nature electronics

Article

https://doi.org/10.1038/s41928-024-01279-w

Scaled vertical-nanowire heterojunction tunnelling transistors with extreme quantum confinement

Received: 14 July 2023

Accepted: 3 October 2024

Published online: 4 November 2024

Check for updates

Yanjie Shao [1,2] ✉, Marco Pala [3,4] ✉, Hao Tang [5], Baoming Wang [5], Ju Li [5,6] ✉, David Esseni [4] ✉ & Jesús A. del Alamo [1,2] ✉

The development of data-centric computing requires new energy-efficient electronics that can overcome the fundamental limitations of conventional silicon transistors. A range of novel transistor concepts have been explored, but an approach that can simultaneously offer high drive current and steep slope switching while delivering the necessary scaling in footprint is still lacking. Here, we report scaled vertical-nanowire heterojunction tunnelling transistors that are based on the broken-band GaSb/InAs system. The devices offer a drive current of 300 μA μm^{-1} and a sub-60 mV dec^{-1} switching slope at an operating voltage of 0.3 V. The approach relies on extreme quantum confinement at the tunnelling junction and is based on an interface-pinned energy band alignment at the tunnelling heterojunction under strong quantization.

Progress in artificial intelligence and the internet of things has led to an increase in computing demand[1–3] and an increase in the need for energy-efficient nanoelectronics[4,5]. What is now required is a transistor technology with a tiny footprint that can deliver high performance with low-voltage operation[4]. However, conventional silicon complementary metal–oxide–semiconductor (CMOS) technology is approaching its fundamental limits[6–8], so that achieving these aims through further CMOS scaling is challenging. To start, the sharpness of the turn-on characteristics of CMOS transistors, which is characterized by the subthreshold swing (S; defined as the increase in the gate-to-source voltage V_{gs} needed to increase the drain current I_d by a factor of 10), is limited to 60 mV dec^{-1} at room temperature. This is known as the Boltzmann tyranny and arises from Fermi–Dirac electron statistics[4]. Furthermore, short-channel effects and direct source-to-drain tunnelling impose a minimum gate length (L_g) and with this a scaling limit to the transistor footprint[9,10]. Further reductions in the supply voltage (V_{dd}) require breaking the 60 mV dec^{-1} thermionic limit of metal–oxide–semiconductor field-effect transistors (MOSFETs)[1]. To achieve this, transistors based on negative capacitance[11,12], cold source[13] and phase change[14] have been explored. Alternatively, theory predicts that tunnelling in a heterojunction could be used to create a transistor with a drive current of over 100 μA μm^{-1} at V_{dd} = 0.3 V and a deep-subthermionic switching slope[15–19]. However, experimental demonstrations of heterojunction tunnelling transistors have not offered these two features simultaneously in the same device[20–27].

One approach for creating tunnelling transistors is to use vertical-nanowire geometry heterojunctions based on broken-band GaSb/InAs (ref. 16). In a vertical transistor structure, L_g scaling is uncoupled from footprint scaling, which leads to flexible management of short-channel effects. A vertical-nanowire geometry with a gate-all-around configuration offers the tightest possible electrostatic charge control on the smallest footprint[28]. Furthermore, a vertical configuration of a heterojunction device can benefit from abrupt epitaxial growth[29]. However, given the nanofabrication challenges, there is a limited understanding of the tunnelling physics in extremely quantum-confined heterojunctions, especially the roles of surfaces and

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, US. [2]Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA, US. [3]Université Paris-Saclay, CNRS, C2N, Palaiseau, France. [4]DPIA, University of Udine, Udine, Italy. [5]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, US. [6]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, US. ✉e-mail: shaoyj@mit.edu; marco.pala@uniud.it; liju@mit.edu; david.esseni@uniud.it; alamo@mit.edu

https://doi.org/10.1038/s41928-024-01279-w

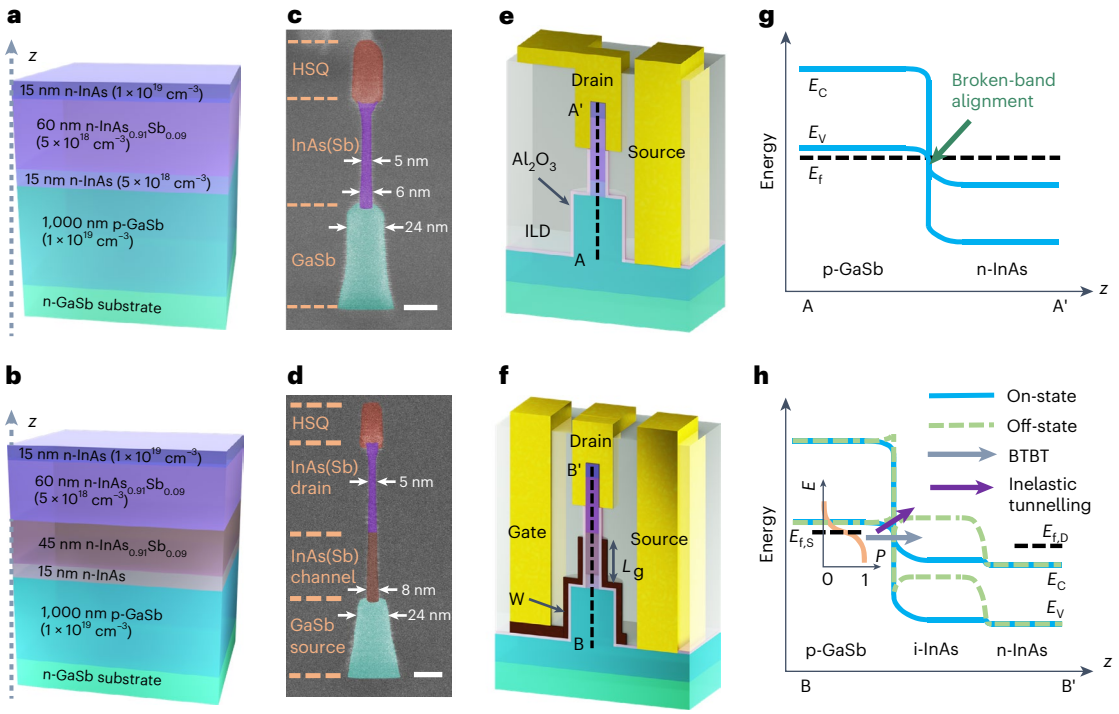

Fig. 1 | Ultra-scaled vertical-nanowire device design. a,**b**, Heterostructures in the fabrication of a heterojunction Esaki diode (**a**) and a tunnelling transistor (**b**). **c**,**d**, False-coloured 30°-tilted SEM images of vertical nanowires in the Esaki diode (**c**) and the tunnelling transistor (**d**), which were fabricated by dry-etching and citric acid:H_2O_2 wet-etching with D_{InAs} of around 6 and 8 nm at the tunnelling junction, respectively. **e**,**f**, Schematics of the finished device structures for the vertical-nanowire Esaki diode (**e**) and vertical-nanowire tunnelling transistor (**f**). L_g denotes the gate length, which is around 50 nm. W denotes tungsten, the gate metal. **g**,**h**, Schematic energy band diagrams for a vertical-nanowire diode and a vertical-nanowire transistor along the cut lines A–A' (**e**) and B–B' (**f**). E_c, E_v, $E_{f,S}$ and $E_{f,D}$ denote the lowest conduction sub-band, the highest valence sub-band, the source Fermi level and the drain Fermi level, respectively. Inset of **h**, Fermi–Dirac distribution for a finite temperature in the source. E and P denote the energy and occupation probability, respectively. Scale bars, 20 nm (**c**,**d**).

interfaces, and such a regime remains unexplored for heterojunction tunnelling-transistor designs.

In this article, we describe scaled vertical-nanowire heterojunction tunnelling transistors that are based on the broken-band GaSb/InAs system. The transistors offer a drive current of 300 μA μm^{-1}, a peak transconductance of 1,050 μS μm^{-1} and a sub-60 mV dec^{-1} switching slope at a 0.3 V drain-to-source voltage (V_{ds}) and at room temperature. The approach relies on extreme quantum confinement that enhances the tunnelling current density and is the result of an interface-pinned band alignment at the heterojunction in the presence of strong transverse quantum confinement.

Design and fabrication of device structures

Our vertical device design exploits the ability of molecular-beam epitaxy (MBE) to grow abrupt heterostructures. Vertical-nanowire devices were fabricated by a top-down approach with active device regions defined by etching technology with nanometre accuracy (Methods). We first studied the tunnelling physics in two-terminal heterojunction Esaki diodes and then the fabrication and performance characterization of three-terminal heterojunction tunnelling transistors. The semiconductor heterostructures of the diodes (Fig. 1a) and transistors (Fig. 1b) share the same design principles and include a GaSb substrate, an atomically sharp GaSb/InAs interface (Supplementary Fig. 1), fully strained heterostructures made using a combination of thin strained InAs and thick InAsSb layers lattice-matched to GaSb, and the use of Si as, simultaneously, a p-type dopant for GaSb and an n-type dopant for InAs and InAsSb.

The vertical nanowires in both heterostructures were fabricated using chlorine-based dry-etching[30], selective citric-acid-based wet-etching (Methods) and alcohol-based digital-etching[31]. To explore the impact of extreme quantization, we scaled the InAs diameter (D_{InAs}) down to 5 nm, while leaving a thicker GaSb portion to ensure the mechanical stability of the devices (Fig. 1c,d).

For the Esaki diodes, after etching the vertical nanowires and growing a 3 nm Al_2O_3 passivation layer, an interlayer dielectric (ILD) was applied to isolate the source and drain contacts (Fig. 1e). The contacts were made using a Ni/Mo bilayer. An image of a finished vertical-nanowire Esaki diode taken with high-angle annular dark-field imaging and scanning transmission electron microscopy (STEM) and the corresponding element mappings captured with energy-dispersive X-ray spectroscopy (Supplementary Fig. 2) match the expected device structure illustrated in Fig. 1e.

