

The van der Waals Gap: a Hidden Showstopper in Semiconductor Device Scaling

Mahdi Pourfath^{1*} and Tibor Grasser^{1*}

¹Institute for Microelectronics, TU Wien, Gusshausstrasse 27-29/E360, 1040 Vienna, Austria.

*Corresponding author. Email: pourfath@iue.tuwien.ac.at and grasser@iue.tuwien.ac.at

Continued miniaturization of transistors is critical for sustaining advances in computing performance, energy efficiency, and integration density. One of the most critical challenges at the nanoscale is controlling gate leakage through ultrathin dielectrics. In the search for suitable insulators, their permittivity and bandgap appear to be the most commonly considered performance indicators. However, while two-dimensional (2D) semiconductors provide ultimate electrostatic control, their interfaces with gate dielectrics often form a van der Waals (vdW) gap. Interestingly, despite its widely acknowledged presence, the electronic properties of this vdW gap and its impact on device performance have not received the attention they require. First-principles calculations and analytical modeling supported by experimental data, indicate that typical vdW gaps measure around 1.4 Å and exhibit a low dielectric constant of approximately 2, effectively adding about 2.7 Å to the equivalent oxide thickness (EOT). Although the vdW gap acts as an additional tunneling barrier and reduces gate leakage currents by about one to two orders of magnitude, it also contributes parasitic capacitance that can offset the advantages of high-permittivity dielectrics. To support material selection, a dimensionless figure of merit is introduced that integrates dielectric screening, tunneling suppression, and the thickness-dependent permittivity of ultrathin oxides, offering a predictive framework for identifying the minimum achievable EOT for a certain gate leakage current in the presence of a vdW gap. Although

certain materials may benefit from vdW gaps, the results presented here demonstrate that they frequently impose serious constraints on further device scaling. In particular, our results show that due to the presence of the vdW gap, most currently considered insulators will not be scaleable down to an EOT of 5Å as required by the IRDS roadmap for future device nodes. As a potential alternative, zippered structures are explored, in which quasi-covalent bonding between two-dimensional layers eliminates the vdW gap entirely while avoiding the formation of dangling bonds.

Introduction

Two-dimensional (2D) semiconductors, such as MoS₂ and other layered materials, are intensively explored for ultra-scaled field-effect transistors (FETs) (1). Achieving sub-1 nm equivalent oxide thickness (EOT) in the gate stack is essential for continued scaling beyond 2030, as outlined in the International Roadmap for Devices and Systems (IRDS) (2). This requires gate insulators with high permittivity (κ) and large band offsets to suppress leakage currents (3). Conventional screening of candidate dielectrics typically focuses on bulk properties such as dielectric constant and band gap, while assuming ideal, abrupt interfaces (4, 5), as illustrated in Fig. 1(A). However, in real devices, several non-ideal interfacial phenomena—such as reduced effective permittivity in thin films (particularly due to interface-induced dead layers (6)), gaps between the insulator and the channel, interface dipoles, and remote phonon scattering in high- κ materials (7)—as illustrated in Fig. 1(B), can significantly alter benchmarking conclusions.

One key limitation arises from interfacial dead layers (6), which are commonly observed in high- κ insulators. Structural or chemical perturbations—such as strain, defects, or incomplete crystallinity—can induce thin regions with significantly reduced permittivity. As shown in Fig. 1(C), these dead layers substantially increase the minimum achievable EOT for a given maximum leakage current by degrading the effective permittivity at small thicknesses. Another critical factor is the weak van der Waals (vdW) interaction (8) between 2D semiconductors and deposited insulators, which creates an interfacial vacuum-like gap that is often overlooked. Due to small but non-zero charge redistribution, this gap exhibits weak polarization and an effective dielectric constant

significantly lower than that of the insulator. The vdW gap plays a dual role: it increases EOT by introducing a low- κ series layer, but also suppresses tunneling by acting as an additional barrier. As shown in Fig. 1(C), whether the vdW gap increases or reduces the minimum achievable EOT depends on which of these two effects dominates.

Although the vdW gap may seem negligible in size, this study shows that even a sub-nanometer gap can substantially impact device performance. The tunneling current through a gate stack decreases exponentially with both the barrier height and the insulator thickness, scaled by its permittivity. In the Wentzel–Kramers–Brillouin (WKB) approximation, this exponential dependence is captured by the inverse decay length β , which depends on the tunneling effective mass and the conduction band offset ΔE between the insulator and the channel (for nFETs), following $\beta \propto \sqrt{m^* \Delta E}$, with m^* being the effective tunneling mass and ΔE the height of the tunneling barrier (details in the supplementary text, Section S4). To facilitate the screening and benchmarking of potential dielectrics, the insulator figure of merit (FoM) (9) is defined as the ratio of $\varepsilon_r \beta$ for a candidate insulator relative to SiO_2 (definition in the supplementary text, Section S5):

$$\text{FoM} = \frac{\varepsilon_{\text{ins}} \beta_{\text{ins}}}{\varepsilon_{\text{SiO}_2} \beta_{\text{SiO}_2}}.$$

This metric quantifies how effectively an insulator can suppress tunneling while maintaining strong electrostatic control, relative to SiO_2 . A higher FoM indicates lower leakage for the same EOT, enabling more aggressive scaling. Fig. 1(D) shows the idealized case where a large permittivity is assumed to directly reduce EOT. However, as shown in the following sections, the presence of a vdW gap can significantly shift this picture and alter benchmarking outcomes.

A clear example showing that a high- κ (high-FoM) insulator does not necessarily result in a low EOT is the SrTiO_3 (STO) stack. Fig. 1(E) schematically shows an STO– MoS_2 heterostructure, where interfacial dead layers form near the boundaries. Their impact is captured by the degradation factor D , defined in eq. S21, which introduces a thickness-dependent correction to the effective permittivity and scales with the dead-layer thickness (6). As the physical thickness of the insulator decreases, the relative impact of these regions becomes more pronounced. Although bulk STO exhibits a dielectric constant as high as $\kappa \approx 270$, ultrathin films fall short of their ideal electrostatic performance because a significant portion of the electric field is dropped across the low- κ interfacial regions. As shown in Fig. 1(F), the macroscopic, planar-averaged electrostatic potential in bulk STO remains nearly

flat under an applied field, consistent with strong dielectric screening. In contrast, strong potential variations near the interfaces and across the vdW gap reveal regions of weakened screening. The layer-resolved density of states (DOS), shown in fig. S5, indicates electronic interactions between sulfur atoms in MoS₂ and the surface layers of STO. These interactions alter the interfacial DOS and correlate with an enhanced potential drop. Furthermore, Fig. 1(G) shows that the charge density within the vdW gap determines its polarizability under an electric field, resulting in a small effective permittivity—slightly higher than that of vacuum.

Fig. 1(E) summarizes these effects using an equivalent circuit model, where both the vdW gap and dead layers act as series capacitive bottlenecks that limit the minimum achievable EOT. To estimate their contribution quantitatively, the approximation in eq. S26 can be used. For a degradation parameter of $D = 1.6 \text{ \AA}$, representative of interfacial dead layers in STO (10), the corresponding areal capacitance is approximately $C_{\text{DL}}^{\text{tot}} \approx \epsilon_0/D \approx 5.5 \mu\text{F}/\text{cm}^2$. For the vdW gap between STO and MoS₂, assuming an average thickness of $\bar{t}_{\text{vdW}} = 1.4 \text{ \AA}$ and an effective permittivity of $\epsilon_{\text{vdW}} = 2.7$ (see table S1), the resulting areal capacitance is $C_{\text{vdW}} = \epsilon_{\text{vdW}}/\bar{t}_{\text{vdW}} \approx 17.0 \mu\text{F}/\text{cm}^2$.

vdW Gap Characteristics

The vdW gap and its physical origin are first characterized using ab initio calculations, revealing that the binding energy of a vdW–bonded interface between a 2D semiconductor channel and an insulator is typically one to two orders of magnitude smaller than that of a fully covalent interface, as shown in Fig. 2(A). For example, layered dielectrics such as hexagonal boron nitride (hBN) or oxide crystals like STO adhere to MoS₂ with typical surface binding energies of 15–30 meV/Å² (11), whereas a covalent Si–SiO₂ interface exhibits binding energies on the order of eV/Å². This weaker vdW bonding results in a larger equilibrium separation between materials, as illustrated in Fig. 2(B). This separation creates a vacuum-like region between solids, where the crystal wavefunctions decay into evanescent tails. However, not all interfaces are purely vdW or purely covalent. For instance, the STO–MoS₂ interface appears more strongly vdW bonded than hBN–MoS₂, while certain structures such as β -Bi₂SeO₅–Bi₂O₂Se (BSO–BOS) form so-called zippered interfaces (12). These zipper-like bonds fall between the extremes of purely vdW and fully covalent bonding and are of particular

interest because they may enable strong interfacial adhesion without introducing a vdW gap or creating dangling bonds—an essential characteristic for scalable device integration. Fig. 2(A) and Fig. 2(B) contrast vdW-bonded and covalent interfaces, showing both the lower binding energy and the more uniform interfacial charge in vdW systems.

The vdW gap, t_{vdW} , is defined as the distance between adjacent atomic planes minus the sum of their covalent radii, as shown in Fig. 2(A). Using ab initio calculations, the equilibrium positions and relative distances between neighboring atoms in a vdW stack can be determined. By subtracting the corresponding covalent radii from these interatomic distances, the vdW gap can be extracted. For example, as reported in table S1, an average value of $\bar{t}_{\text{vdW}} \approx 1.4 \text{ \AA}$ is obtained for the studied insulator–MoS₂ stacks. An alternative approach to estimating the vdW gap is to use reported average values of vdW and covalent radii, as illustrated in Fig. 2(C). The difference between these two radii provides an approximate vdW gap for each element. By considering all possible heterostructure combinations of these elements, the statistical distribution of vdW gaps is shown in Fig. 2(D). Consistent with the values reported in table S1, the average vdW gap across more than 4,000 heterostructure combinations falls within the range of 1–2 Å, with a mean value of $1.40 \pm 0.22 \text{ \AA}$.

