beyond[6,29], and are expected to survive at least one more generation. The increasing challenges for FinFET technology to continue scaling beyond the 5-nm technology node include the 3D-structure-induced large parasitic capacitance, the high-aspect-ratio fins that are mechanically unstable, the small fin pitch that makes HKMG formation and raised source/drain epitaxy[44] in between the fins very difficult, and the minimum fin width (thickness) limited to about 4 nm (ref. 45), beyond which device performance undergoes rapid degradation, as shown in Fig. 4b. In other words, FinFET technology might find it hard to survive when the physical gate length becomes smaller than 10 nm.

Scaling theory[34] (equation (8)) indicated that the gate-all-around nanowire (NW) structure provides the best electrostatics from the gate-count point of view. This has been confirmed by both experimental and theoretical studies[46,47]. Moreover, vertically stacked NW (VSNW) FETs[48] and laterally packed NW (LPNW) FETs[49] (Fig. 4a), have been demonstrated, which effectively increases the current drive capability with respect to a single NW FET. VSNWs can be made based on FinFET technology, that is, this structure can benefit from the maturity of FinFET process, and hence can save tremendous process development efforts. In contrast, the fabrication processes (bottom up or top down) of LPNW FETs need a lot more investment. The vertical channel of LPNW FETs enables the relaxation of the channel length scaling, and an area and cost reduction without a leakage penalty. However, as technology node shrinks, their scaling advantage could diminish owing to the trend that the effective NW thickness (T_{NW}) and the total width of the laterally placed source, drain and gate contacts (Fig. 4a) become comparable to the gate length. Recently, vertically stacked nanosheet (VSNS) FETs[50] (Fig. 4a) and laterally packed nanosheet (LPNS) FETs[51] (Fig. 4a), derivatives of VSNW and LPNW FETs, respectively, have been developed. The wider nanosheet (NS) with respect to NW, while sacrificing a bit of electrostatics (thus, higher DIBL and SS) owing to the deviation from the ideal cylindrical gate-all-around structure[34], provides more surface area for current conduction and hence higher drive current and performance (Fig. 4c). Interuniversity Microelectronics Centre (IMEC) introduced a dielectric wall in a single NS stack to separate n-type and p-type devices, forming a 'forksheet' structure[52], which allows for tighter n-to-p spacing, and hence more area reduction, at the cost of degraded electrostatics owing to the missing gate in the wall region. It is noted that these NW and NS devices have to inherit the raised and silicided source/drain and low-*k* spacer techniques[6] from FinFETs to minimize parasitic resistances and capacitances, respectively.



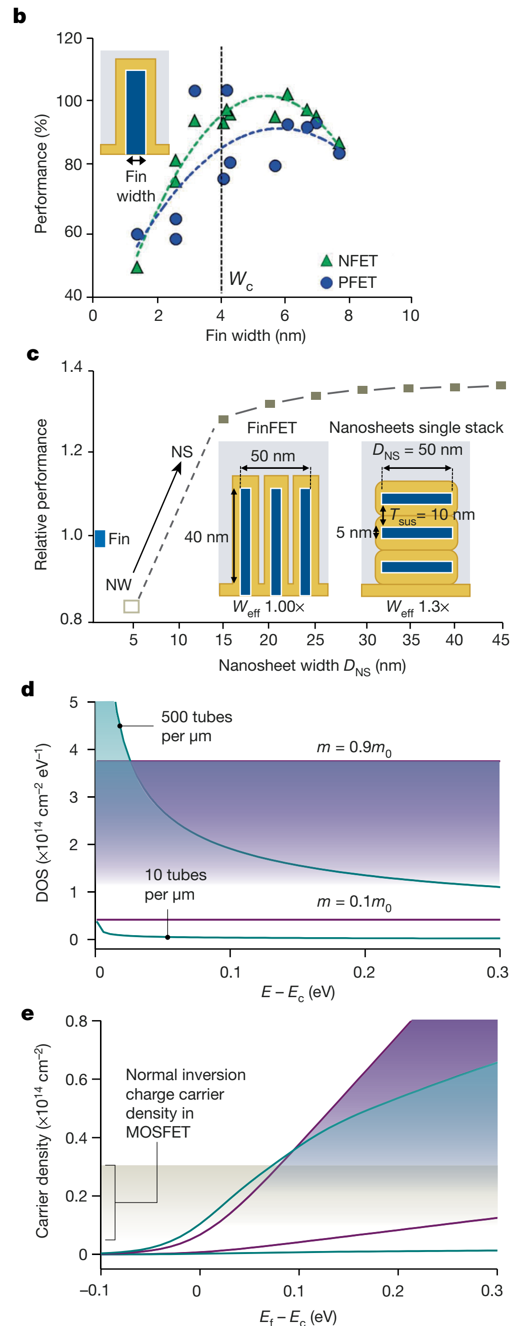


Fig. 4 | Transistors go beyond FinFET. a, Schematics of a FinFET, a VSNW FET, an LPNW FET, a VSNS FET, an LPNS FET, a VS2D FET and an LPSWNT FET. S/G/D represents source/gate/drain, $L_{contact}$/L_{spacer}/L_{gap} is the length of contact/spacer/S-to-D gap. n is the number of channels in the lateral direction. **b**, FinFET performance (operation frequency) versus fin width (thickness), suggesting that 4 nm is the optimal fin width (W_c). The performance percentage values shown are relative to a production baseline 14 nm FinFET technology with Fin width of 7 nm. NFET/PFET represents n-/p-type FET. **c**, The advantages of VSNS with respect to FinFET in terms of effective device width W_{eff} owing to increased total perimeter of channel surface for charge-carrier conduction (white)[50]. The relative performance values shown are with respect to that of the single data point (Fin) for a FinFET. D_{NS} is the nanosheet width, and T_{sus} is the inter-sheet distance. **d**,**e**, DOS (**d**) and carrier density (**e**) of a SWNT array (teal) and a 2DS (purple). E_f is the Fermi level. Panel **b** reproduced with permission from ref. 45, IEEE. Panel **c** adapted with permission from ref. 50, The Japan Society of Applied Physics.

Although NW and NS FETs are promising for replacing FinFETs at the 5-nm node and even beyond, in terms of process maturity and device performance, they are also facing challenges, such as the self-heating effect[53] caused by their one-dimensional heat transport, multiple surface-orientation-induced interface issues and large variability[54]. Moreover, similar to the FinFETs, the channel thickness of NW and NS FETs cannot be scaled below 3 nm, owing to quantum confinement (leading to reduced DOS), mobility degradation, large variability and fabrication difficulties. Therefore, it remains over optimistic to expect that NW and NS FETs can extend scaling to sub-1-nm nodes where physical gate lengths are expected to be smaller than 10 nm.

The third direction, atomic-scale channel thickness, is under research. Single-wall carbon nanotubes (SWNTs)[55] and the emerging 2D layered semiconductors (2DS), in particular, transition-metal dichalcogenides[56,57], have the capability of scaling FETs to the 1-nm node and beyond because of their intrinsic atom-scale thicknesses (≤1 nm).

SWNTs and 2DS also have the advantages (with respect to bulk materials) of a dangling-bond-free surface and atomically uniform thickness, which effectively suppress interface trap generation and variability, respectively, in FET applications. It is worth noting that future FETs are unlikely to go back to a planar structure. Even though SWNTs and 2DS have promising material properties, they have to be integrated into the modern 3D structures (see the suggested vertically stacked 2D (VS2D) FET and the laterally packed SWNT (LPSWNT) FET in Fig. 4a) to make them relevant to VLSI technology. Figure 4d,e shows the calculated DOS and carrier density, respectively, for a SWNT (based on a tight-binding model) array and 2DS (based on effective mass model). As shown, as long as the tube density of the SWNT array and the effective mass of the 2DS are reasonably large, SWNT and 2DS channels can provide sufficiently large DOS and comparable inversion carrier density with respect to the Si channel.

The first semiconducting SWNT FET was demonstrated in 1998 with a back-gate structure[58]. Subsequently, an atomic-layer-deposition-based process was developed to deposit zirconium dioxide (ZrO_2) on the pristine surface of a carbon lattice that lacked nucleation centres, allowing the demonstration of a high-k top-gated SWNT FET together with a hole mobility reaching 3,000 $cm^2\,V^{-1}\,s^{-1}$ (ref. 59). The mechanism behind this success was later attributed to the wetting functionality of several specific metals (scandium, yttrium and zirconium) to the SWNT surface[60]. It was also found that palladium (Pd) was suitable to form a p-type ohmic contact with the SWNTs, and could deliver a ballistic-transport current of about 25 μA per tube[61]. Excellent scalability of the SWNT FET was experimentally proved by showing negligible SS degradation when the channel length was scaled from 15 μm to 15 nm (ref. 62). Recently, a top-gated SWNT FET with 5-nm channel/gate length and graphene contact was demonstrated[63]. The measured SS was as low as 73 mV dec^{-1}, indicating that negligible SDT leakage was involved, probably owing to the fact that carrier effective mass of 1-nm SWNT can be as high as $0.68m_0$ for certain chiralities[64]. In 2015, an IBM group[65], developed an end-contact approach for SWNT by forming a molybdenum carbide (Mo_2C) alloy in the contact area, which shows negligible Schottky barrier and contact-length dependence of contact resistance. With this technology, the IBM group successfully scaled the footprint of SWNT FETs to 40 nm (ref. 66). It is noted that the ON–OFF current ratio of most demonstrated SWNT FETs is relatively low[59–62,66], owing to the small bandgap (about 0.7 eV) of 1-nm SWNTs[64]. In other words, SWNT FETs should be targeted for HP rather than for LP applications.

No matter how good a single SWNT FET is, SWNT FET technology can be competent only when a high-density, high-purity (semiconducting), well aligned wafer-scale SWNT array can be achieved with a CMOS-compatible process. Metallic SWNTs can be removed by several approaches, including thermocapillary flows[67], electrical breakdown[68] and density-gradient ultracentrifugation[69]. A surface-electrochemistry-assisted self-assembly[70,71] technique has been developed for SWNT placement and alignment[70]. Recently, Liu et al.[72], using a solution process, realized a decently high density (120 tubes per μm) and an average alignment degree of 9°, simultaneously. These achievements are exciting, but significant efforts are needed to enhance the n-type device performance (currently not on par with the p-type device), control variability[63,66,73] and develop CMOS-compatible processes.

2DS, such as molybdenum disulfide (MoS_2) and black phosphorus (BP), arose in the surge of graphene (a 2D semimetal) research[56,57,74]. Their planar structure and relatively large bandgaps offer great advantages, with respect to SWNTs, in terms of developing CMOS-compatible process and LP FETs[57]. Therefore, the 2D FETs have witnessed explosive growth[75,76] and are considered more feasible for high-volume production. It is noted that although graphene nanoribbons can be considered as a type of 2DS, their potential for FET application is not on par with other 2DS owing to the difficulty in fabricating large arrays of uniform graphene nanoribbons with width ≤1 nm needed to achieve a bandgap >0.5 eV. Interested readers are referred to a recent review[77] for more details.