For the tunnelling transistors, after etching the vertical nanowires, a 2.4 nm Al_2O_3 film was deposited as the gate dielectric with an equivalent oxide thickness (the equivalent thickness of SiO_2 that provides the same oxide capacitance) of 1.2 nm (Fig. 1f). A gate-all-around configuration was fabricated using tungsten with an effective L_g of around 50 nm. The source, gate and drain Ni/Mo contacts were fabricated and isolated by several ILD layers.

Heterojunction tunnelling with strong quantum confinement

We studied the heterojunction tunnelling physics through the electrical characterization of our ultra-scaled Esaki diodes. The broken-band alignment at the heterojunction interface and heavy doping in both the n- and p-regions in the Esaki diode yield the energy band diagram illustrated in Fig. 1g. Throughout this paper, we use a transistor-like notation for the voltage and define V_{ds} as the voltage of the InAs region with respect to the GaSb region (Supplementary Fig. 3a,b). With this notation, a negative differential resistance (NDR) regime (Supplementary Fig. 3c–e) and a thermally activated regime (Supplementary Fig. 3f) are expected with increasing negative V_{ds}, whereas a pure band-to-band

https://doi.org/10.1038/s41928-024-01279-w

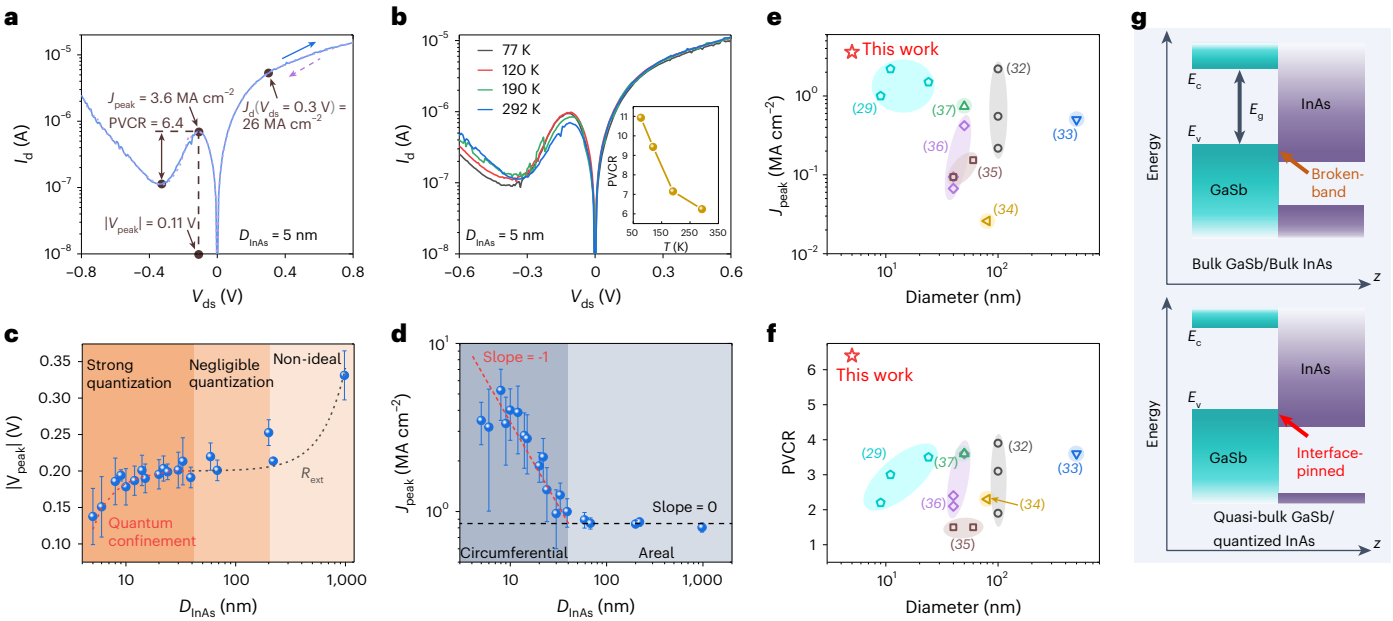

Fig. 2 | Electrical characteristics and benchmarking of vertical-nanowire Esaki diodes. a, Current versus voltage (I–V) sweep of an Esaki diode with D_{InAs} = 5 nm and D_{GaSb} = 24 nm featuring negligible hysteresis, ultra-high peak current density (J_{peak}) of 3.6 MA cm^{-2} and large PVCR = 6.4. The current density was normalized by the InAs cross-sectional area. D_{InAs} and D_{GaSb} denote the InAs and GaSb diameters, respectively. **b**, Temperature-dependent I–V characteristics of the device shown in **a** from 77 K to room temperature. Inset, PVCR extracted from the I–V characteristics as a function of temperature. **c**, Peak voltage ($|V_{peak}|$) versus D_{InAs}, showing a strong quantum confinement regime at small diameters that is dominated by $E_f - E_c$ in InAs (dark orange region), a negligible quantum confinement regime at medium diameters with nearly constant $|V_{peak}|$ that is determined by doping levels (medium dark orange region) and a non-ideal regime at large diameters that is dominated by a non-scalable extrinsic resistance (R_{ext}) from probe pads and probe tips (light orange region). The red dashed line is a trend line from D_{InAs} = 5 nm to 40 nm, whereas the black dashed line was calculated from $|V_{peak}| = |V_{peak,0}| + I_{peak}R_{ext}$, where $|V_{peak,0}|$ = 0.2 V is the intrinsic peak voltage of bulk-like devices without quantization and R_{ext} = 20 Ω. **d**, Experimental J_{peak} versus D_{InAs}, showing areal scaling for D_{InAs} larger than 40 nm (light blue region) and circumferential scaling for D_{InAs} less than 40 nm (dark blue region). The black and red dashed lines are trend lines with slopes of 0 and −1, respectively, in a log-log scale. Each data point in **c** and **d** is the mean value from six different devices. Error bars in **c** and **d** denote one standard deviation on each side. **e**,**f**, Benchmarking of J_{peak} (**e**) and PVCR (**f**), respectively, versus diameter at the tunnelling junction with other reported Esaki diodes in the GaSb/InAs material system. Devices with PVCR ≥ 1.5 are benchmarked to avoid the impact of a non-ideal thermal current on the BTBT current. **g**, Schematic band alignments at the GaSb/InAs interface without (top) and with (bottom) quantum confinement in InAs under interface-pinned conditions.

tunnelling (BTBT) regime should occur for positive V_{ds} (Supplementary Fig. 3g). NDR is usually characterized by peak-to-valley current ratio (PVCR), defined as peak current (I_{peak})/valley current (I_{valley}).

The experimental current-voltage (I–V) characteristics of an Esaki diode with D_{InAs} = 5 nm are shown in Fig. 2a. A clear NDR region is observed. An ultra-high peak current density (J_{peak}) of 3.6 MA cm^{-2} when normalized with respect to the InAs cross-sectional area was achieved, with negligible hysteresis. To the best of our knowledge, this is one of the highest values achieved among all Esaki diodes with a clear NDR reported, irrespective of the configuration or material system (Fig. 2e)[29,32–37]. We found that a good ohmic contact can be obtained despite the small diameter of our vertical nanowire, as confirmed by the sharp linear turn-on at positive V_{ds} (Supplementary Fig. 4a). The NDR regime shows a high PVCR = 6.4, one of the highest attained for GaSb/InAs Esaki diodes (Fig. 2f)[29,32–37]. The high-current potential of tunnelling transistors based on this material system at a reduced V_{dd} is evident by the current density of the Esaki diode $J_d(V_{ds}$ = 0.3 V), which yields 26 MA cm^{-2} or 325 μA μm^{-1} (normalized by the InAs circumference) (Fig. 2a).

The near-ideal nature of the ultra-scaled Esaki diode was further demonstrated when we cooled the sample down to cryogenic temperatures. As shown in Fig. 2b, the current at positive V_{ds} is independent of temperature, confirming its BTBT nature. Thermally activated current components that dominate I_{valley} and the more negative V_{ds} regime decreased with decreasing temperature (T), whereas I_{peak} increased, leading to an increasing PVCR (Fig. 2b, inset). The increase of I_{peak} as T was decreased occurred because the Fermi tail shortened with decreasing T, resulting in a larger Fermi window determined by $\int[F_n(E) - F_p(E)]\,dE$, where E denotes energy and $F_{n(p)}$ denote the Fermi–Dirac distribution in n-InAs (p-GaSb)[38].