Dielectric Properties of vdW Gaps

The vdW gap acts as a low-permittivity interfacial layer, significantly reducing the out-of-plane dielectric response of 2D heterostructures. Its dielectric properties were extracted from first-principles simulations, as described in Materials and Methods and Supplementary Text (Section S1).

Fig. 3(A)-(C) focuses on an hBN-MoS₂ heterostructure (a widely studied 2D gate stack (13)) and reveals how the presence of the vdW gap reduces the effective dielectric response. For reference, the responses of isolated hBN and isolated monolayer MoS₂ are plotted as well. A clear enhancement of polarization in the gap region is observed [Fig. 3 (B)] in the heterostructure compared to the isolated cases, corresponding to a non-uniform $\varepsilon(z)$ that drops to a minimum in the middle of the vdW gap [Fig. 3(C)]. Physically, the gap behaves nearly like vacuum, with only modest polarizability from evanescent electronic tails near the interfaces. Treating the vdW gap as a thin dielectric slab and integrating the spatial permittivity as a series capacitance yields an effective dielectric constant (see eq. S15). For the hBN–MoS₂ stack, this yields an effective permittivity of $\varepsilon_{\text{vdW}}^{\text{eff}}/\varepsilon_0 \approx 1.7$.

Across a range of representative systems, the average value is $\bar{\epsilon}_{\text{vdW}}/\epsilon_0 \approx 2$ (see table S1). Such low- κ interfacial layers hinder electrostatic coupling and fundamentally limit dielectric scaling by increasing the EOT, even in the presence of high- κ materials.

A practical consequence of the vdW gap's low permittivity is that it introduces a nearly constant series thickness to the total EOT, regardless of the bulk insulator. Various insulator–MoS₂ interfaces were analyzed using DFT-relaxed geometries. The equilibrium vdW gap thickness, t_{vdW} , varies moderately (typically within ± 0.2 Å), as shown in Fig. 3(D), and the extracted permittivity values exhibit a clear trend with t_{vdW} . This trend is accurately captured by the analytical model $\epsilon_{\text{vdW}}^{-1} = 1 - c/t_{\text{vdW}}$, derived in eq. S5, with c being a fitting parameter. The model provides a reliable means to estimate how non-equilibrium interface spacing (e.g., due to fabrication-induced strain) may affect ϵ_{vdW} . Based on this model and an average fitted value of $\bar{c} = 0.72$ Å, the vdW gap contributes an average dielectric constant of $\bar{\epsilon}_{\text{vdW}} \approx 2$ and an equivalent oxide thickness of $\overline{\text{EOT}}_{\text{vdW}} \approx 2.7$ Å, as shown in Fig. 3(E) and listed in table S1. These results indicate that over a quarter nanometer of EOT is intrinsically lost to the vdW gap in typical 2D transistors—a substantial penalty when striving for sub-nanometer scaling, given the projected 0.5 nm total EOT target in the 2035 IRDS roadmap (2).

While vdW gaps appear at 2D channel–insulator interfaces, they also exist within layered materials themselves, where internal gaps between atomic planes reduce the bulk dielectric constant. Fig. 3(C) illustrates that in multilayer hBN and MoS₂, the local dielectric constant peaks near atomic layers and drops in the vdW gaps between layers. A simple analytical model treats a layered crystal as a sequence of alternating high- κ slabs (representing atomic planes) and low- κ vdW gaps. For example, an N -layer MoS₂ stack can be modeled as two half-layers at the surfaces plus $(N - 1)$ internal layer-gap bilayers arranged in series (see eq. S16). This model reproduces the increase of the out-of-plane dielectric constant with N and its convergence toward the bulk limit for $N \rightarrow \infty$, as shown in Fig. 3(F). A reduction in the interlayer gap—e.g., due to strain or pressure—is predicted to significantly enhance the effective dielectric constant. Interestingly, one experimental study (14), also shown in Fig. 3(F), reported an anomalously large out-of-plane dielectric constant in few-layer MoS₂, possibly due to unintentional strain that narrows the vdW gaps.

Tunneling Through vdW Gaps

Since the vdW gap is essentially a nanoscale vacuum barrier, this has a profound impact on electron tunneling between the channel and the gate. Both analytical WKB estimates within the Tsu-Esaki tunneling model and explicit quantum transport simulations illustrate this effect, as shown in Fig. 4(A) and Fig. 4(B). A representative graphene–hBN–graphene stack serves as a model system to benchmark the tunneling model against experimental conditions. In this structure, the planar-averaged electrostatic potential across the vdW gap closely approaches the vacuum level, as shown in Fig. 4(B), resulting in an effective barrier height for electrons tunneling into graphene that is approximately equal to its work function (about 4.5 eV (15)).

To analytically estimate the tunneling probability, the WKB approximation is applied. For thin insulators, coherent transport can be assumed, and the total transmission probability is expressed as the product of the individual contributions from the insulator and the vdW gap: $T = T_{\text{ins}}T_{\text{vdW}}$. The transmission through the vdW gap is given by $T_{\text{vdW}} = \exp(-2\alpha\sqrt{m_0\Delta E} t_{\text{vdW}}/\hbar)$, where a free electron mass m_0 is assumed for electrons traversing the vdW gap due to the absence of crystal periodicity. The parameter α represents a shape factor that accounts for deviations from an ideal rectangular potential barrier. While $\alpha = 1$ corresponds to an ideal rectangular profile, smoother barriers arising from electrostatic or band-bending effects can reduce α to values as low as 2/3. A representative value of $\alpha \approx 0.8$ is adopted for typical vdW gap geometries.

For the graphene–hBN–graphene system, the DFT-calculated vdW gap is approximately 1.6 Å, yielding a single-gap transmission probability of about $T_{\text{vdW}} = 6.2\%$. Since there are two vdW gaps on the opposing sides of the hBN film, the overall transmission is approximately the product of the two, leading to an estimated ~ 260 -fold suppression of the tunneling current. This substantial reduction occurs even though the vdW gap is atomically thin, because its barrier height is nearly as high as the vacuum level. Fig. 4(A) confirms this effect: *ab initio* non-equilibrium Green's function (NEGF) quantum transport simulations show that the local density of states (LDOS) decays sharply within hBN and drops significantly at the graphene–hBN interfaces, particularly near the Fermi level of graphene. The calculated transmission remains consistent with the WKB estimate.

The proposed model is compared to experimental data on tunneling through 2D insulators. Fig. 4(C) shows the current–voltage characteristics for graphene–hBN–graphene structures with

one to four layers of hBN as the tunneling barrier. The solid lines (model) and symbols (experiment) show excellent agreement across all thicknesses when the effect of the vdW gaps is included. Without accounting for the gaps, the measured resistance versus hBN thickness is under-predicted by orders of magnitude. Including the two vdW gaps in the simulation increases the calculated resistance by nearly 260×, matching the experimental trend [Fig. 4(D)]. A slight deviation remains for monolayer hBN at high bias, but this is resolved by including a small series resistance in the model (red dashed line in Fig. 4(C)), which accounts for voltage drops in the contacts at high current levels.

Fig. 4(E) compares the developed model with experimental data for a gate stack composed of a monolayer MoS₂ channel, an STO insulator, and an Au metal gate. Since MoS₂ is the channel material in most 2D transistors considered here, its electron affinity (~ 4.3 eV) determines the effective tunneling barrier into the conduction band. Applying a simple WKB model across a vdW gap of $\sim 1\text{--}2$ Å, this barrier yields a transmission probability of only $\sim 3\text{--}20\%$, depending on the gap width and the barrier shape. A single vdW gap at the STO–MoS₂ interface, estimated in this work to be approximately 1.4 Å, results in a transmission probability of $T_{\text{vdW}} \approx 9.1\%$. Although the details of the Au–STO interface are not provided in the experimental data, assuming a transmission probability of 20%—representative of the smallest vdW gaps—yields excellent agreement with experiment.

Another relevant scenario is the metal–2D semiconductor contact, which plays a critical role in enabling low source and drain contact resistances to the channel. In the absence of any vdW gap, such contacts can approach the quantum resistance limit, given by $R_{\text{cq}}W = \frac{\hbar}{2e^2} \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{g_v n_{2D}}}$ (16), where the valley degeneracy $g_v = 2$ for monolayer MoS₂ is due to its two equivalent valleys in the Brillouin zone. However, when a metal electrode is placed on a 2D semiconductor, a vdW gap of a few angstroms typically exists, unless specific processes such as surface treatments or interfacial alloying are used to promote covalent bonding and eliminate the vdW gap.

Quantum contact resistance for a metal–MoS₂ interface was evaluated across a range of vdW gap thicknesses and shape factor values α from 2/3 to 1, as shown in Fig. 4(F). The vdW gap range considered here is slightly larger than the typical equilibrium values discussed earlier, because in practice, many device fabrication methods (such as transfer printing of contacts) can inadvertently introduce larger-than-equilibrium gaps at interfaces, making this issue particularly relevant. It should also be noted that for wider vdW gaps, the barrier profile tends to become more rectangular, making

shape factor values closer to $\alpha = 1$ more appropriate. The analyses conducted here reveal that even a modest vdW gap thickness can result in orders-of-magnitude increases in contact resistance. The model aligns well with experimental data for purely vdW-bonded metal–MoS₂ contacts reported in the literature. The key takeaway is that the tunneling barrier introduced by the vdW gap must be accounted for to accurately predict current flow in 2D heterostructure devices.