In 2011, the first top-gated monolayer (1L) MoS_2 FET was demonstrated, realizing an ON–OFF current ratio and SS of 10^8 and 74 mV dec^{-1}, respectively[78]. This success greatly stimulated the enthusiasm of the device community on 2D FET research. A key limitation was that the measured electron mobility in 1L-MoS_2 (refs. 78,79) was as low as 13 $cm^2\,V^{-1}\,s^{-1}$, which is far below the predicted phonon scattering limited mobility[80] of about 410 $cm^2\,V^{-1}\,s^{-1}$. Such a large mobility loss is attributed to scattering with high-density defects (mostly charged) in MoS_2 and dielectric surface roughness (multilayer 2D channel is less affected)[81]. In 2013, inspired by the idea[82] of employing a high-k dielectric to suppress Coulomb scattering, Liu et al.[83] and Fang et al.[84] managed to obtain nearly 200 $cm^2\,V^{-1}\,s^{-1}$ electron and hole mobilities from 1L tungsten diselenide (WSe_2).

2D FETs have been plagued by large contact resistance (R_c)[74,78,79,81,83–85]. A theoretical study uncovered that the major contributors include the van der Waals gap, the lack of effective doping and the Fermi-level pinning effect, and that contact metals with d orbitals are preferred[86]. English et al.[87] found that an ultrahigh vacuum was essential for achieving low R_c on MoS_2. Recently, inspired by the idea of in-plane seamless graphene edge contact[88], Yeh et al.[89], demonstrated edge-contacted graphene–2DS–graphene FETs, which have the capability of dynamically modulating (by gate bias) the Schottky barrier at graphene–2DS contacts, and achieved a low R_c of 0.67 kΩ μm for 1L-WS_2.

Converting the semiconducting 2H-MoS_2 into metallic 1T-MoS_2 in the contact area[90] has been found to lower the R_c to 0.24 kΩ μm. The semi-metallic bismuth (Bi) has been found to form an ultralow n-type R_c (0.123 kΩ μm)[91] to 1L-MoS_2 owing to its low DOS near the conduction band edge of MoS_2 that suppresses Fermi-level pinning. Although the values achieved in the above two methods are encouraging, the metastability of 1T-phase MoS_2 and low melting point (271 °C) of Bi present great challenges in practical process integration. Intel[92] and Taiwan Semiconductor Manufacturing Company (TSMC)[93] found another semimetal, antimony (Sb), that has a much higher melting point (631 °C) than that of Bi and could deliver an equally low n-type R_c of 0.145 kΩ μm (ref. 92) to 1L-MoS_2, but with relatively low I_{on}. Recently, Li et al.[94] reported an Sb–MoS_2 R_c of 0.042 kΩ μm, which is close to the quantum limit (~0.0296 kΩ μm at a carrier density of 3×10^{13} cm^{-2}). Efforts for such Sb-MoS_2 contacts to be feasible for industry usage in the future include stability at high temperature (>400 °C), and VLSI compatible doping technique to replace strong back gate biasing that induces large parasitic capacitance.

The ambipolar conduction observed in most 2D FETs arises from the Schottky barrier nature of their source/drain contacts. Although such ambipolarity provides an opportunity for designing certain devices such as reconfigurable FETs for low-cost electronics[95], it introduces large leakage currents in CMOS logic gates, which degrades energy efficiency and/or causes logic operation failure. To suppress the ambipolar current and enable energy-efficient 2D complementary (n-type and p-type) FETs, developing transparent (ohmic) contacts via effective doping is essential. Fang et al.[84,96] introduced surface adsorbates, such as nitrogen dioxide and potassium, that can transfer charge to 2DS, and achieved a degenerate doping level. However, owing to the pristine surface of 2DS, surface adsorbates are usually unstable. In contrast, intercalation doping[97], in which dopants are inserted in between the 2D layers, is more reliable, but limited by the relatively long dopant diffusion time. Therefore, non-traditional approaches should be innovated for doping 2DS.

There have been a few attempts to fabricate ultrashort channel (10-nm-scale gate length (L_g)) 2D FETs[98–102]. Cao et al.[101] employed an ultrathin metallic NW as the top gate as well as a self-aligned mask, instead of relying on electron beam lithography, to define the channel length, thereby demonstrating a top-gated 10-nm-scale 1L-MoS_2 FET. Desai et al.[102] fabricated a similar device, by using an SWNT, instead of

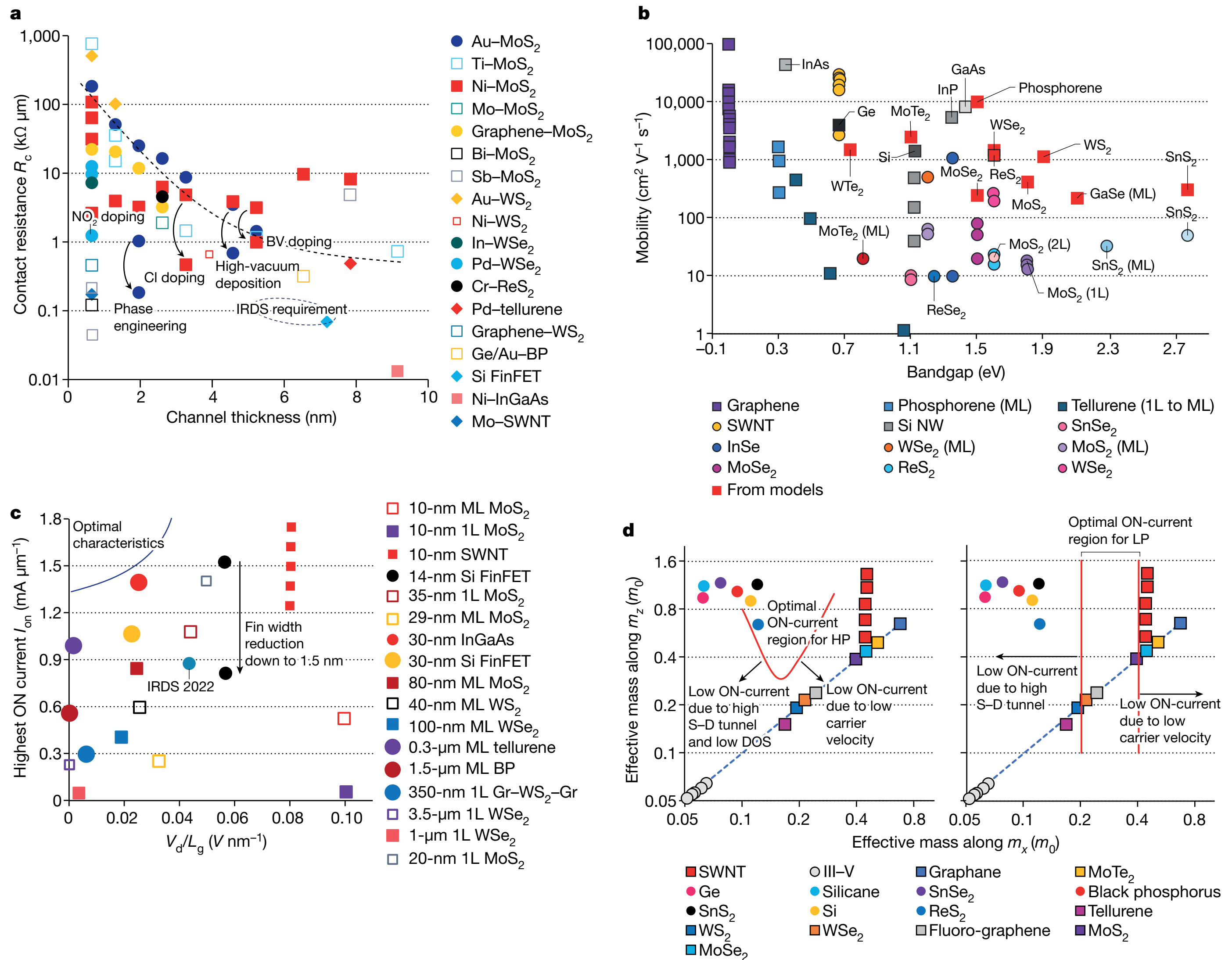


Fig. 5 | Comparative analysis of state-of-the-art and emerging MOSFETs. **a**, Collected contact resistance (in FET ON-state region, that is, at a carrier density level of about 10^{13} cm^{-2}) data from the literature: IRDS[15], InGaAs[28], Si FinFET[44], SWNTs[65] and 2D[81,83–85,89,91,92,94,107–114]. The dashed curve indicates the general trend of increasing contact resistance with reduced channel thickness. BV is benzyl viologen. **b**, Collected mobility data from the literature: SWNTs[59], graphene (Gr)[74,115], MoS_2 (refs. 79,116–118), WS_2 (refs. 106,119–121), SnS_2 (refs. 119,122), ReS_2 (ref. 123), $MoSe_2$ (refs. 119,124), WSe_2 (refs. 83,84,119,125,126), $SnSe_2$ (refs. 119,127), $ReSe_2$ (refs. 128), InSe (ref. 129), BP[130–132], GaSe (ref. 133) and tellurene[107]; ML represents multilayer. All data except those that are 'from models' are experimental results. **c**, Collected highest obtainable ON current from various materials and devices[15,27,45,66,83,94,99,101,134–138]. The dimensions in the legend represent L_g. The IRDS 2022 data corresponds to an HP ON current (for V_d = 0.7 V; L_g = 16 nm; and I_{off} = 10 nA μm^{-1}). For the SWNT I_{on} data points, only the bottom-most data is from experiment, the remaining data are projections based on improving SWNT assembly density. **d**, Projected I_{on} versus effective mass along (m_x) and perpendicular to (m_z) transport direction in units of the free electron mass (m_0) for HP (left) and LP (right) at L_g = 5.9 nm. Si, Ge, III–V, SWNT and some typical 2DS[107,140–142] are used for benchmarking, purely from an effective mass perspective. It is noted that low-band-gap materials (particularly III–V and SWNT) suffer from high leakage current, and hence degrade SS and the attainable ON current for a fixed I_{off} and V_{dd}. The dashed blue line represents the isotropic case, that is, $m_x = m_z$. It is noted that for longer (shorter) L_g with respect to 5.9 nm, the optimal regions will shift to the left (right), accordingly.

an NW, as the back gate. Such a method presents two problems. One is the loss of self-alignment function of gate for source/drain formation, which makes the source-to-drain distance unacceptably large (about 1 μm), and the CGP (Fig. 3a) is still reliant on electron beam lithography. The other is the excess gate voltage consumption in the low-DOS SWNT gate (Fig. 4d) to achieve the required drive current level, which fundamentally limits the voltage scalability and energy efficiency. These two problems could be even more challenging in the recently reported graphene edge-gated FET[103].