In our devices, the diameter of the GaSb portion (D_{GaSb}) is sufficiently large (smallest value is around 24 nm) such that quantization in this region is negligible. Therefore, it is the quantization in the InAs region that matters. As quantum confinement results in quantized energy sub-bands, throughout this article unless otherwise indicated, $E_{c,X}$ ($E_{v,X}$) refers to the lowest (highest) quantized conduction (valence) sub-band edges of material X. The diameter-dependent I–V characteristics of the Esaki diodes (Supplementary Fig. 4b) act as a sensitive probe for studying the effect of strong quantum confinement on the energy band structure and current transport. One distinctive feature was the decrease of the absolute value of the peak voltage ($|V_{peak}|$, the voltage at which the Esaki peak current occurred) as D_{InAs} was decreased below around 40 nm (Fig. 2c). In bulk-like devices where D_{InAs} is in the range 40 to 200 nm, an average $|V_{peak}|$ of around 0.2 V was observed. This is well understood, as $|V_{peak}|$ is mostly determined by the difference between the Fermi levels in the source and drain regions, which, in the absence of quantization, is purely set by the respective doping levels. When D_{InAs} was increased beyond around 200 nm, a quadratic increase of $|V_{peak}|$ was observed due to the non-scalable extrinsic resistance (R_{ext}) of around 20 Ω that originated from the probe pads and probe tips[29]. When D_{InAs} was decreased below 40 nm, $|V_{peak}|$ gradually decreased to the lowest average value of around 0.13 V at D_{InAs} = 5 nm. This is direct evidence of the decrease in the energy difference between the Fermi level in the drain side ($E_{f,D}$) and E_c in InAs with decreasing D_{InAs}, as the electron effective mass (m_e^*) in InAs dramatically increased (Supplementary Fig. 5a,b) whereas the GaSb source remained bulk-like. The

https://doi.org/10.1038/s41928-024-01279-w

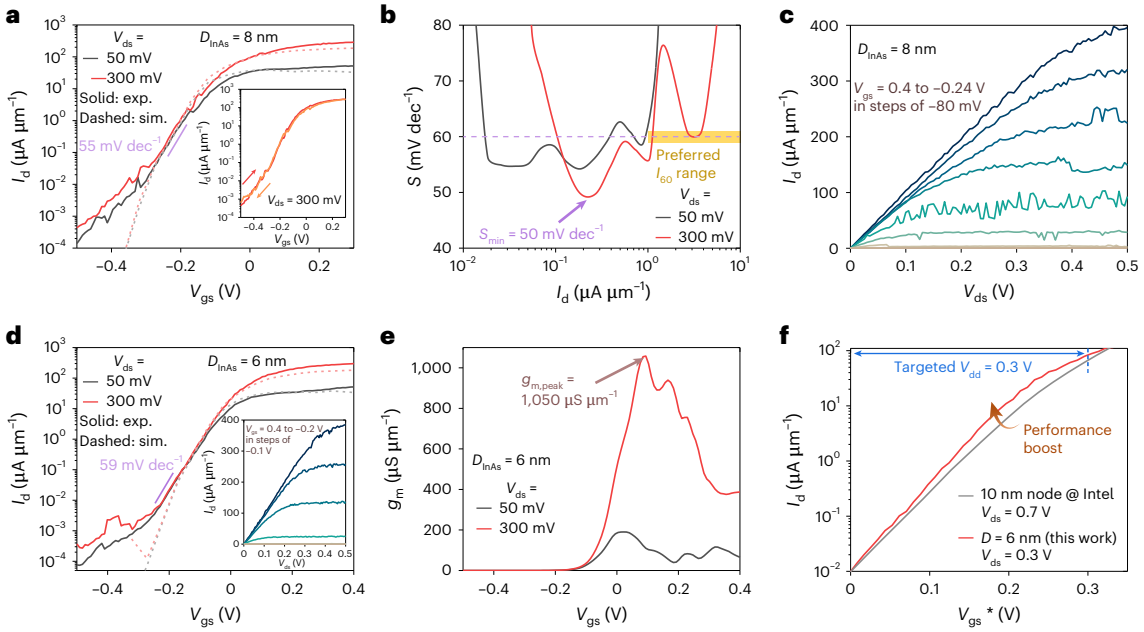

Fig. 3 | Electrical characteristics of ultra-scaled vertical-nanowire tunnelling transistors. a–**c**, Transfer characteristics (**a**), subthreshold swing (S) versus I_d (**b**) and output characteristics (**c**) of a transistor with D_{InAs} = 8 nm. Inset in **a**, Hysteresis sweep. S_{min} denotes the minimum S, and I_{60} denotes the maximum I_d for sub-60 mV dec^{-1} operation. **d**–**f**, Transfer characteristics (**d**), transconductance (g_m) characteristics (**e**) and performance comparison with CMOS technology at the 10 nm node[43] (**f**) of a transistor with D_{InAs} = 6 nm. $g_{m,peak}$ denotes the maximum g_m. In **f**, each of the transfer curves has been shifted by a constant gate voltage to achieve an off-state current I_{off} = 10 nA μm^{-1} at V_{gs} = 0 (denoted as V_{gs}*). Inset in **d**, Output characteristics. The 3D quantum-transport simulations in **a** and **d** show good agreement with experimental results in both subthreshold and on-state regimes (Methods). Note that the reported I_d and g_m have been normalized with respect to the active gate periphery, that is, the circumference of the InAs channel. exp., experimental; sim., simulated.

consequence was that the tunnelling window for electrons to tunnel from InAs to GaSb narrowed so that $|V_{peak}|$ was reduced.

The other distinctive feature of the extreme quantization regime is an unexpected circumferential scaling of I_{peak}. Whereas normal areal scaling prevailed in bulk-like devices with D_{InAs} larger than 40 nm, in devices with D_{InAs} less than 40 nm, we found that I_{peak} scaled with the InAs periphery (Fig. 2d). This anomalous tunnelling current behaviour emerged from quantization, as confirmed by the $|V_{peak}|$ scaling in the same D_{InAs} range. A consequence was that strong quantum confinement was not only not catastrophic for tunnelling[22] but actually favourable for intense BTBT at the GaSb/InAs heterojunction, as it gave rise to a much larger current than otherwise predicted[18]. Our observation of a high BTBT current density together with large PVCR in heterojunction Esaki diodes strongly suggests the feasibility of simultaneously obtaining a high drive current and steep S in ultra-scaled heterojunction tunnelling transistors.

Vertical-nanowire heterojunction tunnelling transistors

In our transistor design, the E_v of the p-GaSb source was used to filter out electrons in the Fermi tail at the source, so that when the E_c of the InAs channel was driven by an increasing V_{gs} from above to below the E_v of GaSb, there was a sharp rise in the BTBT current, which defeated the thermionic limit. In the off state (Fig. 1h, dashed line), BTBT was drastically suppressed, although thermally activated inelastic tunnelling processes[39] often degrade the off-state leakage. In the on state of a tunnelling transistor (Fig. 1h, solid line), the tunnelling physics are identical to those in Esaki diodes and a high on-current is expected.

The transfer characteristics of a representative transistor with D_{InAs} = 8 nm are shown in Fig. 3a. This device had a maximum drain current $I_{d,max}$(V_{ds} = 0.3 V) = 300 μA μm^{-1}, together with an average S, S_{avg} = 55 mV dec^{-1}, at both V_{ds} = 0.05 and 0.3 V, over one decade of current or more. A maximum to minimum drain current ratio ($I_{d,max}/I_{d,min}$) of 6 × 10^5 was observed at V_{ds} = 0.3 V. The minimum S, S_{min} = 50 mV dec^{-1} (Fig. 3b), which breaks the thermionic limit in MOSFETs. The output characteristics of the same device display well-behaved gate-controllable current switching, as well as good current saturation at large V_{ds} (Fig. 3c). Note the noisy nature of the measured current. This has been widely observed in vertical-nanowire transistors with relatively small diameters[40,41], and it has been attributed to discrete trapping events[41,42]. Nevertheless, we observed negligible hysteresis in the transfer curves with upward and downward sweeping, and the subthermionic operation was preserved in both sweeps (Fig. 3a, inset).

By further scaling D_{InAs} down to 6 nm, we found that $I_{d,max}$(V_{ds} = 0.3 V) could be maintained at 300 μA μm^{-1}, with $I_{d,max}/I_{d,min}$ = 10^6 and S_{avg} = 59 mV dec^{-1} at both V_{ds} = 0.05 and 0.3 V (Fig. 3d). The combination of steep switching slope and high BTBT current yielded a peak transconductance ($g_{m,peak}$) of 1,050 μS μm^{-1} at V_{ds} = 0.3 V (Fig. 3e), a key transistor figure of merit for on-state performance. The negligible drain-induced barrier lowering, as observed in the subthreshold characteristics (Fig. 3d), combined with excellent current saturation, as observed in the output curves (Fig. 3d, inset), show the immunity of our transistor architecture to short-channel effects.

The performance of the 6-nm-diameter transistor relative to state-of-the-art Si CMOS technology at the 10 nm node[43] is assessed in Fig. 3f. At a fixed off-state current I_{off} = 10 nA μm^{-1}, this figure highlights the performance boost delivered by heterojunction tunnelling. An on-state current I_{on} = 84 μA μm^{-1} was obtained at the targeted V_{dd} = V_{gs}* = V_{ds} = 0.3 V in our transistor (V_{gs}* denotes the shifted V_{gs} to accommodate the different threshold voltages of these transistors), whereas around 65 μA μm^{-1} was obtained for an Si MOSFET at V_{gs}* = 0.3 V with V_{ds} = 0.7 V, indicating at least 30% I_{on} enhancement despite the reduced V_{ds}.

The necessity of diameter scaling into the sub-10 nm regime to realize the steep turn-on potential of GaSb/InAs tunnelling transistors is illustrated in Fig. 4. Subthreshold characteristics of typical devices with various diameters are shown in Fig. 4a. S improved from over 500 to sub-60 mV dec^{-1} as D_{InAs} was scaled down from 45 to 6 nm, and there

https://doi.org/10.1038/s41928-024-01279-w

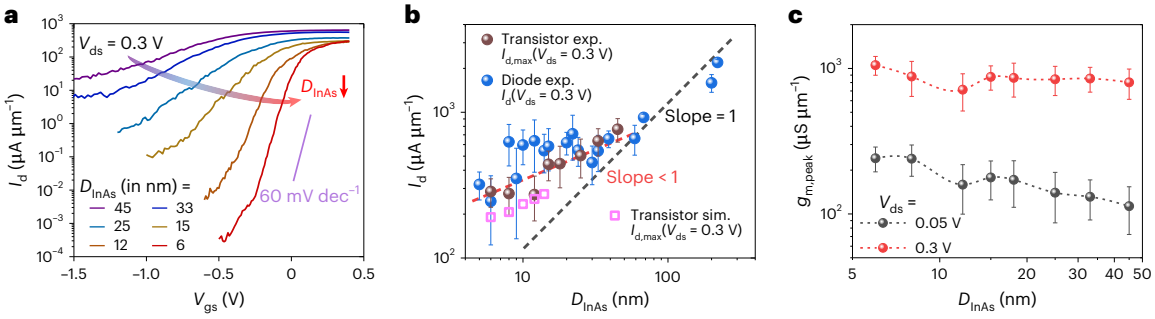

Fig. 4 | Scaling the diameter of vertical-nanowire devices. a, Transfer characteristics of typical vertical-nanowire tunnelling transistors with D_{InAs} ranging from 45 nm to 6 nm. **b**, Diameter scaling of current at V_{ds} = 0.3 V, denoted as I_d(V_{ds} = 0.3 V), for Esaki diodes, as well as the maximum current achieved at V_{ds} = 0.3 V, denoted as $I_{d,max}$(V_{ds} = 0.3 V), for tunnelling transistors (solid symbols). The black and red dashed lines are trend lines with a slope of 1 and less than 1 indicating areal and near-periphery scaling, respectively, in a log-log scale. Quantum-transport simulation results using the interface-pinned band alignment are also shown (open squares). **c**, $g_{m,peak}$ versus D_{InAs} in vertical-nanowire tunnelling transistors for V_{ds} = 0.05 and 0.3 V. Each data point for the experimental results in **b** and **c** is the mean value from six different devices. Error bars in **b** and **c** denote one standard deviation on each side.