Scaling Limits Imposed by vdW Gaps

Integration of the electrostatic and tunneling effects of vdW gaps enables evaluation of their impact on the scaling limits of various gate insulators. Fig. 5(A) summarizes the competing effects, showing leakage current versus EOT curves for two example dielectrics, hBN and STO, with MoS₂ as the channel. The calculated J –EOT relationship for hBN–MoS₂ is presented under three different scenarios: (i) ignoring the vdW gap entirely (only the bulk insulator considered), (ii) including the vdW gap’s tunneling barrier effect but not its capacitance penalty, and (iii) including both the barrier and the dielectric cost of the vdW gap (realistic model). In the case of hBN (low- κ , wide-band-gap) (17), the dotted curve (tunneling only) shows significantly reduced leakage compared to the dashed curve (no gap), thanks to the additional barrier. This shift enables scaling to a smaller EOT for a given J_{target} . However, when the finite permittivity of the gap is accounted for (solid curve), the EOT cannot actually reach competitive values: the minimum achievable EOT in the solid curve is set by the fact that about 3.5 Å of EOT come from the vdW gap itself. In the case of hBN, the net outcome remains positive: the lowest attainable EOT with the gap included is slightly smaller than it would be without the gap, meaning the leakage suppression outweighs the capacitance drawback.

For STO (ultra-high κ , moderate band offset), the situation is reversed. If interface effects are erroneously neglected, STO appears capable of achieving extremely small EOT values (well below 0.5 nm) before reaching the leakage floor (gray curve). However, in practice, STO films exhibit dead layer effects that increase the total minimum achievable EOT. In the limiting case where the electronic contribution to the permittivity is negligible, eq. S55 implies that the dead layer offset to the minimum EOT is approximately $D\varepsilon_{\text{SiO}_2}$. Using a dead layer parameter of $D = 0.16$ nm (10), the corresponding offset in EOT_{min} is around 0.5 nm. If the electronic contribution to the permittivity

remains unaffected by scaling, the offset is smaller—as shown in Fig. 5(A), the minimum achievable EOT in the presence of a dead layer becomes approximately 0.5 nm. In addition, the vdW gap adds an additional series thickness. When these effects are included in the realistic model, the minimum achievable EOT is much larger. Indeed, under the parameters used in this study (18), the STO stack fails to reach the IRDS target of approximately 0.5 nm EOT. Thus, the interfacial dead layer and vdW gap worsen the scaling limit, forfeiting the advantage STO would otherwise offer due to its very high bulk κ . Meanwhile, a zippered β -Bi₂SeO₅–Bi₂O₂Se (also known as β -BSO–BOS), which forms a continuous bond network with the channel (neither purely vdW-bonded nor fully covalent), shows no such penalty and can, in principle, meet the sub-0.5 nm EOT target, as it is also verified by experiments, achieving a minimum EOT of 0.4 nm (19). These examples suggest that vdW interface effects can significantly alter the assessment of material suitability for extreme scaling.

Fig. 5(B) compiles the minimum achievable EOT for a range of gate dielectrics, comparing the values obtained with and without considering the vdW gap. The vdW gap values used in this study are obtained from ab initio calculations. However, in practice, the vdW gap can be larger due to interface imperfections (20) or fabrication-related effects (21), with values in the range of 0.6 to 2 Å or more being possible. Therefore, to analyze more realistic scenarios, this study considers both the ab initio-calculated vdW gap and a case where the gap is increased by 50%. Given the variability in reported values of electron affinity, tunneling effective mass, and permittivity, each material is represented by a range of minimum achievable EOT values. For each case, literature values and physically reasonable parameters are used to compute the tunneling current as a function of EOT, as shown in fig. S6. The intersection of these curves with the target leakage current defines the corresponding range of minimum achievable EOTs. In nearly all cases, including a vdW gap shifts the EOT to higher values. For high- κ insulators, this shift can critically limit their viability, pushing the EOT beyond the required acceptable limits. For instance, CaF₂ and STO both end up above 0.5 nm EOT when the vdW gap is accounted for, whereas without the gap they may seem to be able to reach 0.5 nm.

On the other hand, for a low- κ material like hBN (which inherently has a larger tunneling current per EOT, i.e., low FoM), the vdW gap’s ability to suppress leakage actually allows a somewhat smaller final EOT than otherwise. However, due to the intrinsically low- κ of such insulators,

they remain promising candidates primarily in applications where extreme leakage suppression outweighs the capacitance penalty. A possible golden mean appears to be LaF_3 , which offers an intermediate permittivity of around 13 (22) and favorable band alignment. Although the inclusion of the vdW gap increases its minimum EOT from approximately 0.17 nm to 0.39 nm—a notable penalty—the resulting values remain highly attractive and could still meet projected requirements. It should also be noted that in the presence of a dead layer for this material, it is possible that LaF_3 may no longer satisfy the minimum EOT condition. However, what is clear is that as the electronic contribution to the permittivity increases, the relative impact of interfacial dead layer degradation diminishes, since interface effects predominantly suppress the ionic response while leaving the electronic part largely unaffected. As a result, materials with a higher fraction of electronic polarizability are more resilient to interface-induced permittivity loss and can potentially support further EOT scaling.

These trends can be understood quantitatively by examining the minimum achievable EOT as a function of the insulator FoM, as shown in Fig. 5(C). In general, if the insulator's FoM is smaller than that of the vdW gap, introducing a vdW gap can actually improve the leakage–EOT trade-off, because the gap acts as a better insulator per unit thickness than the material itself. Conversely, if $\text{FoM}_{\text{ins}} > \text{FoM}_{\text{vdW}}$, the vdW gap adds a disproportionate EOT cost (see eq. S68). The vdW gap is modeled using MoS_2 as the channel, assuming an electron affinity of 4.3 eV, the free electron mass, a permittivity of approximately 2, and a shape factor of 0.8. Under these conditions, an estimated value of $\text{FoM}_{\text{vdW}} \approx 1$ is obtained (see eq. S71). In simpler terms, ultra-high- κ insulators generally do not satisfy this condition, making the vdW gap an EOT penalty rather than a benefit.

In the earlier examples, hBN (with low ε_{ins} and modest band offset) yields a FoM_{ins} of about 0.8–1.2, suggesting that the vdW gap can be beneficial. For LaF_3 , FoM_{ins} is around 5–7, far above the vdW gap's FoM, so the gap erodes its scaling advantage. Fig. 5(C) shows the calculated minimum EOT as a function of the insulator FoM, both with and without a vdW gap. Both the analytical model (see eq. S72) and numerical calculations based on the Tsu–Esaki model confirm that the vdW gap imposes the most significant EOT penalty for high–FoM insulators. In such cases, the minimum achievable EOT asymptotically approaches the vdW-gap-only limit, as illustrated in Fig. 5(D). Notably, for vdW gaps larger than 2 Å, even high- κ insulators yield a minimum achievable EOT greater than 0.5 nm. Finally, for materials with FoM below $\text{FoM}_{\text{vdW}} \approx 1$, the vdW gap can

reduce the total EOT by suppressing leakage more than it adds in series capacitance. This analysis underscores why including vdW interface effects is crucial for predicting realistic scaling limits.

Implications and Outlook

The vdW gap emerges as a double-edged feature in the scaling of 2D semiconductor devices. On one hand, it provides a substantial advantage by suppressing direct tunneling across gate insulators—acting as a built-in vacuum-like barrier that can reduce leakage currents by up to two orders of magnitude. On the other hand, the same gap introduces a severe electrostatic penalty: even a sub-nanometer vdW gap with modest permittivity adds an EOT of approximately 2.7 Å, undermining the benefits of high-permittivity materials and limiting overall gate control.

These results challenge the prevailing assumption that a high dielectric constant and large band gap are sufficient for an insulator to support deep scaling. In practice, the presence of interfacial effects—such as vdW gaps and dead layers—can dominate the effective electrostatics and must be explicitly considered. The findings of this study reveal that these effects set a lower bound on the achievable EOT, regardless of how favorable the bulk dielectric properties may be.

A crucial implication is that dielectric materials whose high- κ values are dominated by lattice (ionic) contributions may be particularly vulnerable to scaling-induced degradation. At reduced thicknesses, these materials often experience suppressed lattice polarization near interfaces—modeled effectively as dead layers—leading to permittivity loss. In contrast, dielectrics with a larger electronic polarization component are more robust against such effects and should be prioritized in materials screening.

To overcome the vdW gap bottleneck, new interfacial engineering strategies are needed. One potential strategy involves the application of strain to the insulator. The out-of-plane dielectric constant of vdW insulators is largely limited by the vdW gap between adjacent layers, whereas the permittivity in the vicinity of the atomic layers is typically much higher. The overall permittivity can therefore be significantly enhanced by reducing the interlayer vdW gap—for example, through the application of out-of-plane strain. This approach may offer a viable pathway for permittivity enhancement in layered insulators.

Another particularly promising approach is the use of zipper interfaces—engineered to transition

from purely vdW to partially covalent bonding. This interfacial bonding continuity eliminates the vacuum-like gap and strengthens dielectric coupling. Native oxide approaches, where crystalline insulators are lattice-matched to the 2D semiconductor, exemplify this strategy. For instance, the β -BSO–BOS system demonstrates sub-0.5 nm EOT without interfacial gaps (19), offering a proof-of-concept for scalable zipper integration. A schematic illustration of this concept is shown in fig. S7.

The framework developed in this work provides a quantitative basis for incorporating such interfacial phenomena directly into design strategies. By extending the insulator FoM to include interface-specific factors—such as vdW gaps, and interfacial permittivity suppression—more realistic scaling predictions can be made. Materials once considered ideal may prove unsuitable under these refined criteria, while others, previously overlooked, may emerge as viable candidates for next-generation devices.

Ultimately, overcoming the limitations imposed by vdW gaps is essential to realizing the full potential of 2D semiconductors in nanoscale electronics. By eliminating the interfacial vacuum gap and restoring dielectric continuity, zipper interfaces offer a promising pathway to approach the scaling limits of electrostatic control in 2D semiconductor devices.