The pristine surface of 2DS makes the direct growth of a high-quality, thin gate dielectric on it challenging. An effective seed layer[78] has to be identified to wet the surface of the 2DS. In fact, the best scenario is to develop 2D layered high-*k* dielectrics. The emerging perovskites have been found able to generate high capacitance in a capacitor structure[104]. However, it remains unclear whether thin perovskites in the FET environment can provide the same capacitance, without introducing any hysteresis (memory effect of perovskites) into the current–voltage curves. Recently, Chamlagain et al.[105] reported that 2D insulating tantalum pentoxide (Ta_2O_5), thermally oxidized from tantalum sulfide (TaS_2), has a *k* of about 15.5, and could serve as an effective gate dielectric on MoS_2, which is a positive step towards achieving a 2D high-*k* dielectric.

On the 2DS synthesis front, Kang et al.[106] successfully demonstrated a wafer-scale full-coverage synthesis of 1L-MoS_2 by using metal–organic chemical vapour deposition. Yeh et al.[89] reported an area-selective growth scheme, in which the 2DS is grown at predefined channel

Table 1 | Benchmarking emerging/future MOSFETs

	Si Fin	Si NW/NS	Ge/III–V	SWNT	2DS
Electrostatics	◆◆	◆◆◆	◊◊◊	◆◆◆◆	◆◆◆◆
Source-to-drain tunnelling	◆◆◆	◆◆◆	◆◆	◆◆◆	◆◆◆◆
Manufacturability	◆◆◆◆	◆◆◆	◆◆	◆	◆◆◆
$I_{on,channel}$	◆◆◆	◆◆◆◆	◆◆◆◆	◆◆◆◆	◆◆◆
I_{off}	◆◆◆◊	◆◆◆◆	◆◆	◆◆	◆◆◆◆
Parasitic *R*	◆◆◆	◆◆◆	◆◆◆◆	◆◆◆◆	◆◆◆
Parasitic *C*	◆◆	◆◆◆	◊◊◊	NA	NA
W flexibility	◆	◆◆	◊◊	◆◆◆	◆◆◆◆
Self-heating effect	◆◆◆	◆	◊◊	◆	◆◆
Reliability	◆◆◆	◆◆	◆	◆◆◆◊	◆◆◆◊
Variability	◆◆◆	◆◆◆	◊◊◊	◆◆	◆◆◆◆
CMOS design	◆◆◆◆	◆◆◆◆	◆◆◆	◆	◆◆◆◊

A summary table to quantify (by the number of diamonds) the prospect of various emerging MOSFET technologies (Si FinFET is used as reference) for device scaling. Unfilled diamonds indicate the maximum potential with the aid of other technology. Specifically, the I_{off} of FinFET can be improved with SOI substrate; unfilled diamonds for Ge/III–V represents their maximum potential with fin or NW transistor topology. The reliability of SWNTs and 2DS can be significantly improved with van der Waals-type insulators as the gate dielectric. NA, not available.

regions (smaller than 2DS grain size) in a graphene canvas. This scheme strategically avoids the necessity of wafer-scale, single-crystal 2DS synthesis, and thus provides an alternative route for 2DS towards industrialization.

Emerging technology benchmarking with CMOS

This section provides a comprehensive benchmarking of the several promising transistor technologies discussed above, in terms of contact resistance, mobility, ON-current level and the upper limit of short-channel performance. Although it is unfair to compare emerging technologies (such as 2DS) with almost mature (III–V/Ge and NW and NS) or even commercialized (Si FinFET) technologies, it can provide practical guidelines on how much effort is needed for these technologies to approach or exceed state-of-the-art performance. It is noted that the parasitic capacitance data are not available for emerging transistors (SWNTs and 2D), thus they are not benchmarked here.

Figure 5a shows the R_c data versus channel thickness for various FETs based on different materials (InGaAs[28], Si (in FinFET)[44], SWNT[65] and 2D[81,83–85,89,91,92,94,107–114]), as well as the IRDS requirements. For SWNTs, the Mo-based end-contact approach has reduced the single-tube R_c to 15 kΩ. The mature Si/Ge and III–V devices unsurprisingly have the lowest R_c, and can comfortably meet the IRDS requirements. Recent progress on 2DS contacts has closed the n-type R_c gap of 2D FETs with respect to those of other competing materials (Fig. 5a). However, simultaneously achieving low R_c for both n-type and p-type 2DS FETs remains challenging.

Figure 5b provides a benchmarking on mobility versus bandgap for SWNTs[59], graphene[74,115], MoS_2 (refs. 79,116–118), WS_2 (refs. 106,119–121), SnS_2 (refs. 119,122), ReS_2 (ref. 123), $MoSe_2$ (refs. 119,124), WSe_2 (refs. 83,84,119,125,126), $SnSe_2$ (refs. 119,127), $ReSe_2$ (ref. 128), InSe (ref. 129), BP[130–132], GaSe (ref. 133) and tellurene[107]. A general trend that can be observed is that large-bandgap materials have low mobilities. Among all semiconductors, SWNTs and III–V materials exhibit the highest mobilities, owing to their relatively small bandgaps. Mobility in Si NWs rapidly decreases with reduced thickness. Among all experimentally measured 2DS, only WS_2, WSe_2, tellurene and BP offer high mobilities. It is noted that big gaps still exist between measured and theoretically predicted mobilities for 2DS, indicating plenty of room for further improvements in the material and device quality.

Contact resistance and mobility are indeed two widely employed metrics to quantify device performance. However, it is not rare to find that they are incorrectly measured or calculated, especially when device structure and/or material is prepared in a non-traditional manner[78,79]. The most meaningful metric for device performance is the obtainable highest current level, as collected from the various devices[15,27,45,66,83,94,99,101,134–138] and plotted in Fig. 5c, versus V_d/L_g, which is essentially the average lateral electric field along the channel. The best condition is obviously at the top-left corner, that is, obtaining the highest possible current at the lowest lateral electric field. Compared with SWNTs and other even more mature materials, most 2DS devices, except the Sb-contacted MoS_2 FET, show a lower ON-current level, especially for chemical vapour deposition (CVD) samples, because of their high contact resistance and defect-rich material or imperfect device fabrication.

Rigorous quantum transport simulations[139] have been performed to explore the full benefit of the broad 2DS family[107,119,140–142] for sub-10-nm FETs. Performance maps (Fig. 5d) (see ref. 139 for colour maps) for HP and LP at L_g = 5.9 nm were generated by using effective masses (m_x, along transport direction; m_z, along device width direction) as variables. It is noted that m_x determines the carrier velocity and SDT, whereas m_z serves as a DOS modulator. Materials with an effective mass around $0.3m_0$, such as WSe_2, tellurene and WS_2, are all desirable for LP. Compared with LP, HP allows much higher I_{off} (Fig. 2d), and hence is less sensitive to SDT. Therefore, anisotropic materials, such as $SnSe_2$, BP, ReS_2 and so on, that have small m_x but large m_z, are preferred for HP. It is worthwhile noting that although these two maps are made for 2D FETs, they are also of a certain reference value to other material systems. Therefore, Si, Ge, III–V and SWNTs are also placed in relevant positions within Fig. 5d. A unique advantage of 2DS is that they can cover almost the entire map (see ref. 139 for colour maps), that is, 2DS are capable of a wide range of applications at different technology nodes.

On the basis of the review, analysis and benchmarking above, a summary table (Table 1) to quantify (with the number of diamonds) the prospect of various emerging and future MOSFET technologies can be generated, in terms of several critical metrics including electrostatics, SDT, manufacturability, intrinsic ON current ($I_{on,channel}$), which excludes the effect of contacts, I_{off}, parasitic resistance (parasitic *R*), parasitic capacitance (parasitic *C*), circuit design flexibility in choosing device width (*W* flexibility), self-heating effect, reliability, variability, and the performance balance between n-type and p-type devices (CMOS design). As clearly reflected, the low-dimensional SWNTs and 2DS offer remarkable scalability advantage, with respect to bulk semiconductors. At this stage, the semiconductor industry seems more interested in 2DS[76,92,143], likely owing to the fact that 2DS are much more manufacturable than SWNTs.

The future 'trans-resistors' beyond MOSFETs

MOSFETs are limited by their operation mechanism; therefore, the power consumption and energy efficiency fail to scale at the same pace as the device size. The CMOS community has been searching for 'beyond MOSFET' transistors, to break this energy-efficiency bottleneck. In this context, it is instructive to review the fundamental aspects of a transistor—'trans' (arising from 'transfer' or modulation) and 'resistor' (arising from resistance of a channel). Thus, 'trans' captures the approach to manipulate the information state or carrier—which is the 'resistor'. For commercialized MOSFETs and BJTs, 'trans' is realized with the electric-field effect through a static gate capacitor and a p–n junction barrier modulation, respectively, whereas the 'resistor' is implemented in the form of thermionic emission over a barrier for both devices. There have been plenty of efforts to introduce innovations into one (or both) of these two aspects.

The various 'trans' approaches, negative capacitance (NC)[144], suspended-gate (SG)[145] and Mott-phase-change-material gated