was a positive threshold voltage shift. In addition, $I_{d,max}/I_{d,min}$ increased as the diameter was scaled down. These observations can be attributed to a tighter electrostatic charge control in the radial direction as D_{InAs} was decreased, leading to well-controlled short-channel effects[7,9]. Note that our process yielded four ultra-scaled devices with subthermionic operation. The subthreshold characteristics of these devices are shown in Supplementary Fig. 6. Figure 4b compares the on-state behaviour of Esaki diodes and tunnelling transistors, where I_d(V_{ds} = 0.3 V) for diodes and $I_{d,max}$(V_{ds} = 0.3 V) for transistors are plotted against D_{InAs}. A nearly circumferential scaling behaviour for both types of devices emerged in the extreme-scaled regime, like what was observed for J_{peak} in the Esaki diodes in Fig. 2d. This is because BTBT dominated in all three situations. The transistor average $g_{m,peak}$ at V_{ds} = 0.3 V improved to larger than 1 mS μm^{-1} as D_{InAs} was scaled down to 6 nm (Fig. 4c). Note that the observed increase in $g_{m,peak}$ is not commonly seen in highly scaled MOSFETs[44]. This is because the velocity of the charge carriers in the channel dominates the on-state current transport in MOSFETs, whereas BTBT dominates in tunnelling transistors. As the critical dimension is scaled down, although the electron velocity in the channel tends to decrease due to surface scattering, the heterojunction tunnelling rate improves drastically leading to an increased $g_{m,peak}$.

First-principles and quantum-transport modelling

The origin of the scaling of the tunnelling current in the extreme quantum confinement regime that we experimentally observed requires a theoretical examination. The tunnelling current density at a given bias is directly related to the product of the BTBT probability, determined by the band alignment at the interface, and the joint density of states (DOS) of the GaSb valence band and the InAs conduction band[45]. From a conventional understanding, confinement in the transverse direction of a heterojunction nanowire results in a change of the band alignment between quantized sub-bands[18,46]. A transition from broken- to staggered-band alignment is predicted at a D_{InAs} of around 7 nm (Supplementary Fig. 7a). Quantum-transport simulations (Methods), however, show that such a band alignment would result in a substantial decrease of the tunnelling current density as D_{InAs} is scaled down to the smallest dimensions explored here (Supplementary Fig. 7c,d). This is due to an increase in the tunnelling barrier width and the resulting exponential decrease in BTBT probability.

We performed two types of calculations: (1) first-principles calculations of the band alignment at the heterojunction interface and (2) quantum-transport simulations under the band alignment condition revealed by the first-principles results. First-principles electronic-structure calculations were carried out for a bulk-GaSb/bulk-InAs heterojunction as well as for a confined-InAs/bulk-GaSb geometry (Supplementary Fig. 8a,b). Unlike the conventional understanding, an interface-pinned band alignment (Fig. 2g) with a nearly constant energy difference between $E_{v,GaSb}$ and $E_{c,InAs}$ was observed in both bulk and confined geometries (Supplementary Figs. 8c,d and 9). Subsequently, quantum-transport simulations predict that such a band alignment (Supplementary Fig. 7b) leads to an increased areal density of the tunnelling current as D_{InAs} decreases (Supplementary Figs. 7c,d). The modelled current levels agree well with experimental results (Fig. 4b).

The physical reason for the observed enhancement in the tunnelling current areal density in the extreme quantum confinement regime is twofold. With an interface-pinned band alignment, the BTBT probability at the heterojunction is only weakly reduced by diameter scaling as the thickness of the tunnelling barrier changes little. However, for very small D_{InAs}, the enlarged m_e^* of InAs results in an enhanced DOS for InAs (Supplementary Fig. 5a) and, therefore, an increased current density. In addition, although in the large D_{InAs} regime the number of conduction modes (the number of quantized sub-bands that lie in a certain Fermi window) in an InAs nanowire scales linearly with its cross-sectional area, in the ultra-scaled-diameter regime, that dependence becomes sublinear. The above two mechanisms jointly result in an approximately linear dependence of absolute current on D_{InAs}, as observed in both the experimental and calculation results (Fig. 4b and Supplementary Figs. 7c,d).

The subthreshold characteristics between experiments and self-consistent quantum-transport simulations of D_{InAs} = 8 and 6 nm transistors are compared in Fig. 3a,d, respectively. The good agreement in both the on state and the subthreshold regime (Fig. 3a,d) at both V_{ds} = 0.05 and 0.3 V validates the interface-pinned band alignment hypothesis. The higher leakage current floor in the off state in the experiments was due to unavoidable trap-assisted generation and inelastic tunnelling processes, which were not considered in the simulations. More detailed quantum-transport simulation results are presented in Supplementary Fig. 10. The residual difference of J_{peak} in the Esaki diodes between experiments (Fig. 2d) and quantum-transport simulations with the interface-pinned band alignment (Supplementary Fig. 7c) was probably due to the positive fixed charges in Al_2O_3 (ref. 47), which led to electron accumulation in InAs (Supplementary Fig. 5c–e) and the opening of a larger tunnelling window.

Benchmarking

Future CMOS technology must meet stringent requirements in terms of transistor operating voltage and active area. We assess the relative potential of our tunnelling transistors by normalizing the on-state performance metrics with the tunnelling junction active area (g_m/InAs cross-sectional area). To the best of our knowledge, our

https://doi.org/10.1038/s41928-024-01279-w

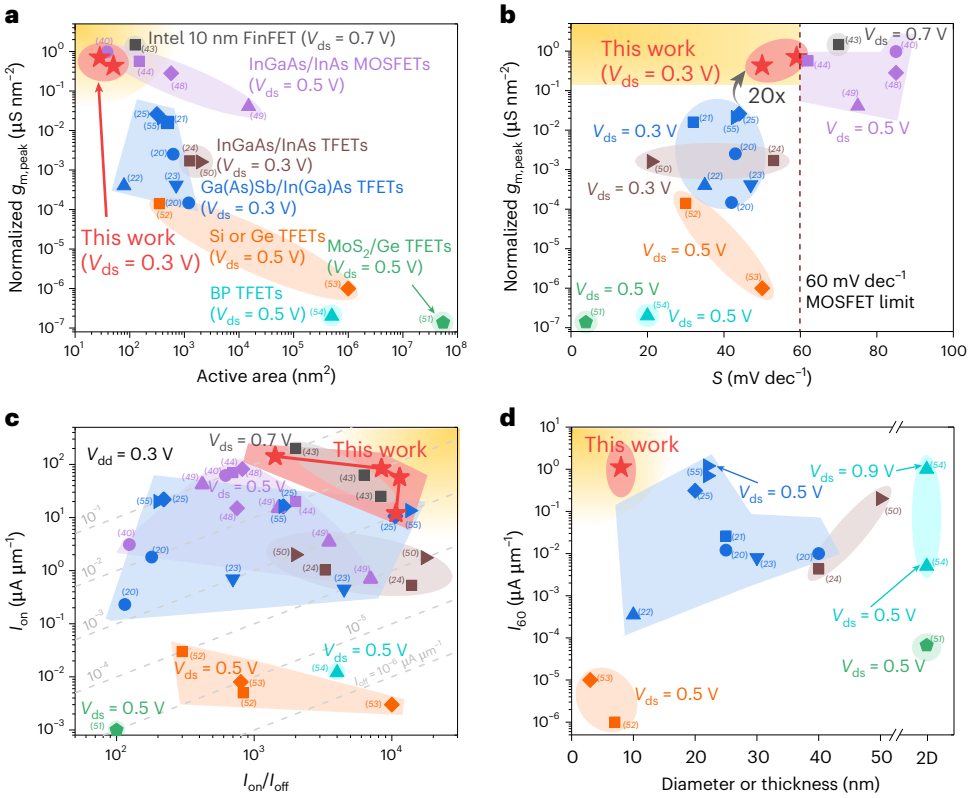

Fig. 5 | Benchmarking transistor technologies. a, Normalized $g_{m,peak}$ versus active area of our heterojunction tunnelling transistors against state-of-the-art tunnel field-effect transistors (TFETs) that feature subthermionic operation and 3D MOSFETs, including 3D InGaAs/InAs MOSFETs and commercial 3D Si MOSFETs at the '10 nm' technology node. The active area is defined as the channel footprint for planar devices and the channel cross-sectional area parallel to the chip surface for vertical ones (Methods). Our 6-nm-diameter heterojunction tunnelling transistors have a small active area compared to TFETs and 3D MOSFETs. Their performance is comparable to the best MOSFETs but at a reduced voltage. **b**, Normalized $g_{m,peak}$ versus S of our heterojunction tunnelling transistors against the same set of TFETs and MOSFETs as in **a**, showing a factor of 20 improvement in on-state performance of our device compared to other TFETs as well as a steeper switching slope compared to MOSFETs. **c**, I_{on} versus I_{on}/I_{off} of our heterojunction tunnelling transistors against the same set of TFETs and MOSFETs as in **a** (unless labelled, V_{dd} = 0.3 V), showing the high I_{on} with I_{on}/I_{off} larger than 10^4 at V_{dd} = 0.3 V. **d**, I_{60} versus channel critical dimension (two-dimensional material channels are plotted on the right) of our heterojunction tunnelling transistors against the same set of TFETs in **a** (unless labelled, V_{ds} = 0.3 V), showing the high I_{60} at V_{ds} = 0.3 V. The two red stars in each of **a** and **b** indicate the two devices reported in Fig. 3. The red stars in **c** indicate the performance of the 6-nm-diameter device shown in Fig. 3d–f at different I_{on} levels. The red star in **d** indicates the performance of the 8-nm-diameter device shown in Fig. 3a–c. Desired corners for optimum performance are indicated in yellow for all the panels. BP, black phosphorus; FinFET, fin field-effect transistor.