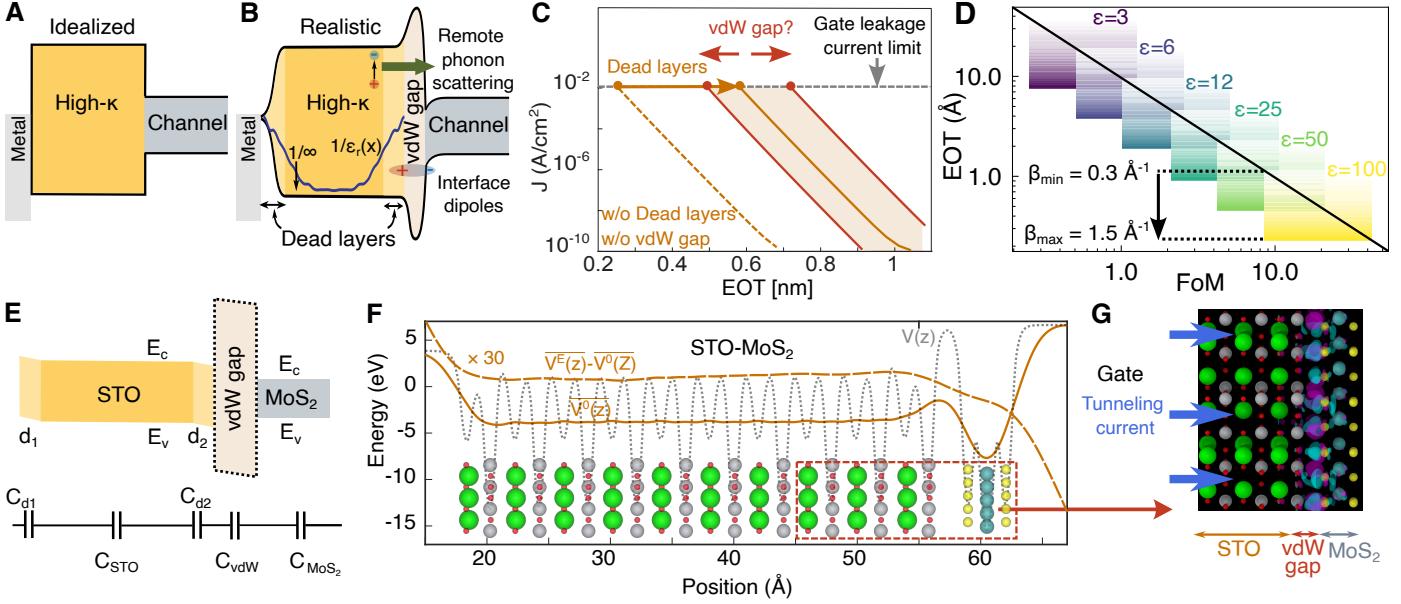


Figure 1: Interface effects in gate–insulator–channel stacks. (A) Idealized gate–insulator–channel stack. (B) Realistic stack highlighting the vdW gap, interface dead layers, dipoles, and remote phonon scattering. (C) Gate tunneling current versus EOT. The minimum achievable EOT is set by the intersection of the tunneling current curve with the maximum allowable leakage for a given application. Dead layers, which lower the effective permittivity in thin films, increase this minimum EOT, while the vdW gap simultaneously suppresses tunneling and adds electrostatic thickness, with one effect potentially outweighing the other. (D) Minimum achievable EOT as a function of the insulator FoM, based on an analytical model from Section S5. Higher permittivity enables scaling to sub-nanometer EOTs. Because the FoM depends on both permittivity and inverse decay length (β), a range of β values from 0.3 \AA^{-1} (leaky insulators) to 1.5 \AA^{-1} (good insulators) is considered. (E) Schematic circuit model illustrating how dead layers and vdW gaps act as series capacitors, thereby increasing the total EOT. (F) Planar-averaged electrostatic potentials in an STO–MoS₂ heterostructure, showing the atomistic potential, zero-field macroscopic potential, and field-induced potential differences. The flat profile within bulk STO reflects its high permittivity, while sharp variations near the interfaces and within the vdW gap reveal low-permittivity regions. (G) Atomistic structure of the STO–MoS₂ interface, highlighting the presence of a vdW gap. Iso-surfaces of charge density difference in the gap region illustrate partial polarization, which affects the gap’s effective permittivity. The vdW gap functions simultaneously as a tunneling barrier and a low- κ interfacial layer.

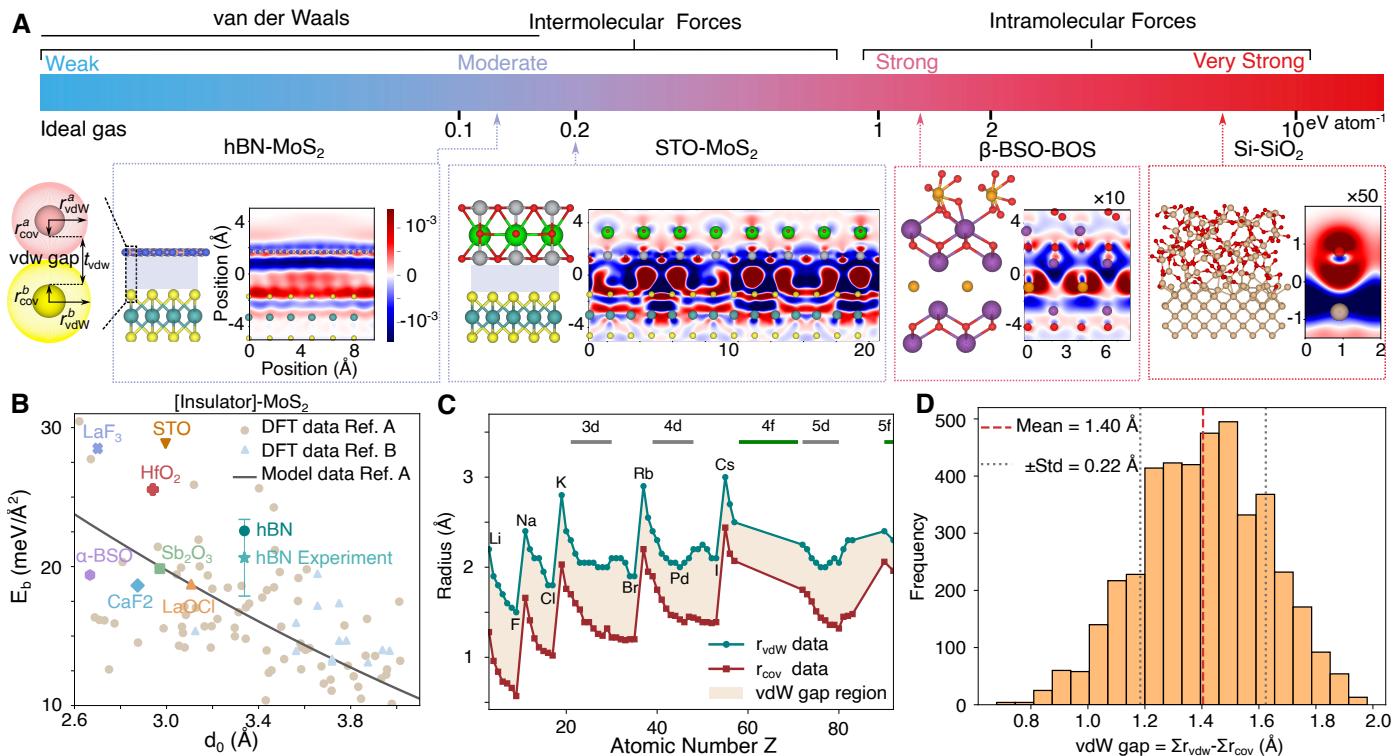


Figure 2: vdW gap definition and statistics. (A) Schematic comparison of interfacial bonding in vdW versus covalent systems. For a vdW-bonded interface (here hBN on MoS₂), the binding energy per area is much lower than for a covalently bonded interface such as SiO₂–Si. The schematic shows charge density distributions: vdW interfaces (hBN–MoS₂) exhibit nearly uniform, delocalized charge, while covalent bonds (Si–SiO₂) display localized, directional density between atoms. The STO–MoS₂ interface appears slightly more strongly bonded than hBN–MoS₂. In contrast, β -BSO–BOS forms a so-called zippered structure, lying between purely vdW and fully covalent bonding. The inset defines the vdW gap as the distance between two surfaces minus the sum of their atomic (covalent) radii, representing the vacuum-like region where no bonding occurs. (B) Binding energy versus vdW gap distance for various 2D heterostructures shows that smaller gaps generally correlate with stronger binding. The DFT data for insulator–MoS₂ stacks (points) generally follow this trend, though some deviations are observed: polar or ionic materials like STO exhibit higher binding due to electrostatic interactions, while atomically flat hBN shows elevated binding from enhanced vdW overlap. DFT data are sourced from Reference A (23) and Reference B (24). (C) Atomic radii r_{cov} and vdW radii r_{vdW} are plotted against atomic number Z for 65 elements, with their difference defining the intrinsic single-element vdW gap. Data for r_{cov} are from Cordero et al. (25), and for r_{vdW} from Batsanov (26). (D) Statistical distribution of vdW gap lengths across binary heterointerfaces, revealing a roughly normal distribution with an average gap near 1.40 ± 0.22 Å.