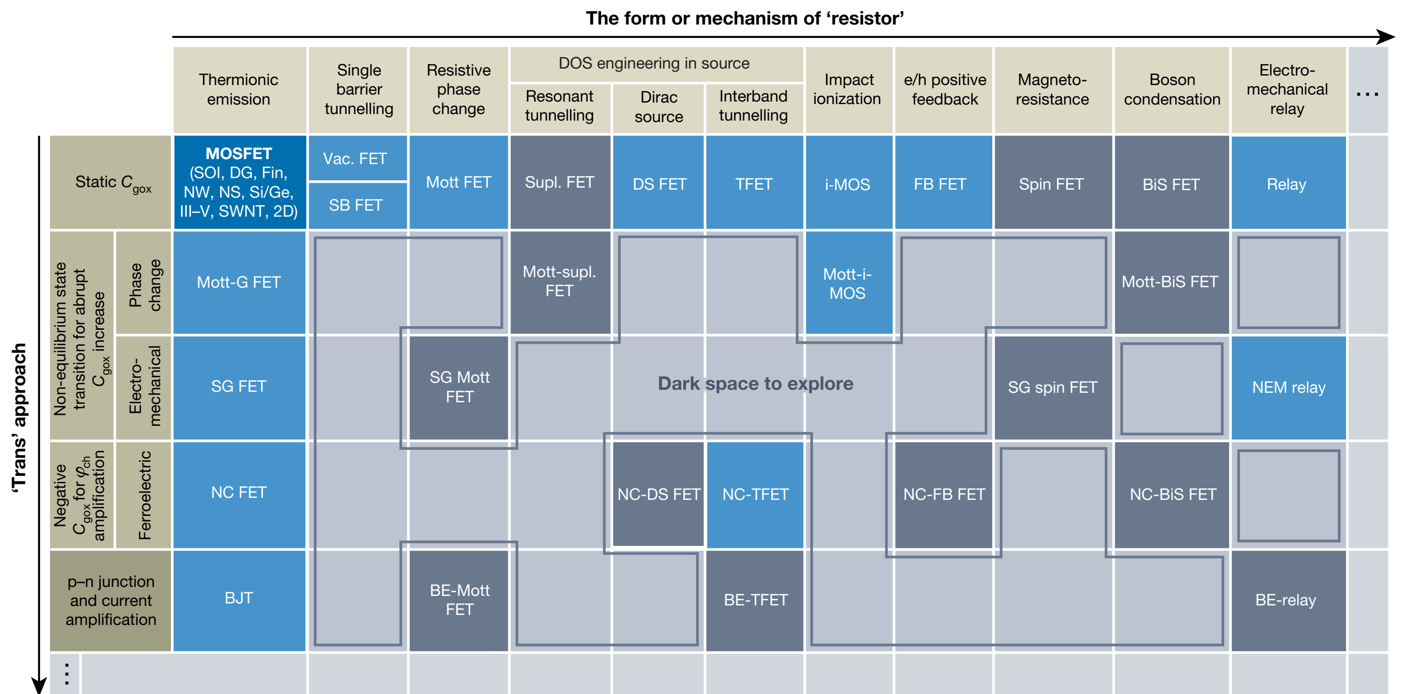


Fig. 6 | Transistors go beyond MOSFETs. Revisiting the two aspects of transistor—'trans' and 'resistor' for exploring the design space of future FETs. An NC FET is a negative capacitance FET[144]; an SG FET is a suspended-gate FET[145]; a Mott-G FET is phase-change-material gated FET[146]; an SB FET is a Schottky barrier FET; a vac. FET is a vacuum-channel FET[147]; a Mott FET employs a phase-change material as the channel or to connect in series to the source[148]; a Supl. FET is superlattice FET[149]; a DS FET is a Dirac source FET[150]; a TFET is an interband or BTBT FET[151,152]; a BE-TFET is a bipolar-enhanced TFET[153]; an NC-TFET is an NC-gated TFET[154]; an i-MOS is an impact-ionization FET[155]; a Mott-i-MOS is a Mott-gated i-MOS FET; an FB FET is charge feedback FET[156]; a BiS FET is bilayer pseudospin FET[159]; an NEM relay is a nanoelectromechanical relay-type FET[160]. Dark grey indicates device concepts (Supl., BiS, Spin[157,158] FETs, BE-TFETs, etc) that have not been experimentally proven. Among all the experimentally demonstrated (light blue) beyond-MOSFET transistors, TFETs and NC FETs are the most promising low-SS devices, and have received the most extensive studies. A more detailed version of this figure (Supplementary Fig. 1) with band diagrams and charge carrier illustrations has been provided in Supplementary Section 3 for interested readers. Note that a dark blue colour is used for the MOSFET to highlight that it is the benchmark device. The light grey colour indicates 'dark space' or unexplored device concepts. The listed 'trans' and 'resistor' options are not exhaustive, and hence, '…' is used along the two axes to indicate opportunities for more innovations.

(Mott-G)[146] FETs (Fig. 6), were proposed and have been experimentally explored. NC FETs utilize the negative capacitance state of ferroelectric materials to construct a negative C_{gox}, targeting to overcome the unity upper limit of gate efficiency. SG and Mott-G FETs introduce a nanoelectromechanical (NEM) switch and an insulator-to-metal phase-change memory cell, respectively, into the gate stack, and utilize their non-equilibrium-state switching transients to realize an abrupt (with respect to the gate voltage) C_{gox} increase, which translates to an abrupt increase of charge density and drain current, that is, an ultrasmall SS. It is worth mentioning that the memory nature of the NEM switch and phase-change memory cell inevitably result in a memory effect, that is, hysteresis, in the transfer characteristics, which limits SG and Mott-G FETs to memory applications.

In contrast, the 'resistor' has been implemented in many beyond-thermionic-emission forms and mechanisms. In Schottky barrier and vacuum[147] FETs, charge carriers tunnel through the Schottky barrier between the metallic source and the semiconducting or vacuum channel, respectively. Mott FETs[148] employ a phase-change memory cell connected in series with the source, to achieve an abrupt change of resistance of the 'resistor' and hence the drain current, during the memory-state switching, which also introduces hysteresis in the I–V curve. Superlattice FETs[149] employ a multi-quantum well in the source region to form an artificial resonant tunnelling band, which is narrow enough to filter the thermionic emission of high-energy (with respect to the Fermi level) carriers. Dirac source FETs[150] utilize the decreasing DOS towards the Dirac point of a graphene source to realize reduced high-energy carrier injection. The minimum achievable SS of this device remains larger than 30 mV dec^{-1}, simply owing to the fact that graphene is gapless, that is, the DOS near the Dirac point, although low, still allows high-energy carrier leakage. Tunnelling FETs (TFETs)[151,152] utilize the bandgap of the source to filter the thermionic emission of high-energy carriers. However, the low band-to-band tunnelling probability sets a constraint on its ON current. Bipolar-enhanced TFETs (BE-TFETs)[153] have been proposed to alleviate this issue, by amplifying the drain current with the large current gain of the BJT. Experimental demonstration is needed to prove this device concept. A previous study[154] employed a ferroelectric gate (expected to function as an NC) on a TFET, and found that the TFET performance was enhanced. Whether such an enhancement is due to the NC effect or simply due to the high k of the ferroelectric material needs to be examined. Essentially, superlattice FETs, Dirac source FETs and TFETs have a similar form of 'resistor', as they are all based on DOS engineering in the source to achieve localized carrier injection near the Fermi level. Impact-ionization MOSFETs (i-MOS)[155] employ superexponential generation (versus gate voltage) of electron–hole pairs during impact ionization to realize a superexponential increase of the drain current. However, a large V_d is required to activate the impact-ionization process, and hence cannot help reduce supply voltage. In feedback FETs[156], electron and hole potential barriers are intentionally introduced near the n-type source and p-type drain sides, respectively, to form a positive feedback loop between electron–hole redistribution and barrier modulation, which dynamically accelerates the electron–hole injection rate. This device also requires a large V_d to activate the feedback loop.

Compared with the abovementioned charge-carrier-based FETs, spin FETs[157,158] employ spin as the information carrier, and implement the 'resistor' in the form of magnetoresistance. It is worthwhile noting

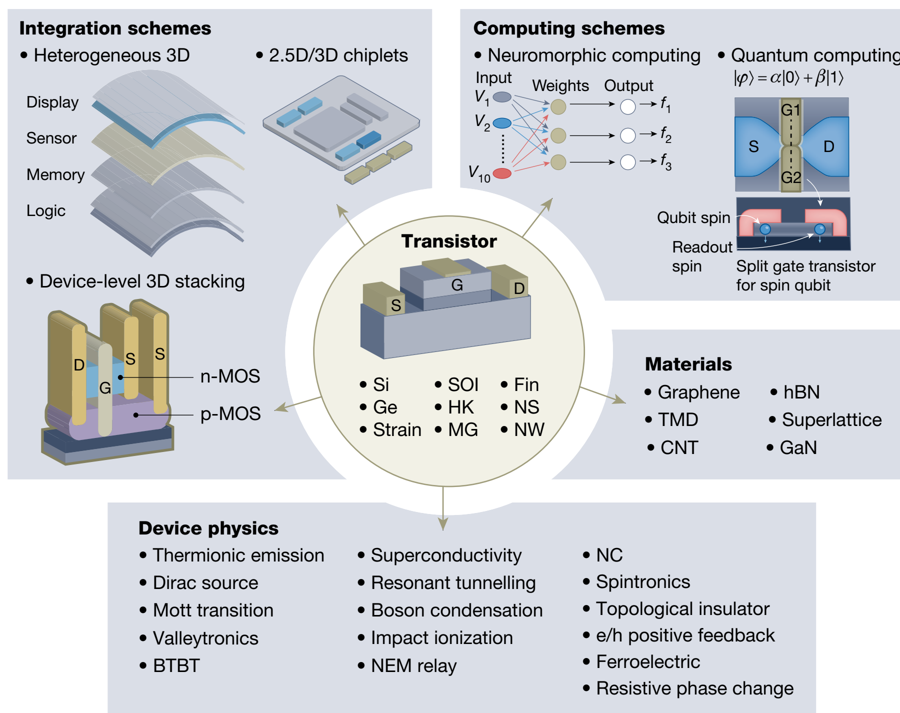


Fig. 7 | A transistor is the powerful engine that is driving all aspects of the information technology industry. Transistors have been igniting and driving forward the research and development of information technologies, in terms of integration schemes—including heterogeneous 3D stacking by means of either TSV or m-3D integration; 2.5D (or 3D[198]) chiplets that allow ultraclose packaging of dies; and device-level 3D stacking (n-type device on top of p-type device or vice versa), as well as computing schemes, device physics, and materials. The centre of the figure shows various transistor topologies and materials that have already been adopted by the industry. The schematic on the right of 'Computing schemes' box shows a spin qubit directly constructed on a FET with split gate (bottom) that allows a quantum computer to benefit from CMOS scaling. The conceivable interactions open up a multitude of exciting possibilities for studying novel physics and designing novel structures and functionalities. For example, TMDs from 'Materials' can be employed to build NS FETs for sub-1-nm CMOS technologies; graphene and TMDs can be useful for studying 'Device physics' such as spintronics and valleytronics, or qubits can be designed with TMDs; BTBT devices and circuits can be used for neuromorphic computing; 1D and 2D materials can be utilized for monolithic heterogeneous 3D integration and 3D chiplet architectures with advanced packaging and connectivity, power delivery, thermal management, and so on. HK, high-k; hBN, hexagonal boron nitride. V_{1-10} and f_{1-3} are the input and output signals of a neural network, respectively. $|\varphi\rangle$ refers to the superimposed qubit state, consisting of qubit basis states $|0\rangle$ and $|1\rangle$, with probability amplitudes of α and β, respectively. MG, TMD, and CNT represent metal gate, transitional metal dichacogenide, and carbon nanotube, respectively.

that spin FETs here refer to Rashba type[158] in which the spin direction is manipulated by a small gate electric field during transport from the source to the drain. Unfortunately, the controllable manipulation of spin direction in experiments is difficult, which has thus far prevented an experimental proof of spin FETs. In bilayer pseudospin (BiS) FETs[159], it was predicted that the condensation of electron–hole pairs (bosons) in the insulator separating the electron–hole bilayer could markedly reduce the interlayer resistance, thereby realizing an abrupt increase in conductance. However, experimental demonstrations are required to validate the device concept of BiS FETs. NEM relay[160], compared with SG FETs, employ a NEM switch at the channel/drain junction, to realize a physical contact and separation between channel and drain. Such mechanical switches suffer heavily from reliability issues arising from stiction, particularly for ultrascaled devices[161]. As reflected in Fig. 6, there remains a large dark space for LP device designers to explore. It is worth noting that to implement those futuristic transistors, judicious selection of material platforms could play a critical role. For example, the pristine surface of 2D materials can be exploited to develop ultrasteep-slope TFETs[162], which are difficult to realize with traditional bulk materials.