6-nm-diameter vertical-nanowire heterojunction tunnelling transistor has one of the smallest active areas among tunnelling transistors and three-dimensional (3D) MOSFETs to date, while delivering a normalized on-state performance at V_{ds} = 0.3 V that is comparable to state-of-the-art group III–V MOSFETs (refs. 40,44,48,49) and commercial Si fin field-effect transistors (ref. 43) operated at higher voltages (0.5 and 0.7 V, respectively) (Fig. 5a). A steeper switching slope can be achieved in our transistors compared to MOSFETs (Fig. 5b). Notably, a comparison with other subthermionic tunnelling transistors[20–25,50–55] shows a 20 times improvement in normalized on-state performance (Fig. 5b).

A long-lasting challenge for any logic technology is to achieve a combination of both high I_{on} and high I_{on}/I_{off} at a low V_{dd} = 0.3 V. This is due to the difficulty of achieving a sharp switching slope across a wide range of current. In fact, this is why V_{dd} scaling has recently become saturated in CMOS technology[1,4,8,15]. Figure 5c benchmarks I_{on} versus I_{on}/I_{off} for the same group of devices shown in Fig. 5a,b. The desirable regime of operation in this diagram is at the top right corner. With I_{on}/I_{off} larger than 10^4, the maximum I_{on} reached in tunnelling transistors so far has been around 15 μA μm^{-1} (ref. 55). On the other hand, although a higher I_{on} can generally be realized in MOSFETs, their I_{on}/I_{off} is always less than 10^4. We show that an I_{on} of around 60 μA μm^{-1} can be obtained with I_{on}/I_{off} larger than 10^4 at V_{dd} = 0.3 V. Besides, I_{on}/I_{off} larger than 10^3 was demonstrated with I_{on} = 150 μA μm^{-1}.

As pointed out in previous studies[56], ideally, a sub-60 mV dec^{-1} operation should be preserved in a current range of up to at least 1 μA μm^{-1}. The maximum I_d for sub-60 mV dec^{-1} operation (I_{60}) of our ultra-scaled transistors was 1.1 μA μm^{-1} (Fig. 3b). Our tunnelling transistor had I_{60} larger than 1 μA μm^{-1} at V_{ds} = 0.3 V (Fig. 5d).

Conclusions

We have reported scaled tunnelling electronics based on a broken-band heterojunction design. Our approach uses extreme quantum confinement at the tunnelling junction and relies on interface-pinned band alignment, a dramatic increase of the DOS in the InAs conduction band and tight gate-all-around electrostatic charge control on the thin transistor body. The 3D transistors have a small active area and exhibit high performance at a reduced operating voltage. Our work could be of use in the development of new devices for future CMOS scaling and highlights the possibilities of nanoscale heterojunction devices in the extreme quantization regime.

Methods

Heterostructure growth

The heterostructures (Figs. 1a,e) used in this work were grown by Intelligent Epitaxy Technology, Inc. (IntelliEPI) using MBE. Samples were grown on 3-inch n-type GaSb substrates with a (001) surface. The use

of $InAs_{0.91}Sb_{0.09}$ layers, which have the same lattice constant as GaSb, resulted in fully strained heterostructures with no visible defects or dislocations (Supplementary Fig. 1a). An atomically sharp 'GaAs-like' interface was demonstrated between GaSb and InAs (Supplementary Fig. 1b,c). This was expected to result in a nearly ideal GaSb/InAs tunnelling junction. Si was used as the only dopant throughout the heterostructures as Si is a p-type dopant in GaSb but an n-type dopant in InAs(Sb). This resulted in an abrupt doping profile at the tunnelling junction for Esaki diodes. In tunnelling transistors, a p-n-i tunnelling junction was expected due to unavoidable Si diffusion into the i-InAs layer. This has been shown to be beneficial for the performance of a tunnelling transistor[57,58]. The doping concentrations indicated in Fig. 1a,b were determined from calibrations during MBE.

Fabrication of ultra-scaled vertical nanowires

A 2.5-nm-thick layer of Si_3N_4 was deposited using plasma-enhanced chemical vapour deposition as the adhesion layer between III–V materials and the electron-beam resist. Hydrogen silsesquioxane (HSQ, 6%, Dow Corning) was spin-coated onto the samples. The spin-coating speed was 3,500 rpm and the duration time was 1 min, resulting in an HSQ film thickness of ~90 nm. Electron-beam lithography with a 125 kV acceleration voltage was carried out to define HSQ pillars. Using these as a hard mask, vertical nanowires were dry-etched using $Ar/BCl_3/SiCl_4$ chemistry in an inductively coupled plasma reactive-ion etching (RIE) etcher[30].

To selectively thin down the diameters of InAs and InAsSb with respect to GaSb in the vertical nanowires, a water solution of citric acid and H_2O_2 was used. One portion of 10% weight per volume (w/v) citric acid and two portions of 30% weight per weight (w/w) H_2O_2 were added into 40 portions of deionized water. The wet-etching rate for InAs and InAsSb was ~1 nm s^{-1}. Depending on the starting diameter of a vertical nanowire, different wet-etching times were employed for different samples. Subsequently, the samples were dipped into a deionized water bath for 30 s, followed by gentle N_2 blow drying.

Right after the citric acid wet-etching, one cycle of alcohol-based digital-etching[31,59] was performed on the samples to further thin down the diameter of the vertical nanowires and also to remove the III–V oxide layer covering the vertical nanowires. The digital-etching consisted of a 3 min exposure to pure O_2 and a 30 s 10% HCl:isopropyl alcohol wet-etching followed by rinsing and drying.

Fabrication of ultra-scaled vertical-nanowire devices

After fabricating the vertical nanowires, the samples were immediately (within 30 s) transferred into an atomic layer deposition chamber to minimize exposure to air. For the vertical-nanowire Esaki diodes, a conformal 3-nm-thick Al_2O_3 layer was deposited by atomic layer deposition at 250 °C with trimethylaluminium and water as precursors. For the vertical-nanowire tunnelling transistors, a 2.4-nm-thick Al_2O_3 layer was deposited at 250 °C with trimethylaluminium as the Al source and oxygen plasma as the O source. The thickness of the Al_2O_3 was confirmed using ellipsometry on Si dummy pieces that were processed at the same time as the III–V device samples. The Al_2O_3 thickness was also validated using scanning electron microscopy (SEM) from the radius increase of vertical nanowires on III–V dummy samples that were processed at the same time. At this point, the fabrication flow differed for the Esaki diodes and the tunnelling transistors (Supplementary Scheme 1).

For the Esaki diodes, FOx flowable oxide (FOx-16, Dow Corning) with a thickness of 490 nm was applied to planarize the vertical nanowires, which was followed by electron-beam lithography and CF_4-based RIE to open the vias for the drain and source contacts. A 10 nm Ni/30 nm Mo bilayer was sputtered for the ohmic contacts. Probe pads with 20 nm Ti/150 nm Au were fabricated using electron-beam lithography, electron-beam evaporation and a lift-off process. The lift-off was carried out using poly(methyl methacrylate) (950 A8, MicroChem). Using the Ti/Au probe pads as a hard mask, Mo was dry-etched with SF_6-based RIE, and Ni was wet-etched with Ni etchant (nickel etchant TFB, Transene). More details about Esaki diode fabrication can be found in our previous work[60].

For the tunnelling transistors, a 40-nm-thick W layer was conformally sputtered as a metal gate, which was followed by annealing in N_2/H_2 forming gas (N_2:H_2 = 0.92:0.08) at 310 °C for 30 min. The samples were then cooled down to room temperature in air. Subsequently, 490-nm-thick FOx-16 (interlayer dielectric 1, ILD1) was spin-coated to planarize the vertical nanowires, which was followed by iterative CF_4- and SF_6-based RIE to define the FOx-16 and W gate position along the vertical-nanowire sidewalls. We intentionally created an ~10 nm gate/drain underlap (Fig. 1f) to suppress the leakage current due to channel-to-drain tunnelling in the off state[61]. A second planarization layer with 145-nm-thick HSQ (ILD2) was applied, which was followed by CF_4-based RIE to sequentially open the vias for the drain, gate and source contacts. As for the Esaki diode fabrication, a Ni/Mo bilayer was sputtered for the ohmic contacts, which was followed by a lift-off of the Ti/Au probe pads and Ni/Mo etching.

All the devices studied in this work feature a single vertical nanowire with different diameters. The diameters of the vertical nanowires were measured using SEM after the atomic layer deposition. The active tunnelling junction dimension is determined by the smaller of the two diameters, InAs or GaSb, at the interface; in this work, for all devices, this was the InAs diameter.

Electrical characterization

Electrical measurements at room temperature were carried out in a probe station (Cascade) in air using a Keysight B1500A semiconductor parameter analyser with high-resolution source measurement units.

Low-temperature electrical measurements were performed in a cryogenic probe station (Lakeshore) in vacuum using a Keysight B1500A semiconductor parameter analyser with the high-resolution source measurement units. For these measurements, the pressure of the chamber was kept at around 10^{-6} Torr. The lowest temperature of 77 K was reached with liquid nitrogen. The temperature of the sample stage was controlled and monitored by a temperature controller (Lakeshore Model 340).