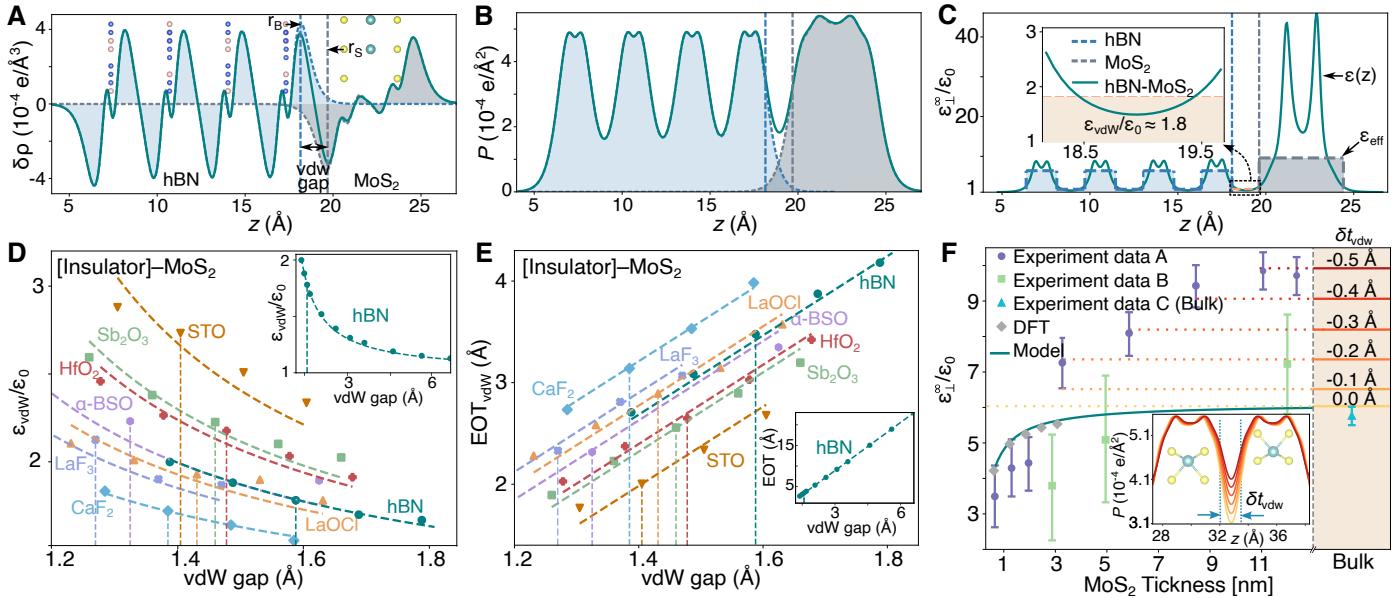


Figure 3: Dielectric properties of the vdW gap. (A) Induced charge density $\Delta\rho(z)$ in an hBN–MoS₂ stack under a small out-of-plane electric field, defined as the difference between charge densities at finite and zero field. Gray and blue curves represent isolated MoS₂ and hBN, respectively; the teal-colored curve corresponds to the combined hBN–MoS₂ stack, showing interfacial charge redistribution. (B) Induced polarization $P(z)$, obtained by integrating $\Delta\rho(z)$, reveals non-zero polarization within the vdW gap when the materials are vdW bonded. (C) Spatial profile of the permittivity $\varepsilon(z)$, peaking near atomic planes and dropping toward unity at the center of the vdW gap. The profile is modeled as alternating slabs: high-permittivity regions representing atomic layers and low-permittivity regions representing the vdW gaps. Effective permittivities for each region are obtained by integrating over the corresponding sections of $\varepsilon(z)$. (D) Effective permittivity versus vdW gap thickness for various insulator–MoS₂ stacks from ab initio calculations, showing that narrower gaps enhance polarization and increase permittivity. The vdW gap thickness was systematically varied to assess how the gap permittivity changes with separation. The inset highlights results for hBN at increased vdW gap thicknesses. (E) EOT contribution of the vdW gap for the same systems. The EOT increases nearly linearly with vdW gap thickness; even a 1.4 Å gap adds approximately 2.7 Å to the total EOT. (F) Effective dielectric constant of multilayer MoS₂ increases with layer number, approaching the bulk limit. Reducing the vdW gap—e.g., via strain—can further enhance polarization and increase the effective permittivity, as shown in the inset. Experimental data are from (14) (A), (27) (B), and (28) (C).

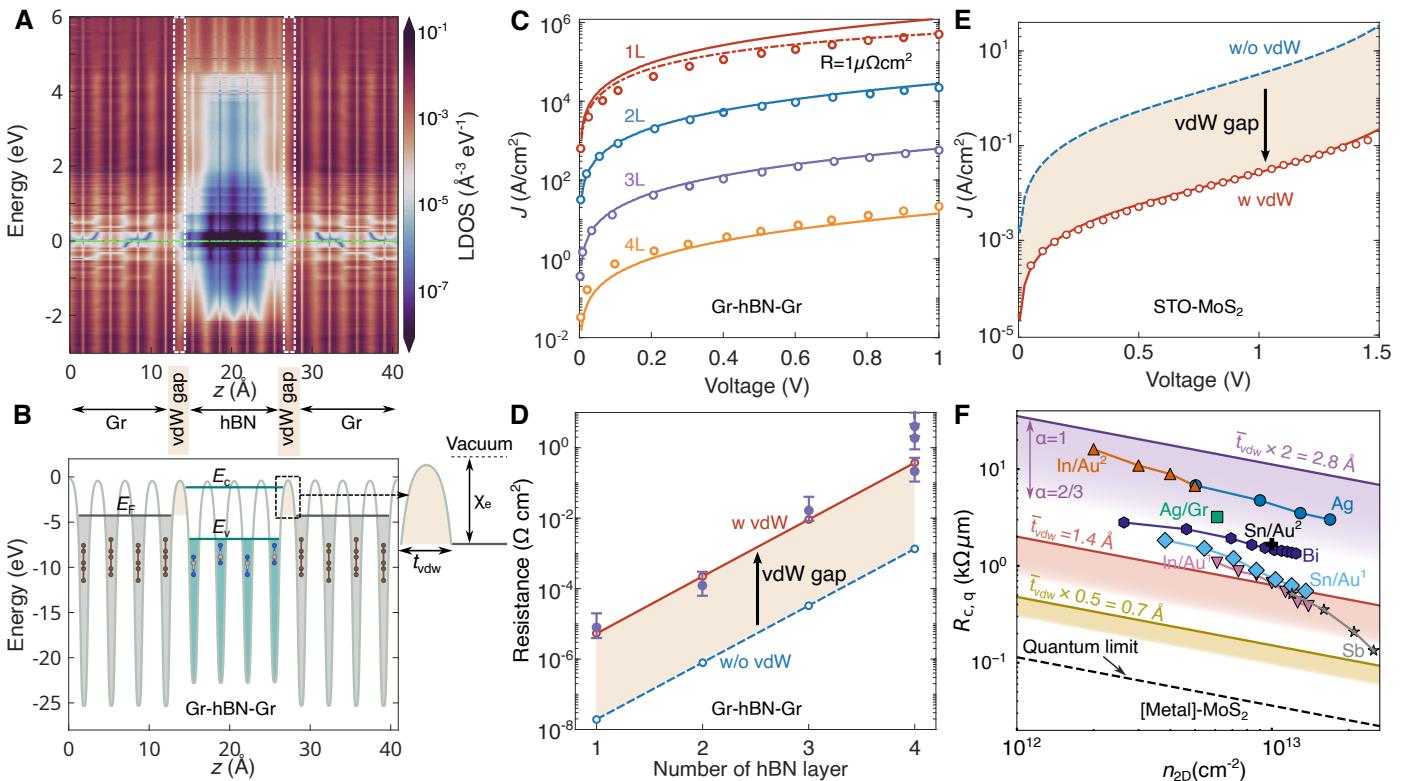


Figure 4: Tunneling suppression due to vdW gaps. **(A)** LDOS from ab initio NEGF simulations for a graphene–hBN–graphene tunnel junction under bias, showing strong LDOS decay within hBN and sharp drops at the graphene–hBN interfaces, indicating a tunneling barrier. **(B)** Planar-averaged electrostatic potential for the same stack, with vdW gap regions approaching the vacuum level, establishing a high barrier height nearly equal to the graphene–hBN band offset. **(C)** Current–voltage (J – V) characteristics for graphene–hBN–graphene tunneling with one to four hBN layers. Solid lines and experimental symbols (from (29)) agree well across current magnitudes. The minor deviation for monolayer hBN at high bias can be corrected by including a series resistance (dashed line). **(D)** Small-bias tunneling resistance vs. hBN thickness. Red and blue curves show results with and without the vdW gap, respectively. Omitting the gap underestimates resistance, while its inclusion increases resistance by nearly 260×, consistent with experimental observations. **(E)** Calculated gate leakage for monolayer MoS₂ with an STO insulator. Including a vdW gap of ~1.4 Å reduces tunneling by an order of magnitude, consistent with experimental data from (10). Red and blue curves show results with and without the vdW gap, respectively. **(F)** Quantum contact resistivity of metal–MoS₂ vdW contacts as a function of vdW gap thickness. Even a modest vdW gap significantly increases resistance. Experimental data for purely vdW-bonded contacts suggest vdW gaps in the range of approximately 1.2–2.7 Å, confirming that such gaps strongly impact contact performance. Experimental data are available for Ag (30); Ag/Gr (31); In/Au¹ and Sn/Au¹ (32); In/Au² (33); Sn/Au² (34); Sb (35); and Bi (36). 17