Among all the experimentally demonstrated novel transistors, TFETs[151] and NC FETs[144] have received the most extensive studies (see Supplementary Section 4). The TFET structure with BTBT was introduced in 1978[151]. In 2004, it was found that sub-60 SS can appear in the BTBT leakage current branch of an SWNT FET structure, demonstrating that sub-60 SS is experimentally achievable[152]. TFETs are essentially gated p–i–n diodes as schematically illustrated in the top half of Supplementary Fig. 2a. Ultralow-SS TFETs have been achieved in many different material systems and device structures, as shown by the collected experimental data[163–168] in Supplementary Fig. 2b. At present, the main challenge for TFETs is the low BTBT probability that limits I_{on}. Innovative structural designs such as using a 2D heterojunction[162] and a gated Esaki diode structure[169] as well as experimental demonstrations are required to overcome this issue.

Since the concept of NC FETs was proposed[144] in 2007–2008, it quickly rose to a notable position, owing to its fabrication-friendly structure (bottom half of Supplementary Fig. 2a), which is basically a MOSFET with an additional NC (typically ferroelectric material, FE) layer inserted between the oxide and the gate. Compared with the FE memory application[170,171] in which the bistable states of the FE was used to store binary information, NC FETs are believed to utilize the metastable state of the FE layer, which provides a negative polarization response to an external electric field, that is, NC, to better (<1) the unity gate efficiency of MOSFETs (equation (1)). Thus far, there have been many claimed 'NC FETs', showing low SS, with relatively small hysteresis in measured d.c. I–V curves[172–177], as shown in Supplementary Fig. 2c. However, according to the fundamental device physics of NC FETs, the design space of NC FETs for hysteresis-free sub-60 SS has recently been found[178] to be very small because of the generally large quantum capacitance, and the excellent electrostatics of modern FETs. In other

words, these measured small SS values cannot be explained by the static NC interpretation[144]. In fact, more and more studies[179–181] have indicated that they can be attributed to the transient effects during the measurement and/or FE polarization dynamics. Recently, a study[182] found an ultrahigh trap density (level of 10^{14} cm^{-2}) between the FE layer and the gate oxide layer. Thus, the trapping and de-trapping dynamics (not controllable) may present an alternative explanation of the steep slopes in the reported 'NC FETs'. Moreover, high trap density in an electron device is generally accompanied by reliability issues. Thus, considerable research efforts are needed to address these issues and thoroughly understand the interplay between the FE and the oxide, before applying NC in CMOS products.

Beyond-Moore integration pathways

In the CMOS scaling history, most efforts have been invested in shrinking the MOSFET feature size, and optimizing the utilization of in-plane area of the logic chip, including the '2.5D' chiplets[183]—a packaging technique, following Moore's law. It is not unfathomable to find out someday that the MOSFET physical size shrinking and the in-plane increase in device integration density have stopped, owing to the fabrication difficulty and cost, as well as power-density constraints. In this regard, the vertical physical space of the logic chip has not been well exploited. Although FinFETs and NW and NS FETs can be considered as one type of such effort, they are only beneficial for improving device current and electrostatics, and are not helpful in increasing the device integration density beyond a certain point. Three-dimensional integration that stacks either devices (usually in a complementary manner)[184] or dies in the vertical space is considered a practical approach to increase device density. Besides the benefit of density scaling, 3D integrated circuits can significantly reduce interconnect delay and power dissipation[185]. Three-dimensional integration can be realized via different pathways, including wire bonding and/or flip-chip based 3D packaging, through-Si-via (TSV)-based 3D die/wafer stacking, and monolithic 3D integration (m-3D). Among the three, m-3D is the most desirable approach, owing to its large layer density and high local connectivity-enabled design possibilities. Although m-3D has thermal budget and heat dissipation issues, progress has been made in conquering these challenges, such as employing solid-phase epitaxy regrowth to activate dopants below 600 °C (ref. 186), cooling 3D integrated circuits with power delivery networks[187] and employing a high-thermal-conductivity hexagonal boron nitride interlayer dielectric[188]. Three-dimensional integration is an inclusive technology. Eventually, it can evolve to the form of heterogeneous 3D integration[169,185,188–191], in which very dissimilar systems of different materials (such as 2DS and Si[188]), devices and functionalities can be integrated together along both vertical and lateral directions, thereby constructing an ultrapowerful and energy-efficient system-on-chip and heterogeneous system-of-chips or chiplets (Fig. 7). Nonetheless, it is important that device and circuit architectures, such as the recently demonstrated 0.5T0.5R hybrid resistive random-access memory[192]—that not only allow computing paradigms such as 'in-memory' computing but also are specifically designed to simultaneously provide higher lateral and vertical stacking densities—remain a central priority for revolutionary advancements in 3D integrated-circuit design.

Beyond von Neumann computing

Information technology has never stopped evolving. Beside the classical computers designed for general-purpose high-performance computation, neuromorphic computing[193] and quantum computing[194] are being actively developed, and expected to usher unprecedented advantages in some domains, such as chemical reaction simulations, or artificial intelligence and machine learning, which are too computationally intensive or power consuming for a conventional von Neumann computer. A common requirement from all of them is a transistor of ultrahigh compactness and ultralow leakage, as well as high current drivability, robustness and energy efficiency, either as core or supporting or interface elements[193–197], as indicated in Fig. 7. Given these desirable characteristics, it is reasonable to argue that transistor scaling and evolution will never stall, and Moore's law will stay with us for a long time in the foreseeable future.

1. Fleming, J. Instrument for converting alternating electric currents into continuous currents. US patent 803684A (1905).
2. Bardeen, J. & Brattain, W. The transistor, a semi-conductor triode. *Phys. Rev.* **74**, 230–231 (1948).
   Demonstration of a solid-state transistor.
3. Lilienfeld, J. E. Method and apparatus for controlling electric currents. US patent 1,745,175 (1930).
   The original idea of a FET.
4. Atalla, M. M. et al. Stabilization of silicon surfaces by thermally grown oxides. *Bell Syst. Tech. J.* **38**, 749–783 (1959).
   The key enabling innovation responsible for the rise of MOSFETs.
5. Kahng, D. Electric field controlled semiconductor device. US patent 3,102,230 (1963).
6. Auth, C. et al. A 10 nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, self-aligned quad patterning, contact over active gate and cobalt local interconnects. In *IEEE International Electron Devices Meeting* 673–676 (IEEE, 2017).
7. Dennard, R. et al. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid State Circuits* **9**, 256–268 (1974).
8. Mistry, K. et al. A 2.0V, 0.35 μm partially depleted SOI-CMOS technology. In *IEEE International Electron Devices Meeting* 583–586 (IEEE, 1997).
9. Tenbroek, B. et al. Self-heating effects in SOI MOSFETs and their measurement by small signal conductance techniques. *IEEE Trans. Electron Devices* **43**, 2240–2248 (1996).
10. Ghani, T. et al. A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors. In *IEEE International Electron Devices Meeting* 978–980 (IEEE, 2003).
    Commercialization of strained-silicon technology.
11. Mistry, K. et al. A 45 nm logic technology with high-*k* + metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging. In *IEEE International Electron Devices Meeting* 247–250 (IEEE, 2007).
    Commercialization of high-*k* + metal-gate CMOS technology.
12. Auth, C. et al. A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors. In *IEEE VLSI Technology Symposium* 131–132 (IEEE, 2012).
    Commercialization of 3D FinFETs.
13. Bohr, M., Chau, R., Ghani, T. & Mistry, K. The high-*k* solution. *IEEE Spectr.* **44**, 29–35 (2007).
14. Kuhn, K. J. Considerations for ultimate CMOS scaling. *IEEE Trans. Electron Devices* **59**, 1813–1828 (2012).
15. More Moore table. (page 13, Table MM9) *International Roadmap for Devices and Systems* IRDS 2022 More Moore (ieee.org) (2022).
16. Luryi, S. Quantum capacitance devices. *Appl. Phys. Lett.* **52**, 501–503 (1988).
17. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* 3rd edn (Wiley-Interscience, 2007).
18. Asenov, A. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: a 3-D atomistic simulation study. *IEEE Trans. Electron Devices* **45**, 2505–2513 (1998).
19. Dadgour, H., Endo, K., De, V. & Banerjee, K. Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability. In *IEEE International Electron Devices Meeting* 705–708 (IEEE, 2008).
20. Kuhn, K. et al. Process technology variation. *IEEE Trans. Electron Devices* **58**, 2197–2208 (2011).
21. Grasser, T. et al. NBTI in nanoscale MOSFETs—the ultimate modeling benchmark. *IEEE Trans. Electron Devices* **61**, 3586–3593 (2014).
22. Cartier, E. et al. Fundamental aspects of HfO_2-based high-*k* metal gate stack reliability and implications on t_{inv}-scaling. In *IEEE International Electron Devices Meeting* 441–444 (IEEE, 2011).
23. Yu, B., Wann, C., Nowak, E. D., Noda, K. & Hu, C. Short-channel effect improved by lateral channel-engineering in deep-submicronmeter MOSFETs. *IEEE Trans. Electron Devices* **44**, 627–634 (1997).
24. Welser, J., Hoyt, J. L. & Gibbons, J. F. NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon–germanium structures. In *IEEE International Electron Devices Meeting* 1000–1002 (IEEE, 1992).
25. Ota, K. et al. Novel locally strained channel technique for high performance 55 nm CMOS. In *IEEE International Electron Devices Meeting* 27–30 (IEEE, 2002).
26. Yeo, Y., Lu, Q., King, T-J. & Hu, C. Enhanced performance in sub-100 nm CMOSFETs using strained epitaxial silicon-germanium. In *IEEE International Electron Devices Meeting* 753–756 (IEEE, 2000).
27. Takagi, S. et al. III–V/Ge channel MOS device technologies in nano CMOS era. *Jpn J. Appl. Phys.* **54**, 06FA01 (2015).
28. del Alamo, J. A. et al. Nanometer-scale III–V MOSFETs. *IEEE J. Electon Devices Soc.* **4**, 205–214 (2016).
29. Yeap, G. et al. 5nm CMOS production technology platform featuring full-fledged EUV and high-mobility channel FinFETs with densest 0.021 μm^2 SRAM cells for mobile SoC and high-performance computing applications. In *IEEE International Electron Devices Meeting* 879–882 (IEEE, 2019).