In all electrical measurements of Esaki diodes, we grounded the GaSb source and applied a bias to the InAs drain (V_{ds}). For the electrical measurements of the tunnelling transistors, we grounded the GaSb source and applied a bias to the W gate (V_{gs}) and the InAs drain (V_{ds}) separately.

Electron microscopy characterization

SEM images were acquired with a Zeiss GeminiSEM 450 and a Zeiss Sigma 300 SEM. The images of vertical nanowires shown in Figs. 1c,d were obtained with a tilt angle of 30°.

Lamellae of the heterostructure and the finished device were prepared in a Helios Nanolab 660 Dual Beam Focused Ion Beam Milling System using a Ga ion beam. The stage was cooled down with liquid nitrogen to suppress the reaction between III–V materials and the Ga beam. A carbon protection layer was deposited by an electron beam to protect the surface from ion-beam damage.

The STEM images and energy-dispersive X-ray spectroscopy results shown in Supplementary Figs. 1 and 2 were acquired with a Thermo Fisher Scientific Themis Z G3 aberration-corrected STEM and the equipped Super-X EDS system.

First-principles calculations

Density functional theory calculations. Geometric optimizations and electronic band structure calculations were carried out using density functional theory as implemented in Vienna ab initio Simulation Package (VASP)[62]. The generalized gradient approximation form of the Perdew–Burke–Ernzerhof exchange–correlation functional[63] was used with a Hubbard U correction (ref. 64). The projector augmented

wave method was used[65]. Structural relaxation was performed until the change of all forces on any atom was below 10^{-2} eV Å^{-1}. The energy cutoff was set to 250 eV. For InAs and GaSb, the optimal values of U have been found to be $U^{In,p} = -0.5$ eV, $U^{As,p} = -7.5$ eV, $U^{Ga,p} = 0.8$ eV and $U^{Sb,p} = -6.9$ eV (ref. 66).

In our calculations for the bulk-like GaSb/InAs structure (Supplementary Fig. 8a, left), we adopted the lattice constant of 6.096 Å for GaSb and placed an epitaxially matched InAs layer on top of the GaSb layer, as expected for our MBE-grown heterostructures. A GaAs-like interface was constructed, identical to what was observed in the STEM characterization, as shown in Supplementary Fig. 1. The simulated structures consist of 11 (15) atomic layers for GaSb (InAs; Supplementary Fig. 8b). A vacuum region of about 15 Å was added to the model to prevent spurious interactions between periodic images. The k-point mesh was sampled by the Monkhorst–Pack method with a separation of 0.2 Å^{-1} (ref. 67).

For the bulk-GaSb/confined-InAs structure (Supplementary Fig. 8a, right), due to the computational complexity, it was intractable to calculate with the nanowire geometry. Therefore, we adopted a confined InAs nanosheet with dangling bonds on the InAs sidewalls passivated by hydrogen atoms. The thickness of the InAs nanosheet was ~20 Å. Despite the geometrical difference between a nanowire and a nanosheet, the physics at the interface should be identical in both cases. The total numbers of layers for the two materials were identical to those in the bulk structure.

The red and blue dots in Supplementary Fig. 8c,d reflect the projection weight of the Bloch state to atoms in the InAs and GaSb regions, respectively. To avoid a contribution from the free surfaces (the two back surfaces of the slab, and for the confined case, the part of GaSb top surface not covered by InAs), four layers of atoms close to the surfaces were not included in the projection. The projection distinguishes bands from InAs and GaSb and provides the band alignment around the interface at an atomic scale.

Calculations with density functional theory coupled with non-equilibrium Green's functions. To include different doping types and doping levels in the InAs and GaSb regions, respectively, so that we could study the band alignment with and without quantum confinement under doping (not possible in VASP calculations), we performed electronic calculations using the Quantum Atomistix ToolKit (QuantumATK). These were based on density functional theory coupled with non-equilibrium Green's functions (NEGFs)[68]. The linear combination of atomic orbitals basis type, the generalized gradient approximation form of the Perdew–Burke–Ernzerhof exchange–correlation functional and the high basis set of the PseudoDojo pseudopotential were adopted[69]. The DFT-1/2 approach was used to correct self-interaction errors[70]. We used 4 × 4 × 100 k-points for the Brillouin zone integration within the Monkhorst–Pack scheme[67]. The density mesh cutoff was set as 105 Hartree in the charge density calculation.

In our calculations, we modelled the same structures as discussed in the previous 'Density functional theory calculations' section, but we added p- and n-type doping in GaSb and InAs regions, respectively, to a level that was close to the nominal doping levels of the experimental Esaki diode structure. We simulated the structures in a device configuration in thermal equilibrium, so that the Fermi levels of both the source and drain contacts was at 0, which was used as the energy reference. The temperature for both electrodes and the broadening of the electron Fermi distribution function were set to 300 K. The projected local DOS was calculated from the converged results. Due to wavefunction broadening, it was not straightforward to determine the band edges and the band alignment from only the projected local DOS spectra. Therefore, from the converged results, we also calculated the zero-bias transmission spectrum[71] so that we could determine the band edge at the two ends of the structures ($E_{c,InAs}$ at the source contact and $E_{v,InAs}$ at the drain contact), and we calculated the electrostatic difference potential so that we could determine the relative position between $E_{c,InAs}$ and $E_{v,InAs}$ right at the interface once we had obtained the band-edge positions at the two ends. The band edge in the electrode-central region interface was first read from the projected local DOS and then extended to the interface according to the average electrostatic difference potential over a unit cell. The calculated results are shown in Supplementary Fig. 9.

Device modelling and simulation

Two-dimensional device simulations. Two-dimensional simulations of circular cross sections (Supplementary Fig. 5a–c,e) that represent nanowires with an infinite length were carried out in nextnano[72] using the eight-band $\mathbf{k}\cdot\mathbf{p}$ method[73], which is suitable for III–V materials with a direct bandgap. Circular areas with different diameters of GaSb or InAs were considered. These were surrounded by a 3-nm-thick cubic Al_2O_3 layer with a relative dielectric permittivity of 8. To make the simulation results more accurate, we considered the penetration of wavefunctions into the dielectric. The material properties and $\mathbf{k}\cdot\mathbf{p}$ parameters used in the simulations were mostly adopted from ref. 74, with the exceptions noted in the caption of Supplementary Table 1. The conduction (E_c) and valence band (E_v) edges under quantization were defined as the lowest electron (conduction) and highest hole (valence) sub-band energies, respectively, at the Γ point in the Brillouin zone. Effective bandgap values were calculated as $E_g = E_c - E_v$. The electron effective mass (m_e^*) shown in Supplementary Fig. 5a was calculated from the dispersion relation of the lowest electron sub-band as $1/m_e^* = 1/\hbar\,\partial^2E/\partial k_z^2$, where $\hbar$ is the reduced Planck constant, E the electron energy and k_z the wavevector in the transport direction ([001] in this study). In the self-consistent Poisson–Schrödinger simulations (Supplementary Fig. 5b,c,e), the total number of simulated conduction sub-bands was 50, which was verified to be sufficient based on sub-band occupation. In certain simulations (Supplementary Fig. 5c,e), a fixed positive charge with a different charge density was also included in the oxide layer.

3D device simulations. Quantum-transport simulations (shown in Figs. 3a,d and 4b and Supplementary Figs. 7 and 10) were carried out on a home-built NEGF simulator developed by the French National Centre for Scientific Research and the University of Udine[75]. The numerical simulations comprised a 3D self-consistent solution to the electrostatic (Poisson) and transport equations (NEGF). The quantum-transport model was based on NEGFs. An eight-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian was used to describe the material properties[75]. The heterojunctions were described by adopting position-dependent $\mathbf{k}\cdot\mathbf{p}$ parameters in the Hamiltonian matrix[76]. The material parameters for InAs, GaSb and InSb were taken from ref. 74 and are reported in Supplementary Table 1. The $\mathbf{k}\cdot\mathbf{p}$ parameters for InAsSb were obtained by interpolating those of InAs and InSb using the bowing parameters recommended in ref. 74.

The slight lattice mismatch between GaSb and InAs was assumed to result in a tensile strain in the InAs layers, which we included in the model by employing the strain interaction matrix detailed in ref. 75 and by using the deformation potentials from ref. 74. Bandgap narrowing due to the high doping concentrations in GaSb and InAs(Sb) was also included according to the experimental values reported in refs. 77,78.

The 3D device structures were designed to closely resemble the fabricated devices sketched in Fig. 1e,f. For the tunnelling transistors, the diameter D and the length L along the transport direction z were L = 40 nm and D = 24 nm for the p-GaSb source (note that 24 nm was the largest possible diameter in our simulator), L = 15 nm and D = 6 nm for the undoped InAs channel, L = 45 nm and D = 6 nm for the undoped InAsSb channel, and L = 23 nm and D = 6 nm for the n-InAs drain. The oxide thickness and doping concentrations are reported in Supplementary Fig. 10a. A dielectric constant of 8 was used for Al_2O_3. To investigate the influence of the gate/source capacitive coupling, we modelled the oxide layer covering the source with different effective oxide thicknesses ($t_{ox,eff}$) of 3, and 12 nm, with or without a smaller dielectric constant $\varepsilon_{r,eff}$. For the simulation results shown in Fig. 3a,d,

the gate dielectric Al_2O_3 thickness was 2.4 nm, $t_{ox,eff}$ = 12 nm and $\varepsilon_{r,eff}$ = 1. For the Esaki diodes, all the parameter values were identical to those in the experimental device structure, except that D_{GaSb} = 22 nm.