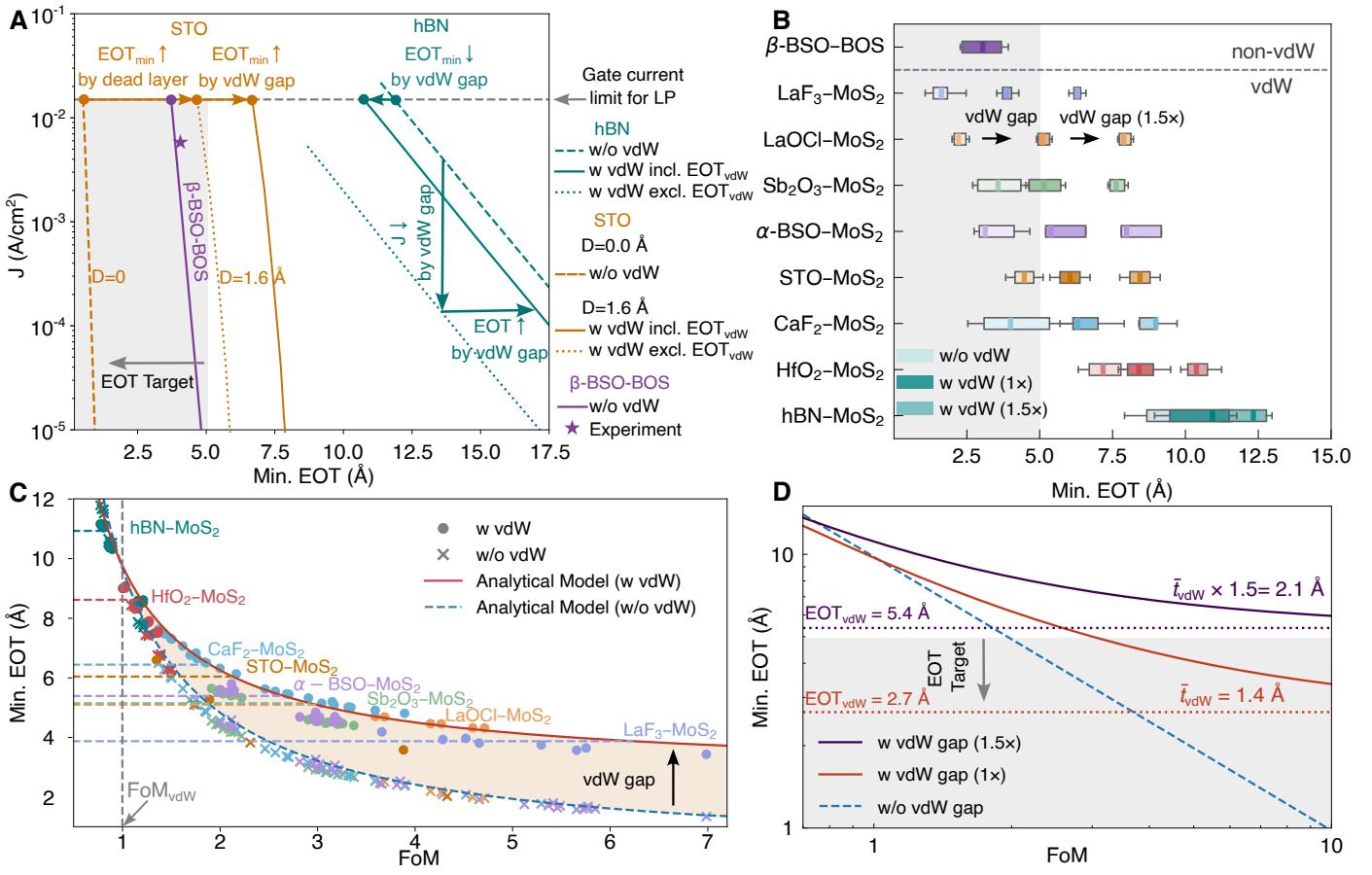


Figure 5: Impact of vdW gaps on scaling limits. (A) Calculated leakage current density vs. EOT for hBN, STO, and β -BSO–BOS on MoS₂. Three cases are shown per insulator with a vdW gap. For hBN, the gap suppresses leakage and slightly reduces the minimum EOT. For STO, dead layers and the vdW gap raise EOT beyond 0.5 nm. In contrast, β -BSO–BOS forms a bonded interface without a vdW gap, maintaining low leakage at small EOT. Dashed lines mark the target leakage current $J_{\text{target}} = 1.5 \times 10^{-2} \text{ A/cm}^2$. (B) Minimum achievable EOT for various insulators on MoS₂, comparing cases without a vdW gap, with a DFT-predicted gap (1x), and with an enlarged gap (1.5x) accounting for non-idealities. For most materials, the vdW gap increases EOT. Only low- κ dielectrics like hBN benefit from leakage suppression. β -BSO–BOS achieves sub-0.5 nm EOT experimentally by avoiding this penalty. (C) Minimum EOT vs. insulator FoM, showing when the vdW gap helps or hinders scaling. Curves show analytical predictions; symbols denote results from numerical Tsu–Esaki calculations. For $\text{FoM} \lesssim 1$, the vdW gap improves scaling; above this threshold, it becomes detrimental. (D) Log-log plot of minimum EOT vs. insulator FoM, comparing no-gap, 1x, and 1.5x vdW gap thicknesses. At high FoM, the EOT saturates to the limit imposed by the vdW gap, establishing it as a fundamental scaling barrier.

References and Notes

1. S. Das, *et al.*, Transistors based on two-dimensional materials for future integrated circuits. *Nat. Electron.* **4** (11), 786–799 (2021), doi:10.1038/s41928-021-00670-1.
2. International Roadmap for Devices and Systems (IRDS) (2023), <https://irds.ieee.org/>.
3. Y. Y. Illarionov, *et al.*, Insulators for 2D nanoelectronics: the gap to bridge. *Nat. Commun.* **11** (1), 3385 (2020), doi:10.1038/s41467-020-16640-8.
4. M. R. Osanloo, M. L. Van de Put, A. Saadat, W. G. Vandenberghe, Identification of two-dimensional layered dielectrics from first principles. *Nat. Commun.* **12** (1), 5051 (2021), doi:10.1038/s41467-021-25310-2.
5. Y. Li, *et al.*, High-throughput screening and machine learning classification of van der Waals dielectrics for 2D nanoelectronics. *Nat. Commun.* **15** (1), 9527 (2024), doi:10.1038/s41467-024-53864-4.
6. M. Stengel, N. A. Spaldin, Origin of the dielectric dead layer in nanoscale capacitors. *Nature* **443** (7112), 679–682 (2006), doi:10.1038/nature05148.
7. M. V. Fischetti, D. A. Neumayer, E. A. Cartier, Effective electron mobility in Si inversion layers in metal–oxide–semiconductor systems with a high- κ insulator: The role of remote phonon scattering. *J. Appl. Phys.* **90** (9), 4587–4608 (2001), doi:10.1063/1.1405826.
8. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, 2D materials and van der Waals heterostructures. *Science* **353** (6298), aac9439 (2016), doi:10.1126/science.aac9439.
9. Y.-C. Yeo, T.-J. King, C. Hu, MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations. *IEEE Trans. Electron Devices* **50** (4), 1027–1035 (2003), doi:10.1109/ted.2003.812504.
10. Huang, Jing-Kai and Wan, Yi and Shi, Junjie and Zhang, Ji and Wang, Zeheng and Wang, Wenxuan and Yang, Ni and Liu, Yang and Lin, Chun-Ho and Guan, Xinwei and Hu, Long and Yang, Zi-Liang and Huang, Bo-Chao and Chiu, Ya-Ping and Yang, Jack and Tung, Vincent and

- Wang, Danyang and Kalantar-Zadeh, Kourosh and Wu, Tom and Zu, Xiaotao and Qiao, Liang and Li, Lain-Jong and Li, Sean, High- κ perovskite membranes as insulators for two-dimensional transistors. *Nature* **605** (7909), 262–267 (2022), doi:10.1038/s41586-022-04588-2.
11. M. Moazzami Gudarzi, S. H. Aboutalebi, Mapping the binding energy of layered crystals to macroscopic observables. *Adv. Sci.* **9** (33), 2204001 (2022), doi:10.1002/advs.202204001.
 12. T. Li, *et al.*, A native oxide high- κ gate dielectric for two-dimensional electronics. *Nat. Electron.* **3** (8), 473–478 (2020), doi:10.1038/s41928-020-0444-6.
 13. Y. Shen, *et al.*, MoS₂ transistors with 4 nm hBN Gate Dielectric and 0.46 V threshold voltage. *ACS Nano* **19** (17), 16903–16912 (2025), doi:10.1021/acsnano.5c02341.
 14. X. Chen, *et al.*, Probing the electron states and metal-insulator transition mechanisms in molybdenum disulphide vertical heterostructures. *Nat. Commun.* **6** (1), 6088 (2015), doi:10.1038/ncomms7088.
 15. G. Giovannetti, *et al.*, Doping graphene with metal contacts. *Phys. Rev. Lett.* **101**, 026803 (2008), doi:10.1103/PhysRevLett.101.026803.
 16. D. Akinwande, C. Biswas, D. Jena, The quantum limits of contact resistance and ballistic transport in 2D transistors. *Nat. Electron.* **8** (2), 96–98 (2025), doi:10.1038/s41928-024-01335-5.
 17. T. Knobloch, *et al.*, The performance limits of hexagonal boron nitride as an insulator for scaled CMOS devices based on two-dimensional materials. *Nat. Electron.* **4** (2), 98–108 (2021), doi:10.1038/s41928-020-00529-x.
 18. Materials and methods are available as supplementary material.
 19. Y. Zhang, *et al.*, A single-crystalline native dielectric for two-dimensional semiconductors with an equivalent oxide thickness below 0.5 nm. *Nat. Electron.* **5** (10), 643–649 (2022), doi:10.1038/s41928-022-00824-9.
 20. A. P. Rooney, *et al.*, Observing Imperfection in Atomic Interfaces for van der Waals Heterostructures. *Nano Letters* **17** (9), 5222–5228 (2017), doi:10.1021/acs.nanolett.7b01248.