30. Skotnicki, T. & Boeuf, F. How can high mobility channel materials boost or degrade performance in advanced CMOS. In *IEEE VLSI Technology Symposium* 153–154 (IEEE, 2010).
31. Jin, D., Kim, D., Kim, T. & del Alamo, J. A. Quantum capacitance in scaled down III–V FETs. In *IEEE International Electron Devices Meeting* 495–498 (IEEE, 2009).
32. Koba, S. et al. Channel length scaling limits of III–V channel MOSFETs governed by source–drain direct tunneling. *Jpn J. Appl. Phys.* **53**, 04EC10 (2014).
33. Zheng, Z. et al. Gallium nitride-based complementary logic integrated circuits. *Nat. Electron.* **4**, 595–603 (2021).
34. Yan, R., Ourmazd, A. & Lee, K. Scaling the Si MOSFET: from bulk to SOI to bulk. *IEEE Trans. Electron Devices* **39**, 1704–1710 (1992).
Provides a simple but important MOSFET scaling guideline.
35. Frank, D., Taur, Y. & Wong, H. Generalized scale length for two-dimensional effects in MOSFET's. *IEEE Electron Device Lett.* **19**, 385–387 (1998).
36. Ando, T. Ultimate scaling of high-*k* gate dielectrics: higher-*κ* or interfacial layer scavenging? *Materials* **5**, 478–500 (2012).
37. Robertson, J. & Wallace, R. M. High-*k* materials and metal gates for CMOS applications. *Mater. Sci. Eng. R* **88**, 1–41 (2014).
38. Hayashi, Y. Gate insulator type field effect transistor. Japanese patent JP1791730 (filed 24 June 1980) (1993).
Proposal for a 3D transistor.
39. Hisamoto, D., Kaga, T., Kawamoto, Y. & Takeda, E. A fully depleted lean-channel transistor (DELTA)—a novel vertical ultrathin SOI MOSFET. In *IEEE International Electron Devices Meeting* 833–836 (IEEE, 1989).
40. Hisamoto, D. et al. A folded-channel MOSFET for deep-sub-tenth micron era. In *IEEE International Electron Devices Meeting* 1032–1034 (IEEE, 1998).
A prototype demonstration of an ultrathin-body 3D FinFET.
41. Huang, X. et al. Sub 50-nm FinFET: PMOS. In *IEEE International Electron Devices Meeting* 67–70 (IEEE, 1999).
42. Rasouli, S., Dadgour, H., Endo, K., Koike, H. & Banerjee, K. Design optimization of FinFET domino logic considering the width quantization property. *IEEE Trans. Electon Devices* **57**, 2934–2943 (2010).
43. Kawasaki, H. Challenges and solutions of FinFET integration in an SRAM cell and a logic circuit for 22 nm node and beyond. In *IEEE International Electron Devices Meeting* 289–292 (IEEE, 2009).
44. Niimi, H. et al. Sub-10^{-9} Ω-cm^2 n-type contact resistivity for FinFET technology. *IEEE Electron Device Lett.* **37**, 1371–1374 (2016).
45. He, X. et al. Impact of aggressive fin width scaling on FinFET device characteristics. In *IEEE International Electron Devices Meeting* 493–496 (IEEE, 2017).
46. Appenzeller, J. et al. Toward nanowire electronics. *IEEE Trans. Electron Devices* **55**, 2827–2845 (2008).
47. Mizutani, T. et al. Threshold voltage and current variability of extremely narrow silicon nanowire MOSFETs with width down to 2 nm. In *Silicon Nanoelectronics Workshop* 1–2 (Publishers are Japan Society of Applied Physics, IEEE, 2015).
48. Mertens, H. et al. Gate-all-around MOSFETs based on vertically stacked horizontal Si nanowires in a replacement metal gate process on bulk Si substrates. In *IEEE VLSI Technology Symposium* 1–2 (IEEE, 2016).
49. Yang, B. et al. Vertical silicon-nanowire formation and gate-all-around MOSFET. *IEEE Electron Device Lett.* **29**, 791–794 (2008).
50. Loubet, N. et al. Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET. In *IEEE VLSI Technology Symposium* 230–231 (IEEE, 2017).
51. Jagannathan, H. et al. Vertical-transport nanosheet technology for CMOS scaling beyond lateral-transport devices. In *IEEE International Electron Devices Meeting* 557–560 (IEEE, 2021).
52. Weckx, P. et al. Novel forksheet device architecture as ultimate logic scaling device towards 2nm. In *IEEE International Electron Devices Meeting* 36.5.1–36.5.4 (IEEE, 2019).
53. Wang, R. et al. Experimental study on quasi-ballistic transport in silicon nanowire transistors and the impact of self-heating effects. In *IEEE International Electron Devices Meeting* 1.4.1–1.4.4 (IEEE, 2008).
54. Zhuge, J. et al. Experimental investigation and design optimization guidelines of characteristic variability in silicon nanowire CMOS technology. In *IEEE International Electron Devices Meeting* 61–64 (IEEE, 2009).
55. Iijima, S. & Ichihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603–605 (1993).
56. Novoselov, K. S. et al. Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).
57. Ajayan, P., Kim, P. & Banerjee, K. Two-dimensional van der Waals materials. *Phys. Today* **69**, 38–44 (2016).
58. Tans, S., Verschueren, A. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
Demonstration of a carbon nanotube transistor.
59. Javey, A. et al. High-*k* dielectrics for advanced carbon nanotube transistors and logic gates. *Nat. Mater.* **1**, 241–246 (2002).
60. Wang, Z. et al. Growth and performance of yttrium oxide as an ideal high-*k* gate dielectric for carbon-based electronics. *Nano Lett.* **10**, 2024–2030 (2010).
61. Javey, A., Guo, J., Wang, Q., Lundstrom, M. & Dai, H. Ballistic carbon nanotube field-effect transistors. *Nature* **424**, 654–657 (2003).
62. Franklin, A. & Chen, Z. Length scaling of carbon nanotube transistors. *Nat. Nanotechnol.* **5**, 858–862 (2010).
63. Qiu, C. et al. Scaling carbon nanotube complementary transistors to 5-nm gate length. *Science* **355**, 271–276 (2017).
64. Shabrawy, K., Maharatna, K., Bagnall, D. & Hashimi, B. Modeling SWCNT bandgap and effective mass variation using Monte Carlo approach. *IEEE Trans. Nanotechnol.* **9**, 184–193 (2010).
65. Cao, Q. et al. End-bonded contacts for carbon nanotube transistors with low, size-independent resistance. *Science* **350**, 68–72 (2015).
66. Cao, Q., Tersoff, J., Farmer, D., Zhu, Y. & Han, S. Carbon nanotube transistors scaled to a 40-nanometer footprint. *Science* **356**, 1369–1372 (2017).
Demonstration of carbon nanotube FETs outperforming state-of-the-art Si MOSFETs.
67. Jin, S. et al. Using nanoscale thermocapillary flows to create arrays of purely semiconducting single-walled carbon nanotubes. *Nat. Nanotechnol.* **8**, 347–355 (2013).
68. Shulaker, M. et al. Carbon nanotube computer. *Nature* **501**, 526–530 (2013).
69. Ghosh, S., Bachilo, S. & Weisman, R. Advanced sorting of single-walled carbon nanotubes by nonlinear density-gradient ultracentrifugation. *Nat. Nanotechnol.* **5**, 443–450 (2010).
70. Cao, Q. et al. Arrays of single-walled carbon nanotubes with full surface coverage for high-performance electronics. *Nat. Nanotechnol.* **8**, 180–186 (2013).
71. Cao, Q., Han, S. & Tulevski, G. Fringing-field dielectrophoretic assembly of ultrahigh-density semiconducting nanotube arrays with a self-limited pitch. *Nat. Commun.* **5**, 5071 (2014).
72. Liu, L. et al. Aligned, high-density semiconducting carbon nanotube arrays for high-performance electronics. *Science* **368**, 850–856 (2020).
73. Islam, A. Variability and reliability of single-walled carbon nanotube field effect transistors. *Electronics* **2**, 332–367 (2013).
74. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
75. Iannaccone, G., Bonaccorso, F., Colombo, L. & Fiori, G. Quantum engineering of transistors based on 2D materials heterostructures. *Nat. Nanotechnol.* **13**, 138–191 (2018).
76. Li, M., Su, S., Wong, H. & Li, L. How 2D semiconductors could extend Moore's law? *Nature* **567**, 169–170 (2019).
77. Wang, H. et al. Graphene nanoribbons for quantum electronics. *Nature Rev. Phys.* **3**, 791–802 (2021).
78. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS_2 transistors. *Nat. Nanotechnol.* **6**, 147–150 (2011).
Demonstration of a top-gated monolayer 2D semiconductor FET.
79. Fuhrer, M. & Hone, J. Measurement of mobility in dual-gated MoS_2 transistors. *Nat. Nanotechnol.* **8**, 146–147 (2013).
80. Kaasbjerg, K., Thygesen, K. & Jacobsen, K. Phonon-limited mobility in n-type single-layer MoS_2 from first principles. *Phys. Rev. B* **85**, 115317 (2012).
81. Liu, W. et al. High-performance few-layer-MoS_2 field-effect-transistor with record low contact-resistance. In *IEEE International Electron Devices Meeting* 499–502 (IEEE, 2013).
82. Jena, D. & Konar, A. Enhancement of carrier mobility in semiconductor nanostructures by dielectric engineering. *Phys. Rev. Lett.* **98**, 136805 (2007).
83. Liu, W. et al. Role of metal contacts in designing high-performance monolayer n-type WSe_2 field effect transistors. *Nano Lett.* **13**, 1983–1990 (2013).
Demonstration of high-performance monolayer 2D-TMD FET.
84. Fang, H. et al. High-performance single layered WSe_2 p-FETs with chemically doped contacts. *Nano Lett.* **12**, 3788–3792 (2012).
85. Das, S., Chen, H., Penumatcha, A. & Appenzeller, J. High performance multilayer MoS_2 transistors with scandium contacts. *Nano Lett.* **13**, 100–105 (2012).
86. Kang, J., Liu, W., Sarkar, D., Jena, D. & Banerjee, K. Computational study of metal contacts to monolayer transition-metal dichalcogenide semiconductors. *Phys. Rev. X* **4**, 031005 (2014).
Comprehensive ab initio analysis of metal contacts to 2D semiconductors.
87. English, C., Shine, G., Dorgan, V., Saraswat, K. & Pop, E. Improved contacts to MoS_2 transistors by ultra-high vacuum metal deposition. *Nano Lett.* **16**, 3824–3830 (2016).
88. Kang, J., Sarkar, D., Khatami, Y. & Banerjee, K. Proposal for all-graphene monolithic logic circuits. *Appl. Phy. Lett.* **103**, 083113 (2013).
89. Yeh, C., Cao, W., Pal, A., Parto, K. & Banerjee, K. Area-selective-CVD technology enabled top-gated and scalable 2D-heterojunction transistors with dynamically tunable Schottky barrier. In *IEEE International Electron Devices Meeting* 23.4.1–23.4.4 (IEEE, 2019).
Demonstration of single-crystal 2D semiconductor growth in pre-designed sites enabling high-performance FETs.
90. Kappera, R. et al. Phase-engineered low-resistance contacts for ultrathin MoS_2 transistors. *Nat. Mater.* **13**, 1128–1134 (2014).
91. Shen, P. et al. Ultralow contact resistance between semimetal and monolayer semiconductors. *Nature* **593**, 211–217 (2021).
92. O'Brien, K. et al. Advancing 2D monolayer CMOS through contact, channel and interface engineering. In *IEEE International Electron Devices Meeting* 146–149 (IEEE, 2021).
93. Chou, A. et al. Antimony semimetal contact with enhanced thermal stability for high performance 2D electronics. In *IEEE International Electron Devices Meeting* 150–153 (IEEE, 2021)
94. Li, W. et al. Approaching the quantum limit in two-dimensional semiconductor contacts. *Nature* **613**, 274–279 (2023).
95. Cao, W., Chu, J.-H., Parto, K. & Banerjee, K. A mode-balanced reconfigurable logic gate built in a van der Waals strata. *npj 2D Mater. Appl.* **5**, 20 (2021).
96. Fang, H. et al. Degenerate n-doping of few-layer transition metal dichalcogenides by potassium. *Nano Lett.* **13**, 1991–1995 (2013).
97. Kang, J. et al. On-chip intercalated-graphene inductors for next-generation radio frequency electronics. *Nat. Electron.* **1**, 46–51 (2018).
98. Smets, Q. et al. Ultra-scaled MOCVD MoS_2 MOSFETs with 42 nm contact pitch and 250 μA/μm drain current. In *IEEE International Electron Devices Meeting* 23.4.1–23.4.4 (IEEE, 2019).
99. Yang, L., Lee, R., Rao, S., Tsai, W. & Ye, P. 10 nm nominal channel length MoS_2 FETs with EOT 2.5 nm and 0.52 mA/μm drain current. *Device Research Conference* 237–238 (IEEE, 2015).
100. Li, K. et al. MoS_2 U-shape MOSFET with 10 nm channel length and poly-Si source/drain serving as seed for full wafer CVD MoS_2 availability. In *IEEE Symposium on VLSI-TSA* 52–53 (IEEE, 2016).
101. Cao, W., Liu, W. & Banerjee, K. Prospects of ultra-thin nanowire gated 2D-FETs for next-generation CMOS technology. In *IEEE International Electron Devices Meeting* 14.7.1–14.7.4 (IEEE, 2016).
102. Desai, S. et al. MoS_2 transistor with 1-nanometer gate lengths. *Science* **354**, 99–102 (2016).