The simulation procedure started with a 3D real-space discretization of the Hamiltonian operator throughout the device structure. The electron wavefunction was allowed to penetrate into the oxide region. In fact, a closed-boundary condition for the electron wavefunction at the oxide interface was expected to overestimate the sub-band quantization effects, in particular in the InAs region, where D_{InAs} = 6 nm and which has a very small electron effective mass[74]. Unfortunately, for relatively long and wide devices, the requirements for memory and the central processing unit tend to become intractable for a real-space representation. Hence, we softened the numerical burden by adopting the coupled mode-space approach[79]. In this respect, we used 70 modes for the valence band of GaSb and 50 modes for the conduction band of InAs(Sb) and verified that this gives results insensitive to a further increase in the number of modes. Inelastic phonon scattering was also included by using a local formulation of phonon self-energies within the self-consistent Born approximation[79]. Acoustic phonons were treated according to the elastic approximation, and polar phonons were described as dispersion-less optical phonons[79]. For all semiconductors, we employed the values of the deformation potentials reported in Supplementary Table 2. The value of the effective deformation potential in polar-optical-phonon scattering, D_{op}, was chosen to best reproduce the electrical measurements of vertical-nanowire GaSb/InAs Esaki diodes in an earlier study[60].

Literature benchmarking

The transistor active area was calculated as the area that the channel region occupied on a chip, that is, the channel width (W) times gate length (L_g) for planar transistors, the fin width (W_f) times gate length (L_g) for fin field-effect transistors, and the cross-sectional area of the active device channel in the transverse direction for vertical-transport devices. In this study, the active area was calculated using the cross-sectional area of the InAs portion, which determines the BTBT process. The transconductance reported in Fig. 5a,b was normalized with respect to the active area.

For devices in the literature, the maximum drain current $I_{d,max}$ (V_{ds} = 0.3 V), I_{on}, I_{off}, S_{min} and $g_{m,peak}$ for transistors were extracted from the published transfer characteristics, unless the corresponding values were explicitly reported. Similarly, the peak current I_{peak}, drain current I_d(V_{ds} = 0.3 V) and the PVCR of Esaki diodes were extracted from the published I–V characteristics unless the corresponding values were explicitly noted. In rare cases[22], in the absence of sufficient information, $I_{d,max}$(V_{ds} = 0.3 V) and $g_{m,peak}$ at V_{ds} = 0.3 V were extracted and estimated from the transfer characteristics at V_{ds} = 0.05 V. Therefore, some minor deviations from the actual figures of merit may have been introduced.

Data processing

Discrete trapping events were observed in the tunnelling transistors. This prevented the accurate extraction of key figures of merit, so for all data points corresponding to a measurement, the measured transfer characteristics were smoothed using the locally weighted scatter plot smoothing (LOWESS) method[80] with a smoothing window of 10%. The transconductance was calculated as $g_m = \partial I_d/\partial V_{gs}$ using the smoothed transfer characteristics. This was followed by another LOWESS smoothing with a smoothing window of 10% for all data points corresponding to a measurement of the transconductance characteristics (g_m versus V_{gs}). Therefore, the extracted $g_{m,peak}$ in this work represents a lower limit of the actual value. For the point-subthreshold-swing calculations at room temperature shown in Fig. 3b, we used the well-known Savitzky–Golay filter[81] with a window of five points to balance between removing noise and preserving the curve shape in the subthreshold regime. The point subthreshold swing of each device was then calculated as $S = \Delta V_{gs}/\Delta(\log_{10} I_d)$.

Data availability

Source data are provided with this paper. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Salahuddin, S., Ni, K. & Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **1**, 442–450 (2018).
2. Ionescu, A. M. Energy efficient computing and sensing in the zettabyte era: from silicon to the cloud. In *Proc. 2017 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 1.2.1–1.2.8 (IEEE, 2017).
3. Zhou, Z. et al. Edge intelligence: paving the last mile of artificial intelligence with edge computing. *Proc. IEEE* **107**, 1738–1762 (2019).
4. Datta, S., Chakraborty, W. & Radosavljevic, M. Toward attojoule switching energy in logic transistors. *Science* **378**, 733–740 (2022).
5. Mehonic, A. & Kenyon, A. J. Brain-inspired computing needs a master plan. *Nature* **604**, 255–260 (2022).
6. Horowitz, M. et al. Scaling, power, and the future of CMOS. In *Proc. 2005 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 7–15 (IEEE, 2005).
7. Kuhn, K. J. Considerations for ultimate CMOS scaling. *IEEE Trans. Electron Devices* **59**, 1813–1828 (2012).
8. Lundstrom, M. S. & Alam, M. A. Moore's law: the journey ahead. *Science* **378**, 722–723 (2022).
9. Ferain, I., Colinge, A. A. & Colinge, J.-P. Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors. *Nature* **479**, 310–316 (2011).
10. Frank, D. J. et al. Device scaling limits of Si MOSFETs and their application dependencies. *Proc. IEEE* **89**, 259–288 (2001).
11. Khan, A. I. et al. Negative capacitance in a ferroelectric capacitor. *Nat. Mater.* **14**, 182–186 (2015).
12. Cheema, S. et al. Ultrathin ferroic HfO_2–ZrO_2 superlattice gate stack for advanced transistors. *Nature* **604**, 65–71 (2022).
13. Qiu, C. et al. Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches. *Science* **361**, 387–392 (2018).
14. Shukla, N. et al. A steep-slope transistor based on abrupt electronic phase transition. *Nat. Commun.* **6**, 7812 (2015).
15. Lu, H. & Seabaugh, A. Tunnel field-effect transistors: state-of-the-art. *IEEE J. Electron Devices Soc.* **2**, 44–49 (2014).
16. Avci, U. E., Morris, D. H. & Young, I. A. Tunnel field-effect transistors: prospects and challenges. *IEEE J. Electron Devices Soc.* **3**, 88–95 (2015).
17. Esseni, D., Pala, M., Palestri, P., Alper, C. & Rollo, T. A review of selected topics in physics based modeling for tunnel field-effect transistors. *Semicond. Sci. Technol.* **32**, 083005 (2017).
18. Luisier, M. & Klimeck, G. Performance comparisons of tunneling field-effect transistors made of InSb, carbon, and GaSb-InAs broken gap heterostructures. In *Proc. 2009 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 37.6.1–37.6.4 (IEEE, 2009).
19. Brocard, S., Pala, M. G. & Esseni, D. Design options for hetero-junction tunnel FETs with high on current and steep sub-threshold voltage slope. In *Proc. 2013 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 5.4.1–5.4.4 (IEEE, 2013).
20. Convertino, C. et al. A hybrid III–V tunnel FET and MOSFET technology platform integrated on silicon. *Nat. Electron.* **4**, 162–170 (2021).
21. Krishnaraja, A. et al. Tuning of source material for InAs/InGaAsSb/GaSb application-specific vertical nanowire tunnel FETs. *ACS Appl. Electron. Mater.* **2**, 2882–2887 (2020).