21. P. Luo, *et al.*, Molybdenum disulfide transistors with enlarged van der Waals gaps at their dielectric interface via oxygen accumulation. *Nat. Electron.* **5** (12), 849–858 (2022), doi: 10.1038/s41928-022-00877-w.
22. Z. Fu, *et al.*, Low-temperature controlled growth of 2D LaOCl with enhanced dielectric properties for advanced electronics. *Adv. Funct. Mater.* p. 2501136 (2025), doi:10.1002/adfm.202501136.
23. D. Willhelm, *et al.*, Predicting van der waals heterostructures by a combined machine learning and density functional theory approach. *ACS Appl. Mater. Interfaces* **14** (22), 25907–25919 (2022), doi:10.1021/acsami.2c04403.
24. K. Tang, W. Qi, Y. Wei, G. Ru, W. Liu, High-throughput calculation of interlayer van der Waals forces validated with experimental measurements. *Research* **2022** (2022), doi:10.34133/2022/9765121.
25. B. Cordero, *et al.*, Covalent radii revisited. *Dalton Trans.* (21), 2832–2838 (2008), doi: 10.1039/B801115J.
26. S. S. Batsanov, Van der Waals radii of elements. *Inorg. Mater.* **37** (9), 871–885 (2001), doi: 10.1023/A:1011625728803.
27. Y. Kang, D. Jeon, T. Kim, Local mapping of the thickness-dependent dielectric constant of MoS₂. *J. Phys. Chem. C* **125** (6), 3611–3615 (2021), doi:10.1021/acs.jpcc.0c11198.
28. B. L. Evans, P. A. Young, Delocalized excitons in thin anisotropic crystals. *Phys. Status Solidi B* **25** (1), 417–425 (1968), doi:10.1002/pssb.19680250139.
29. L. Britnell, *et al.*, Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12** (3), 1707–1710 (2012), doi:10.1021/nl3002205.
30. Y. H. Deng, *et al.*, Improved electrical uniformity and performance via low-cost hybrid wet transfer method for van der Waals source/drain contact formation in MoS₂ field-effect transistors. *IEEE Trans. Electron Devices* **71** (12), 7943–7947 (2024), doi:10.1109/TED.2024.3487821.

31. D. Qi, *et al.*, Graphene-enhanced metal transfer printing for strong van der Waals contacts between 3D metals and 2D semiconductors. *Adv. Funct. Mater.* **33** (27) (2023), doi:10.1002/adfm.202301704.
32. A. Kumar, *et al.*, Sub-200 $\Omega \cdot \mu\text{m}$ alloyed contacts to synthetic monolayer MoS₂, in *2021 IEEE International Electron Devices Meeting (IEDM)* (IEEE) (2021), pp. 7.3.1–7.3.4, doi: 10.1109/IEDM19574.2021.9720609.
33. Y. Wang, *et al.*, Van der Waals contacts between three-dimensional metals and two-dimensional semiconductors. *Nature* **568** (7750), 70–74 (2019), doi:10.1038/s41586-019-1052-3.
34. A.-S. Chou, *et al.*, High on-state current in chemical vapor deposited monolayer MoS₂ nFETs with Sn ohmic contacts. *IEEE Electron Device Lett.* **42** (2), 272–275 (2021), doi:10.1109/LED.2020.3048371.
35. W. Li, *et al.*, Approaching the quantum limit in two-dimensional semiconductor contacts. *Nature* **613** (7943), 274–279 (2023), doi:10.1038/s41586-022-05431-4.
36. Z. Sun, *et al.*, Low contact resistance on monolayer MoS₂ field-effect transistors achieved by CMOS-compatible metal contacts. *ACS Nano* **18** (33), 22444–22453 (2024), doi:10.1021/acsnano.4c07267.
37. G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993), doi:10.1103/PhysRevB.47.558.
38. G. Kresse, J. Hafner, Ab initiomolecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49** (20), 14251–14269 (1994), doi: 10.1103/PhysRevB.49.14251.
39. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132** (15) (2010), doi:10.1063/1.3382344.
40. K. Cherkaoui, *et al.*, Electrical, structural, and chemical properties of HfO₂ films formed by electron beam evaporation. *J. Appl. Phys.* **104** (6), 064113 (2008), doi:10.1063/1.2978209.

41. M. Lee, Y. Youn, K. Yim, S. Han, High-throughput ab initio calculations on dielectric constant and band gap of non-oxide dielectrics. *Sci. Rep.* **8** (1), 14794 (2018), doi: 10.1038/s41598-018-33095-6.
42. A. Sher, R. Solomon, K. Lee, M. W. Muller, Transport properties of LaF₃. *Phys. Rev.* **144** (2), 593–604 (1966), doi:10.1103/PhysRev.144.593.
43. L. Li, *et al.*, Ultrathin van der Waals lanthanum oxychloride dielectric for 2D field-effect transistors. *Adv. Mater.* p. 2309296 (2023), doi:10.1002/adma.202309296.
44. K. Liu, *et al.*, A wafer-scale van der Waals dielectric made from an inorganic molecular crystal film. *Nat. Electron.* **4** (12), 906–913 (2021), doi:10.1038/s41928-021-00683-w.
45. J. Wang, *et al.*, High-temperature relaxations in CaF₂ single crystals. *Mater. Sci. Eng. B* **188**, 31–34 (2014), doi:10.1016/j.mseb.2014.06.004.
46. A. Pierret, *et al.*, Dielectric permittivity, conductivity and breakdown field of hexagonal boron nitride. *Mater. Res. Express* **9** (6), 065901 (2022), doi:10.1088/2053-1591/ac4fe1.
47. A. Laturia, M. L. Van de Put, W. G. Vandenberghe, Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: from monolayer to bulk. *npj 2D Mater. Appl.* **2** (1) (2018), doi:10.1038/s41699-018-0050-x.
48. R. Geick, C. H. Perry, G. Rupprecht, Normal modes in hexagonal noron nitride. *Phys. Rev.* **146** (2), 543–547 (1966), doi:10.1103/PhysRev.146.543.
49. W. Li, *et al.*, Uniform and ultrathin high- κ gate dielectrics for two-dimensional electronic devices. *Nat. Electron.* **2** (12), 563–571 (2019), doi:10.1038/s41928-019-0334-y.
50. Y. Li, J. R. Sun, J. L. Zhao, B. G. Shen, Surface electronic inhomogeneity of the (001)-SrTiO₃:Nb crystal with a terrace-structured morphology. *J. Appl. Phys.* **114** (15), 154303 (2013), doi:10.1063/1.4825047.
51. V. E. Henrich, G. Dresselhaus, H. J. Zeiger, Surface defects and the electronic structure of SrTiO₃ surfaces. *Phys. Rev. B* **17** (12), 4908–4921 (1978), doi:10.1103/PhysRevB.17.4908.

52. W. Pong, C. S. Inouye, Photoemission studies of LaF₃. *J. Opt. Soc. Am.* **68** (4), 521–523 (1978), doi:10.1364/JOSA.68.000521.
53. H. Liu, L. Rao, J. Qi, G. Tang, First-principles insights into Bi₂XO₅ (X=Se,Te) monolayers as high- κ gate dielectrics for 2D electronics. *Appl. Phys. Lett.* **126** (7) (2025), doi:10.1063/5.0242581.
54. R. T. Poole, *et al.*, Electronegativity as a unifying concept in the determination of fermi energies and photoelectric thresholds. *Chem. Phys. Lett.* **36** (3), 401–403 (1975), doi:10.1016/0009-2614(75)80267-3.
55. F. Matusalem, M. Marques, L. K. Teles, A. Filippetti, G. Cappellini, Electronic properties of fluorides by efficient approximated quasiparticle DFT-1/2 and PSIC methods: BaF₂, CaF₂ and CdF₂ as test cases. *J. Phys. Condens. Matter* **30** (36), 365501 (2018), doi:10.1088/1361-648X/aad654.
56. S. A. Marye, R. R. Kumar, A. Useinov, N. Tumilty, Thermal stability, work function and Fermi level analysis of 2D multi-layered hexagonal boron nitride films. *Microelectron. Eng.* **283**, 112106 (2024), doi:10.1016/j.mee.2023.112106.
57. S. W. Lee, J. H. Han, C. S. Hwang, Electronic conduction mechanism of SrTiO₃ thin film grown on Ru electrode by atomic layer deposition. *Electrochim Solid-State Lett* **12** (11), G69 (2009), doi:10.1149/1.3212897.
58. M. A. Pawlak, *et al.*, Towards 1X DRAM: Improved leakage 0.4 nm EOT STO-based MIMcap and explanation of leakage reduction mechanism showing further potential, in *2011 Symposium on VLSI Technology - Digest of Technical Papers* (2011).
59. O. Kochan, *et al.*, Energy structure and luminescence of CeF₃ crystals. *Materials* **14** (15), 4243 (2021), doi:10.3390/ma14154243.
60. P. Khakbaz, *et al.*, Two-dimensional Bi₂SeO₂ and its native insulators for next-generation nanoelectronics. *ACS Nano* **19** (10), 9788–9800 (2025), doi:10.1021/acsnano.4c12160.

61. B. Y. Zhang, *et al.*, Theoretical and experimental characterizations of hot electron emission of n-Si/CaF₂/Au emitter used in hot electron detection experiment. *Phys. B Condens. Matter* **272** (1), 425–427 (1999), doi:10.1016/S0921-4526(99)00388-9.
62. M. I. Vexler, *et al.*, Electrical characterization and modeling of the Au/CaF₂/nSi(111) structures with high-quality tunnel-thin fluoride layer. *J. Appl. Phys.* **105** (8), 083716 (2009), doi:10.1063/1.3110066.
63. J. P. Allen, J. J. Carey, A. Walsh, D. O. Scanlon, G. W. Watson, Electronic structures of antimony oxides. *J. Phys. Chem. C* **117** (28), 14759–14769 (2013), doi:10.1021/jp4026249.
64. J.-B. Liu, *et al.*, Few-layer α -Sb₂O₃ molecular crystals as high- κ van der Waals dielectrics: electronic decoupling and significant surface ionic behaviors. *J. Mater. Chem. C* **12** (24), 8825–8836 (2024), doi:10.1039/D4TC00753K.
65. M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, F. Bechstedt, Linear optical properties in the projector-augmented wave methodology. *Phys. Rev. B* **73** (4), 045112 (2006), doi:10.1103/physrevb.73.045112.
66. F. Giustino, A. Pasquarello, Theory of atomic-scale dielectric permittivity at insulator interfaces. *Phys. Rev. B* **71** (14), 144104 (2005), doi:10.1103/PhysRevB.71.144104.
67. P. Kumar, *et al.*, Thickness and electric-field-dependent polarizability and dielectric constant in phosphorene. *Phys. Rev. B* **93** (19), 195428 (2016), doi:10.1103/PhysRevB.93.195428.
68. M. S. Majdoub, R. Maranganti, P. Sharma, Understanding the origins of the intrinsic dead layer effect in nanocapacitors. *Phys. Rev. B* **79** (11), 115412 (2009), doi:10.1103/PhysRevB.79.115412.
69. J. Maserjian, Tunneling in thin MOS structures. *J. Vac. Sci. Technol.* **11** (6), 996–1003 (1974), doi:10.1116/1.1318719.
70. X. Guan, D. Kim, K. C. Saraswat, H.-S. P. Wong, Complex band structures: From parabolic to elliptic approximation. *IEEE Electron Device Lett.* **32** (9), 1296–1298 (2011), doi:10.1109/LED.2011.2160143.