103. Wu, F. et al. Vertical MoS_2 transistors with sub-1-nm gate lengths. *Nature* **603**, 259–264 (2022).
104. Osada, M. & Sasaki, T. Two-dimensional dielectric nanosheets: novel nanoelectronics from nanocrystal building blocks. *Adv. Mater.* **24**, 210–228 (2012).
105. Chamlagain, B. et al. Thermally oxidized 2D TaS_2 as a high-κ gate dielectric for MoS_2 field-effect transistors. *2D Mater.* **4**, 031002 (2017).
106. Kang, K. et al. High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity. *Nature* **520**, 656–660 (2015).
Wafer-scale (4 inch) growth of 2D semiconductor.
107. Wang, Y. et al. Field-effect transistors made from solution-grown two-dimensional tellurene. *Nat. Electron.* **1**, 228–236 (2018).
108. Kang, J., Liu, W. & Banerjee, K. High performance MoS_2 transistors with low resistance molybdenum contacts. *Appl. Phys. Lett.* **104**, 093106 (2014).
109. Guo, Y. et al. Study on the resistance distribution at the contact between molybdenum disulfide and metals. *ACS Nano* **8**, 7771–7779 (2014).
110. Baugher, B., Churchill, H., Yang, Y. & Jarillo-Herrero, P. Intrinsic electronic transport properties of high-quality monolayer and bilayer MoS_2. *Nano Lett.* **13**, 4212–4216 (2013).
111. Kiriya, D., Tosun, M., Zhao, P., Kang, J. & Javey, A. Air-stable surface charge transfer doping of MoS_2 by benzyl viologen. *J. Am. Chem. Soc.* **136**, 7853–7856 (2014).
112. Cui, X. et al. Multi-terminal transport measurements of MoS_2 using a van der Waals heterostructure device platform. *Nat. Nanotechnol.* **10**, 534–540 (2015).
113. Li, S. et al. Thickness scaling effect on interfacial barrier and electrical contact to two-dimensional MoS_2 layers. *ACS Nano* **8**, 12836–12842 (2014).
114. Allain, A., Kang, J., Banerjee, K. & Kis, A. Electric contacts to two-dimensional semiconductors. *Nat. Mater.* **14**, 1195–1205 (2015).
115. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
116. Liu, Y. et al. Pushing the performance limit of sub-100 nm molybdenum disulfide transistors. *Nano Lett.* **16**, 6337–6342 (2016).
117. Smithe, K., Suryavanshi, S., Rojo, M., Tedjarati, A. & Pop, E. Low variability in synthetic monolayer MoS_2 devices. *ACS Nano* **11**, 8456–8463 (2017).
118. Liu, W., Sarkar, D., Kang, J., Cao, W. & Banerjee, K. Impact of contact on the operation and performance of back-gated monolayer MoS_2 field-effect-transistors. *ACS Nano* **9**, 7904–7912 (2015).
119. Zhang, W., Huang, Z., Zhang, W. & Li, Y. Two-dimensional semiconductors with possible high room temperature mobility. *Nano Res.* **7**, 1731–1737 (2014).
120. Cui, Y. et al. High-performance monolayer WS_2 field-effect transistors on high-κ dielectrics. *Adv. Mater.* **27**, 5230–5234 (2015).
121. Withers, F., Bointon, T. H., Hudson, D. C., Craciun, M. F. & Russo, S. Electron transport of WS_2 transistors in a hexagonal boron nitride dielectric environment. *Sci. Rep.* **4**, 4967 (2014).
122. Song, H. S. et al. High-performance top-gated monolayer SnS_2 field-effect transistors and their integrated logic circuits. *Nanoscale* **5**, 9666–9670 (2013).
123. Lin, Y. et al. Single-layer ReS_2: two-dimensional semiconductor with tunable in-plane anisotropy. *ACS Nano* **9**, 11249–11257 (2015).
124. Wang, X. et al. Chemical vapor deposition growth of crystalline monolayer $MoSe_2$. *ACS Nano* **8**, 5125–5131 (2014).
125. Podzorov, V., Gershenson, M. E., Kloc, Ch., Zeis, R. & Bucher, E. High-mobility field-effect transistors based on transition metal dichalcogenides. *Appl. Phys. Lett.* **84**, 3301–3303 (2004).
126. Li, S. et al. Halide-assisted atmospheric pressure growth of large WSe_2 and WS_2 monolayer crystals. *Appl. Mater. Today* **1**, 60–66 (2015).
127. Pei, T. et al. Few-layer $SnSe_2$ transistors with high on/off ratios. *Appl. Phys. Lett.* **108**, 053506 (2016).
128. Yang, S. et al. Layer-dependent electrical and optoelectronic responses of $ReSe_2$ nanosheet transistors. *Nanoscale* **6**, 7226–7231 (2014).
129. Feng, W., Zheng, W., Cao, W. & Hu, P. Back gated multilayer InSe transistors with enhanced carrier mobilities via the suppression of carrier scattering from a dielectric interface. *Adv. Mater.* **26**, 6587–6593 (2014).
130. Zhang, Z. et al. Two-step heating synthesis of sub-3 millimeter-sized orthorhombic black phosphorus single crystal by chemical vapor transport reaction method. *Sci. China Mater.* **59**, 122–134 (2016).
131. Li, L. et al. Black phosphorus field-effect transistors. *Nat. Nanotechnol.* **9**, 372–377 (2014).
132. Qiao, J., Kong, X., Hu, Z., Yang, F. & Ji, W. High-mobility transport anisotropy and linear dichroism in few-layer black phosphorus. *Nat. Commun.* **5**, 4475 (2014).
133. Late, D. J. et al. GaS and GaSe ultrathin layer transistors. *Adv. Mater.* **24**, 3549–3554 (2012).
134. Chhowalla, M., Jena, D. & Zhang, H. Two-dimensional semiconductors for transistors. *Nat. Rev. Mater.* **1**, 16052 (2016).
135. Singh, J. et al. 14 nm FinFET technology for analog and RF applications. In *IEEE VLSI Technology Symposium* 140–141 (IEEE, 2017).
136. Du, Y., Yang, L., Zhou, H. & Ye, P. Performance enhancement of black phosphorus field-effect transistors by chemical doping. *IEEE Electron Device Lett.* **37**, 429–432 (2016).
137. Yang, L. et al. Chloride molecular doping technique on 2D materials: WS_2 and MoS_2. *Nano Lett.* **14**, 6275–6280 (2014).
138. Pang, C., Wu, P., Appenzeller, J. & Chen, Z. EOT WS_2-FET with I_{DS} > 600 μA/μm at V_{DS} = 1V and SS < 70mV/dec at L_G = 40 nm. In *IEEE International Electron Devices Meeting* 3.4.1–3.4.4 (IEEE, 2020).
139. Cao, W., Kang, J., Sarkar, D., Liu, W. & Banerjee, K. 2D semiconductor FETs- Projections and design for sub-10 nm VLSI. *IEEE Trans. Electron Devices* **62**, 3459–3469 (2015).
A comprehensive scalability analysis of 2D FETs that established the viability of certain 2D semiconductors as optimal channel materials for sub-10-nm FETs.
140. Liu, L., Kumar, S., Ouyang, Y. & Guo, J. Performance limits of monolayer transition metal dichalcogenide transistors. *IEEE Trans. Electron Devices* **58**, 3042–3047 (2011).
141. Liu, R. et al. Integrated digital inverters based on two-dimensional anisotropic ReS_2 field-effect transistors. *Nat. Commun.* **6**, 6991 (2015).
142. Mudd, G. et al. Tuning the bandgap of exfoliated InSe nanosheets by quantum confinement. *Adv. Mater.* **25**, 5714–5718 (2013).
143. Chau, R. Process and packaging innovations for Moore's law continuation and beyond. In *IEEE International Electron Devices Meeting* 1.1.1–1.1.6 (IEEE, 2019).
144. Salahuddin, S. & Datta, S. Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Lett.* **8**, 405–410 (2008).
Proposal of an NC FET.
145. Ionescu, A. M. et al. Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture In *Proceedings International Symposium on Quality Electronic Design* 496–501 (IEEE, 2002).
146. Huang, Q., Huang, R., Pan, Y., Tan, S. & Wang, Y. Resistive-gate field-effect transistor: a novel steep-slope device based on a metal–insulator–metal–oxide gate stack. *IEEE Electron Device Lett.* **35**, 877–879 (2014).
147. Han, J., Moon, D. & Meyyappan, M. Nanoscale vacuum channel transistor. *Nano Lett.* **17**, 2146–2152 (2017).
148. Shukla, N. et al. A steep-slope transistor based on abrupt electronic phase transition. *Nat. Commun.* **6**, 7812 (2015).
149. Gnani, E., Reggiani, S., Gnudi, A. & Baccarani, G. Steep-slope nanowire FET with a superlattice in the source extension. *Solid State Electron.* **65–66**, 108–113 (2011).
150. Qiu, C. et al. Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches. *Science* **361**, 387–392 (2018).
Demonstration of a Dirac source FET.
151. Quinn, J., Kawamoto, G. & McCombe, B. Subband spectroscopy by surface channel tunneling. *Surf. Sci.* **73**, 190–196 (1978).
Proposal for a tunnel FET structure using band-to-band tunnelling.
152. Appenzeller, J., Knoch, Y. & Avouris, P. Band-to-band tunneling in carbon nanotube field-effect transistors. *Phys. Rev. Lett.* **93**, 196805 (2004).
153. Wan, J., Zaslavsky, A., Le Royer, C. & Cristoloveanu, S. Novel bipolar-enhanced tunneling FET with simulated high on-current. *IEEE Electron Device Lett.* **34**, 24–26 (2013).
154. Saeidi, A. et al. Negative capacitance as performance booster for tunnel FETs and MOSFETs: an experimental study. *IEEE Electron Device Lett.* **38**, 1485–1488 (2017).
155. Gopalakrishnan, K., Griffin, P. & Plummer, J. Impact ionization MOS (I-MOS)-part I: device and circuit simulations. *IEEE Trans. Electron Devices* **52**, 69–76 (2005).
156. Padilla, A., Chun Wing, Y., Shin, C., Hu, C. & King Liu, T.-J. Feedback FET: a novel transistor exhibiting steep switching behavior at low bias voltages. In *IEEE International Electron Devices Meeting* 1–4 (IEEE, 2008).
157. Datta, S. & Das, B. Electronic analog of the electro-optic modulator. *Appl. Phys. Lett.* **56**, 665–667 (1990).
Proposal of a spin FET.
158. Manchon, A. et al. New perspectives for Rashba spin–orbital coupling. *Nat. Mater.* **14**, 871–882 (2015).
159. Banerjee, S., Register, L., Tutuc, E., Reddy, D. & MacDonald, A. Bilayer pseudospin field-effect transistor (BiSFET): a proposed new logic device. *IEEE Electron Device Lett.* **30**, 158–160 (2009).
160. Akarvardar, K. et al. Design considerations for complementary nanoelectro-mechanical logic gates. In *IEEE International Electron Devices Meeting* 299–302 (IEEE, 2007).
161. Dadgour, H., Hussain, M., Cassell, A., Singh, N. & Banerjee, K. Impact of scaling on the performance and reliability degradation of metal-contacts in NEMS devices. *IEEE International Reliability Physics Symposium* 280–289 (IEEE, 2011).
162. Cao, W., Sarkar, D., Khatami, Y., Kang, J. & Banerjee, K. Subthreshold-swing physics of tunnel field-effect transistors. *AIP Adv.* **4**, 067141 (2014).
163. Gandhi, R. et al. CMOS-compatible vertical-silicon-nanowire gate-all-around p-type tunneling FETs with ≤50 mV/decade subthreshold swing. *IEEE Electron Device Lett.* **32**, 1504–1506 (2011).
164. Tomioka, K., Yoshimura, M. & Fukui, T. Steep-slope tunnel field-effect transistors using III–V nanowire/Si heterojunction. In *IEEE Symposium on VLSI Technology* 47–48 (IEEE, 2012).
A heterojunction tunnel FET with a steep slope.
165. Knoll, L. et al. Inverters with strained Si nanowire complementary tunnel field-effect transistors. *IEEE Electron Device Lett.* **34**, 813–815 (2013).
166. Sarkar, D. et al. A subthermionic tunnel field-effect transistor with an atomically thin channel. *Nature* **526**, 91–95 (2015).
A 2D semiconductor channel tunnel FET with a steep slope.
167. Memisevic. E., Svensson, J., Hellenbrand, M., Lind, E. & Wernersson, L. Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with S = 48 mV/decade and I_{on} = 10 μA/μm for I_{off} = 1 nA/μm at V_{DS} = 0.3 V. In *IEEE International Electron Devices Meeting* 500–503 (IEEE, 2016).
168. Kim, S. et al. Thickness-controlled black phosphorus tunnel field-effect transistor for low-power switches. *Nat. Nanotechnol.* **15**, 203–206 (2020).
169. Cao, W. et al. Designing band-to-band tunneling field-effect transistors with 2D semiconductors for next-generation low-power VLSI. In *IEEE International Electron Devices Meeting* 12.3.1–12.3.4 (IEEE, 2015).
170. Ross, I. M. Semiconductive translating device. US patent 2791760 (1957).
171. Scott, J. F. & Araujo, C. Ferroelectric memories. *Science* **246**, 1400–1405 (1989).
172. Zhou, J. et al. Frequency dependence of performance in Ge negative capacitance PFETs achieving sub-30 mV/decade swing and 110 mV hysteresis. In *IEEE International Electron Devices Meeting* 373–376 (IEEE, 2017).
173. Z. Yu, et al. Negative capacitance 2D MoS_2 transistors with sub-60 mV/dec subthreshold swing over 6 orders, 250 μA/μm current density, and nearly-hysteresis-free. In *IEEE International Electron Devices Meeting* 577–580 (IEEE, 2017).
174. Si, M. et al. Steep-slope hysteresis-free negative capacitance MoS_2 transistors. *Nat. Nanotechnol.* **13**, 24–28 (2018).
175. Lee, M. et al. Ferroelectric Al:HfO_2 negative capacitance FETs. In *IEEE International Electron Devices Meeting* 565–568 (IEEE, 2017).
176. Fan, C. et al. Energy-efficient $HfAlO_x$ NCFET: using gate strain and defect passivation to realize nearly hysteresis- free sub-25 mV/dec switch with ultralow leakage. In *IEEE International Electron Devices Meeting* 561–564 (IEEE, 2017).
177. Chung, W. et al. Hysteresis-free negative capacitance germanium CMOS FinFETs with bi-directional sub-60 mV/dec. In *IEEE International Electron Devices Meeting* 365–368 (IEEE, 2017).