22. Memisevic, E., Svensson, J., Lind, E. & Wernersson, L.-E. Vertical nanowire TFETs with channel diameter down to 10 nm and point S_{MIN} of 35 mV/decade. *IEEE Electron Device Lett.* **39**, 1089–1091 (2018).
23. Alian, A. et al. Record 47 mV/dec top-down vertical nanowire InGaAs/GaAsSb tunnel FETs. In *Proc. 2018 IEEE Symposium on VLSI Technology* 133–134 (IEEE, 2018).
24. Zhao, X., Vardi, A. & del Alamo, J. A. Sub-thermal subthreshold characteristics in top-down InGaAs/InAs heterojunction vertical nanowire tunnel FETs. *IEEE Electron Device Lett.* **38**, 855–858 (2017).
25. Memisevic, E., Svensson, J., Hellenbrand, M., Lind, E. & Wernersson, L.-E. 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 *Proc. 2016 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 19.1.1–19.1.4 (IEEE, 2016).
26. Bijesh, R. et al. Demonstration of $In_{0.9}Ga_{0.1}As/GaAs_{0.18}Sb_{0.82}$ near broken-gap tunnel FET with I_{ON} = 740 μA/μm, G_M = 700 μS/μm and gigahertz switching performance at V_{DS} = 0.5 V. In *Proc. 2013 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 28.2.1–28.2.4 (IEEE, 2013).
27. Dey, A. W. et al. High-current GaSb/InAs(Sb) nanowire tunnel field-effect transistors. *IEEE Electron Device Lett.* **34**, 211–213 (2013).
28. Yakimets, D. et al. Vertical GAAFETs for the ultimate CMOS scaling. *IEEE Trans. Electron Devices* **62**, 1433–1439 (2015).
29. Shao, Y., Pala, M., Esseni, D. & del Alamo, J. A. Sub-10-nm diameter GaSb/InAs vertical nanowire Esaki diodes with ideal scaling behavior: experiments and simulations. In *Proc. 2021 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 32.1.1–32.1.4 (IEEE, 2021).
30. Zhao, X. & del Alamo, J. A. Nanometer-scale vertical-sidewall reactive ion etching of InGaAs for 3-D III-V MOSFETs. *IEEE Electron Device Lett.* **35**, 521–523 (2014).
31. Lu, W. et al. 10-nm fin-width InGaSb p-channel self-aligned FinFETs using antimonide-compatible digital etch. In *Proc. 2017 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 17.7.1–17.7.4 (IEEE, 2017).
32. Pawlik, D. et al. Benchmarking and improving III-V Esaki diode performance with a record 2.2 MA/cm^2 peak current density to enhance TFET drive current. In *Proc. 2012 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 27.1.1–27.1.3 (IEEE, 2012).
33. Bhatnagar, K. et al. Integration of broken-gap heterojunction InAs/GaSb Esaki tunnel diodes on silicon. *J. Vac. Sci. Technol. B* **33**, 062203 (2015).
34. Convertino, C. et al. Investigation of InAs/GaSb tunnel diodes on SOI. In *Proc. 017 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)* 148–151 (IEEE, 2017).
35. Cutaia, D. et al. Complementary III–V heterojunction lateral NW tunnel FET technology on Si. In *Proc. 2016 Symposium on VLSI Technology* 1–2 (IEEE, 2018).
36. Borg, B. M. et al. Influence of doping on the electronic transport in GaSb/InAs(Sb) nanowire tunnel devices. *Appl. Phys. Lett.* **101**, 043508 (2012).
37. Afzalian, A. et al. Physics and performances of III–V nanowire broken-gap heterojunction TFETs using an efficient tight-binding mode-space NEGF model enabling million-atom nanowire simulations. *J. Phys.: Condens. Matter* **30**, 254002 (2018).
38. Yan, R. et al. Esaki diodes in van der Waals heterojunctions with broken-gap energy band alignment. *Nano Lett.* **9**, 5791–5798 (2015).
39. Sajjad, R. N., Chern, W., Hoyt, J. L. & Antoniadis, D. A. Trap assisted tunneling and its effect on subthreshold swing of tunnel FETs. *IEEE Trans. Electron Devices* **63**, 4380–4387 (2016).
40. Zhao, X. et al. Sub-10 nm diameter InGaAs vertical nanowire MOSFETs. In *Proc. 2017 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 17.2.1–17.2.4 (IEEE, 2017).
41. Memisevic, E. et al. Individual defects in InAs/InGaAsSb/GaSb nanowire tunnel field-effect transistors operating below 60 mV/decade. *Nano Lett.* **17**, 4373–4380 (2017).
42. Pala, M. G. & Esseni, D. Interface traps in InAs nanowire tunnel-FETs and MOSFETs. Part I. Model description and single trap analysis in tunnel-FETs. *IEEE Trans. Electron Devices* **60**, 2795–2801 (2013).
43. 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 *Proc. 2017 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 29.1.1–29.1.4.44 (IEEE, 2017).
44. Lu, W. et al. First transistor demonstration of thermal atomic layer etching: InGaAs FinFETs with sub-5 nm fin-width featuring in situ ALE-ALD. In *Proc. 2018 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 39.1.1–39.1.4 (IEEE, 2018).
45. Esaki, L. New phenomenon in narrow germanium p–n junctions. *Phys. Rev.* **109**, 603–604 (1958).
46. Wang, Z., Wang, L., En, Y. & Jiang, X.-W. Quantum confinement modulation on the performance of nanometer thin body GaSb/InAs tunnel field-effect transistors. *J. Appl. Phys.* **121**, 224503 (2017).
47. Na, J. et al. Low-frequency noise in multilayer MoS_2 field-effect transistors: the effect of high-k passivation. *Nanoscale* **6**, 433–441 (2014).
48. Kilpi, O.-P., Svensson, J. & Wernersson, L.-E. Sub-100-nm gate-length scaling of vertical InAs/InGaAs nanowire MOSFETs on Si. In *Proc. 2017 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 17.3.1–17.3.4 (IEEE, 2017).
49. Tomioka, K., Yoshimura, M. & Fukui, T. A III–V nanowire channel on silicon for high-performance vertical transistors. *Nature* **488**, 189–192 (2012).
50. Tomioka, K., Gamo, H., Motohisa, J. & Fukui, T. Vertical gate-all-around tunnel FETs using InGaAs nanowire/Si with core-multishell structure. In *Proc. 2020 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 21.1.1–21.1.4 (IEEE, 2020).
51. Sarkar, D. et al. A subthermionic tunnel field-effect transistor with an atomically thin channel. *Nature* **526**, 91–95 (2015).
52. Knoll, L. et al. Inverters with strained Si nanowire complementary tunnel field-effect transistors. *IEEE Electron Device Lett.* **34**, 813–815 (2013).
53. Krishnamohan, T., Kim, D., Raghunathan, S. & Saraswat, K. Double-gate strained-Ge heterostructure tunneling FET (TFET) with record high drive currents and ≪60 mV/dec subthreshold slope. In *Proc. 2008 IEEE International Electron Devices Meeting (IEDM) Technical Digest* 1–3 (IEEE, 2008).
54. Kim, S. et al. Thickness-controlled black phosphorus tunnel field-effect transistor for low-power switches. *Nat. Nanotechnol.* **15**, 203–206 (2020).
55. Rangasamy, G., Zhu, Z. & Wernersson, L.-E. High current density vertical nanowire TFETs with I_{60} > 1 μA/μm. *IEEE Access* **11**, 95692–95696 (2023).
56. Vandenberghe, W. G. et al. Figure of merit for and identification of sub-60 mV/decade devices. *Appl. Phys. Lett.* **102**, 013510 (2013).
57. Sharma, A., Goud, A. A. & Roy, K. GaSb-InAs n-TFET with doped source underlap exhibiting low subthreshold swing at sub-10-nm gate-lengths. *IEEE Electron Device Lett.* **35**, 1221–1223 (2014).
58. Afzalian, A., Passlack, M. & Yeo, Y.-C. Atomistic simulation of gate-all-around GaSb/InAs nanowire TFETs using a fast full-band mode-space NEGF model. In *Proc. 2016 IEEE Symposium on VLSI Technology* 1–2 (IEEE, 2016).

59. Lu, W., Zhao, X., Choi, D., El Kazzi, S. & del Alamo, J. A. Alcohol-based digital etch for III–V vertical nanowires with sub-10 nm diameter. *IEEE Electron Device Lett.* **38**, 548–551 (2017).
60. Shao, Y., Pala, M., Esseni, D. & del Alamo, J. A. Scaling of GaSb/InAs vertical nanowire Esaki diodes down to sub-10-nm diameter. *IEEE Trans. Electron Devices* **69**, 2188–2195 (2022).
61. Krishnaraja, A., Svensson, J., Lind, E. & Wernersson, L.-E. Reducing ambipolar off-state leakage currents in III-V vertical nanowire tunnel FETs using gate-drain underlap. *Appl. Phys. Lett.* **115**, 143505 (2019).
62. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
63. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
64. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. *Phys. Rev. B* **57**, 1505–1509 (1998).
65. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
66. Yu, M., Yang, S., Wu, C. & Marom, N. Machine learning the Hubbard *U* parameter in DFT+U using Bayesian optimization. *npj Comput. Mater.* **6**, 180 (2020).
67. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
68. Smidstrup, S. et al. QuantumATK: an integrated platform of electronic and atomic-scale modelling tools. *J. Phys.: Condens. Matter* **32**, 015901 (2020).
69. van Setten, M. J. et al. The PseudoDojo: training and grading a 85 element optimized norm-conserving pseudopotential table. *Comput. Phys. Commun.* **226**, 39–54 (2018).
70. Tao, S. X., Cao, X. & Bobbert, P. A. Accurate and efficient band gap predictions of metal halide perovskites using the DFT-1/2 method: GW accuracy with DFT expense. *Sci. Rep.* **7**, 14386 (2017).
71. Çakır, D. & Peeters, F. M. Dependence of the electronic and transport properties of metal-$MoSe_2$ interfaces on contact structures. *Phys. Rev. B* **89**, 245403 (2014).
72. Birner, S. et al. nextnano: general purpose 3-D simulations. *IEEE Trans. Electron Devices* **54**, 2137–2142 (2007).
73. Cardona, M. & Pollak, F. H. Energy-band structure of germanium and silicon: the k·p method. *Phys. Rev.* **142**, 530–543 (1966).
74. Vurgaftman, I., Meyer, J. R. & Ram-Mohan, L. R. Band parameters for III–V compound semiconductors and their alloys. *J. Appl. Phys.* **89**, 5815 (2001).
75. Bahder, T. B. Eight-band k·p model of strained zinc-blende crystals. *Phys. Rev. B* **41**, 11992–12001 (1990).
76. Brocard, S., Pala, M. & Esseni, D. Large on-current enhancement in hetero-junction tunnel-FETs via molar fraction grading. *IEEE Electron Device Lett.* **35**, 184–186 (2014).
77. Jain, S. C., McGregor, J. M. & Roulston, D. J. Band-gap narrowing in novel III-V semiconductors. *J. Appl. Phys.* **68**, 3747–3749 (1990).
78. Stollwerck, G., Sulima, O. V. & Bett, A. W. Characterization and simulation of GaSb device-related properties. *IEEE Trans. Electron Devices* **47**, 448–457 (2000).
79. Conzatti, F., Pala, M., Esseni, D., Bano, E. & Selmi, L. Strain-induced performance improvements in InAs nanowire tunnel FETs. *IEEE Trans. Electron Devices* **59**, 2085–2092 (2012).
80. Cleveland, W. S. Robust locally weighted fitting and smoothing scatterplots. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
81. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).

Acknowledgements

This work was supported by Intel Corporation and MIT Donner Endowed Chair. B.M. and J.L. acknowledge support from DTRA (Award No. HDTRA1-20-2-0002) through the Interaction of Ionizing Radiation with Matter University Research Alliance. Device fabrication was carried out at MIT.nano, MTL and SEBL of MIT. We thank G. Dewey, U. E. Avci and A. V. Penumatcha at Intel Corporation for fruitful discussions. We thank M. Robinson at Intel Corporation for SIMS measurements. We thank X. Zhao, W. Lu and A. Vardi for insightful suggestions. We thank J. Zhu, Q. Xie and T. Zeng for helpful suggestions during device fabrication and measurements. We thank Z. Hu and P. Zhang for discussions on the physical properties of InAs.

Author contributions

Y.S. and J.A.d.A. designed the study. Y.S. fabricated the devices and performed SEM imaging, electrical characterization and two-dimensional device simulations. M.P. and D.E. carried out the NEGF-based 3D quantum-transport simulations. H.T. and Y.S. carried out the first-principles calculations supervised by J.L. B.W. performed the FIB and TEM imaging of materials and devices supervised by J.L. Y.S. and J.A.d.A. wrote the manuscript. All authors read and revised the manuscript in its current form.

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/s41928-024-01279-w.

Correspondence and requests for materials should be addressed to Yanjie Shao, Marco Pala, Ju Li, David Esseni or Jesús A. del Alamo.

Peer review information *Nature Electronics* thanks Aryan Afzalian and Cezar Zota for their contribution to the peer review of this work.

Reprints and permissions information is available at 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.

© The Author(s), under exclusive licence to Springer Nature Limited 2024