178. Cao, W. & Banerjee, K. Is negative capacitance FET a steep-slope logic switch?. *Nat. Commun.* **11**, 196 (2020).
This study demystifies the fundamental limitations of NC FETs and identifies alternative roles of NC in FET design.
179. Wang, H. et al. New insights into the physical origin of negative capacitance and hysteresis in NCFETs. In *IEEE International Electron Devices Meeting* 31.1.1–31.1.4 (IEEE, 2018).
180. Li, X. & Toriumi, A. Direct relationship between sub-60 mV/dec subthreshold swing and internal potential instability in MOSFET externally connected to ferroelectric capacitor. In *IEEE International Electron Devices Meeting* 31.3.1–31.3.4 (IEEE, 2018).
181. Jin, C., Jang, K., Saraya, T., Hiramoto, T. & Kobayashi, M. Experimental study on the role of polarization switching in subthreshold characteristics of HfO_2-based ferroelectric and anti-ferroelectric FET. In *IEEE International Electron Devices Meeting* 31.5.1–31.5.4 (IEEE, 2018).
182. Toprasertpong, K., Takenaka, M. & Takagi, S. Direct observation of interface charge behaviors in FeFET by quasi-static split C-V and Hall techniques: revealing FeFET operation. In *IEEE International Electron Devices Meeting* 23.7.1–23.7.4 (IEEE, 2019).
183. Su, L. T., Naffziger, S. & Papermaster, M. Multi-chip technologies to unleash computing performance gains over the next decade. In *IEEE International Electron Devices Meeting* 1–8 (IEEE, 2017).
184. Huang, C. et al. 3-D self-aligned stacked NMOS-on-PMOS nanoribbon transistors for continued Moore's law scaling. In *IEEE International Electron Devices Meeting* 20.6.1–20.6.4 (IEEE, 2020).
185. Banerjee, K., Souri, S., Kapur, P. & Saraswat, K. 3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration. *Proc. IEEE* **89**, 602–633 (2001).
A comprehensive treatise on 3D and heterogeneous integration from a circuit, system, thermal and technology perspective.
186. Batude, P. et al. Advances in 3D CMOS sequential integration. In *IEEE International Electron Devices Meeting* 14.1.1–14.1.4 (IEEE, 2009).
187. Wei, H. et al. Cooling three-dimensional integrated circuits using power delivery networks. In *IEEE International Electron Devices Meeting* 327–330 (IEEE, 2012).
188. Jiang, J., Parto, K., Cao, W. & Banerjee, K. Ultimate monolithic-3D integration with 2D materials: rationale, prospects, and challenges. *IEEE J. Electron Devices Soc.* **7**, 878–887 (2019).
A detailed analysis quantifying the benefits of monolithic and heterogeneous 3D integration with 2D materials.
189. Sachid, A. et al. Monolithic 3D CMOS using layered semiconductors. *Adv. Mater.* **28**, 2547–2554 (2016).
190. Jiang, J., Kang, J., Chu, J. & Banerjee, K. All-carbon interconnect scheme integrating graphene-wires and carbon-nanotube-vias. In *IEEE International Electron Devices Meeting* 342–345 (IEEE, 2017).
191. Shulaker, M. et al. Monolithic 3D integration of logic and memory: carbon nanotube FETs, resistive RAM, and silicon FETs. In *IEEE International Electron Devices Meeting* 638–641 (IEEE, 2014).
192. Zhang, D., Yeh, C., Cao, W. & Banerjee, K. 0.5T0.5R—an ultracompact RRAM cell uniquely enabled by van der Waals heterostructures. *IEEE Trans. Electron Devices* **68**, 2033–2040 (2021).
A novel FET-RRAM hybrid device.
193. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
194. Gonzalez-Zalba, M. et al. Scaling silicon-based quantum computing using CMOS technology. *Nat. Electron.* **4**, 872–884 (2021).
195. Maurand, R. et al. A CMOS silicon spin qubit. *Nat. Commun.* **7**, 13575 (2016).
Demonstration of a spin qubit based on CMOS technology.
196. Merolla, P. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
197. Davies, M. et al. Loihi: a neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
198. Burd, T. et al. Zen3: the AMD 2nd-generation 7nm x86-64 microprocessor core. In *IEEE International Solid-State Circuits Conference* 1–3 (IEEE, 2022).

Acknowledgements K.B. acknowledges support from the Army Research Office (grant W911NF1810366), the Air Force Office of Scientific Research (grant FA9550-18-1-0448), the Japan Science and Technology Agency CREST Program (grant SB180064) and the National Science Foundation (grant CCF 2132820). K.B. thanks the following individuals for their selfless support during the organization of the collaboration: T. Ernst, CEA-LETI, Grenoble, France; T. Sakurai, The University of Tokyo, Tokyo, Japan; J. Welser, IBM Almaden Research Centre, San Jose, USA. K.B. also thanks S. Oda, Tokyo Institute of Technology, Ōokayama, Japan, for useful discussions.

Author contributions K.B. organized and led the collaboration. W.C. and K.B. wrote the paper with input from all other authors.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06145-x.
Correspondence and requests for materials should be addressed to Kaustav Banerjee.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
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.

© Springer Nature Limited 2023, corrected publication 2023