71. B. Brar, G. D. Wilk, A. C. Seabaugh, Direct extraction of the electron tunneling effective mass in ultrathin SiO₂. *Appl. Phys. Lett.* **69** (18), 2728–2730 (1996), doi:10.1063/1.117692.
72. D. Esseni, M. Pala, P. Palestri, C. Alper, T. Rollo, A review of selected topics in physics based modeling for tunnel field-effect transistors. *Semicond. Sci. Tech.* **32** (8), 083005 (2017), doi:10.1088/1361-6641/aa6fca.

Acknowledgments

Funding: We acknowledge support by the Austrian Science Fund (FWF) and the European Research Council (ERC) under Grant agreement no. 101055379 (F2GO). We acknowledge fruitful discussions with Deji Akinwande, Mario Lanza, Aftab Nazir, Theresia Knobloch, and Dominic Waldhoer.

Author contributions: M.P. developed the methodology, performed all simulations, analyzed the data, interpreted the results, prepared the figures, and wrote the manuscript with feedback and suggestions from T.G. T.G. conceived the initial idea of the study, which was continuously refined in daily discussions with M.P. Both authors approved the final version of the manuscript.

Competing interests: There are no competing interests to declare.

Data and materials availability: Available on request

Supplementary materials

Materials and Methods

Supplementary Text

Figs. S1 to S7

Tables S1

References (37-72)

Supplementary Materials for

The van der Waals Gap: a Hidden Showstopper in

Semiconductor Device Scaling

Mahdi Pourfath* and Tibor Grassler*

*Corresponding author. Email: pourfath@iue.tuwien.ac.at and grasser@tuwien.ac.at

This PDF file includes:

Materials and Methods

Supplementary Text

Figures S1 to S7

Tables S1

Materials and Methods

All first-principles calculations were performed using the Vienna *Ab initio* Simulation Package (VASP) (37, 38). The exchange-correlation effects were treated within the generalized gradient approximation (GGA) using the Perdew–Burke–Ernzerhof (PBE) functional, with vdW interactions accounted for via the DFT-D3 method of Grimme (39). A plane-wave energy cutoff of at least 580 eV was employed in all calculations.

Heterostructures were constructed by applying strain values below 1.5% to the MoS₂ layer to match the lattice constants of the insulating substrates while minimizing variations in the insulator band gap. Subsequent structural relaxation was performed along the direction perpendicular to the interface to adjust interlayer spacing. Atomic positions were relaxed until residual Hellmann–Feynman forces fell below 1×10^{-2} eV/Å, and electronic self-consistency was achieved with an energy convergence criterion of 10^{-8} eV. For Brillouin zone sampling during geometry relaxations, Monkhorst–Pack \mathbf{k} -point meshes with a density of approximately 0.04 Å⁻¹ were used. For total energy and charge density calculations, denser \mathbf{k} -point meshes were employed with a Γ -centered grid corresponding to a reciprocal-space spacing of about 0.03 Å⁻¹. A vacuum region of at least 30 Å was introduced along the out-of-plane (*z*) direction to eliminate spurious interactions between periodic images, which is especially critical for dielectric property calculations.

Dielectric constants were computed using both macroscopic and microscopic approaches. The in-plane dielectric responses were obtained via macroscopic methods, as described in the Supplementary Information. For out-of-plane components, a microscopic analysis was preferred to avoid numerical errors associated with the series capacitor model and the need for large vacuum separations to suppress artificial interactions. In these simulations, a small external electric field was applied perpendicular to the heterostructure (along the *z*-axis), and the field-induced charge density was computed by subtracting the zero-field electron density. This induced charge distribution enabled calculation of the local polarization and, in turn, the spatially varying dielectric profile across the interface. Details of the first-principles calculations used to extract the vdW gap's dielectric profile, including the applied fields and the analysis of induced charge and polarization, are provided in the Supplementary Text.

For LDOS evaluation, non-equilibrium Green's function (NEGF) simulations were performed

in QuantumATK using a DZP basis, norm-conserving pseudopotentials, a 150 Ry mesh cutoff, and a 150 k-point grid along the transport direction. Energy convergence was set below 10^{-6} eV.

The three key material parameters for all insulator layers—dielectric constant $\varepsilon/\varepsilon_0$, electron affinity χ , and tunneling effective mass m^*/m_0 —were collected from literature as follows. Although the band gap E_g does not directly enter the tunneling expression, it contributes via a two-band correction used for more accurate evaluation of the tunneling current. Since reported values of E_g show less variation than the other three parameters, a representative single value for each material was used.

Bulk dielectric constants ranged from high values in STO ($\varepsilon_{\text{bulk}}/\varepsilon_0 = 270$, with a dead layer thickness $D = 1.6 \text{ \AA}$ (10)) and α -BSO ($\varepsilon_{\text{bulk}}/\varepsilon_0 = 30.1$, with $D = 0.8 \text{ \AA}$, extracted from a dead layer model fit to experimental thickness-dependent data from Ref. (19), shown in fig. S2(C)), to moderate values in HfO₂ (16–25 (40)), LaF₃ (14–16.5 (41, 42)), LaOCl (10.8–13.8 (22, 43)), and Sb₂O₃ (11.5 (44)), and down to lower values in CaF₂ (8.7 (45)) and hBN (3.4–5.1 (46–48)).

Bulk dielectric constants ranged from high values in STO ($\varepsilon_{\text{bulk}}/\varepsilon_0 = 270$, with a dead layer thickness $D = 1.6 \text{ \AA}$ (10)) and α -BSO ($\varepsilon_{\text{bulk}}/\varepsilon_0 = 16.5$ (19)), to moderate values in HfO₂ ($\varepsilon_{\text{bulk}}/\varepsilon_0 = 16$, with $D = 1.18 \text{ \AA}$, extracted from a dead layer model fit to experimental thickness-dependent data from Ref. (49), shown in fig. S2(C)), LaF₃ (14–16.5 (41, 42)), LaOCl (10.8–13.8 (22, 43)), and Sb₂O₃ (11.5 (44)), and down to lower values in CaF₂ (6.8–8.4 (45)) and hBN (3.4–5.1 (46–48)).

Electron affinities χ varied significantly. For STO, values of 3.3 (10), 3.9 (50), and 4.1 (51) were reported. Sb₂O₃ ranged from 2.9 to 3.2 (44), LaF₃ and LaOCl were consistently reported at 2.3 (4, 52), and α -BSO exhibited an electron affinity of 1.89 (53). HfO₂ spanned 1.75 to 2.25 (40), and CaF₂ varied from –0.15 (54) to 1.04 and 1.73 (55), and up to 2.9 (44). hBN exhibited values between 1.59 and 2.26 (56).

Effective masses m^*/m_0 were lowest in HfO₂ (0.08–0.14 (40)) and STO (0.10 (57), 0.50 (58)). Intermediate values were used for LaF₃ (0.90 (59)), LaOCl (1.12 (4)), and α -BSO (0.63 (60)). CaF₂ ranged from 0.31 (61) to 1.20 (62), Sb₂O₃ from 0.85 (63) to 1.04 (64), and hBN was assigned 0.50 (29).

For the β -BSO–BOS stack, no vdW gap is present, as the interface exhibits partial covalent character. A metal gate work function of 4.4 eV was used to match experimental alignment. The

conduction band offset was assumed to be $\Delta\chi = 1.55\text{--}1.86$ eV (60), with β -BSO and BOS band gaps of 3.2 and 1.0 eV, respectively.

Representative band gap values E_g were used for each material, as tunneling current depends only weakly on E_g —entering only through the two-band correction—and reported variations are relatively small. The values used were: 11.8 eV for CaF₂, 9.7 eV for LaF₃, 5.9 eV for hBN, 5.5 eV for LaOCl, 5.3 eV for HfO₂, 4.4 eV for Sb₂O₃, 3.5 eV for α -BSO, 3.3 eV for STO, 3.2 eV for β -BSO.

Supplementary Text

S1 vdW Gap Dielectric Constant

Note: Throughout this section, all dielectric constants refer to the *electronic* contribution, ε^∞ , which describes the system's response to an external electric field via the electronic charge redistribution, as computed by linear-response DFT or field-induced polarization. The superscript ∞ is omitted for notational simplicity. The dielectric response of the vdW gap reflects only electronic polarization, as the region contains no atoms and therefore no lattice contribution.

Macroscopic Approach: Using DFT calculations combined with classical electrostatics, the out-of-plane ($\varepsilon_{\text{vdW}}^\perp$) and in-plane ($\varepsilon_{\text{vdW}}^{\parallel}$) dielectric constants of the vdW gap are derived. The effective dielectric response of composite stacks under a small homogeneous electric field is computed using the macroscopic dielectric framework of Ref. (65), based on perturbative linear-response DFT.

As illustrated in fig. S1, a simulation cell comprising an insulator, a vdW gap, and a monolayer of MoS₂ is constructed to calculate the effective dielectric response of the entire supercell. The system is embedded in at least 20 Å of vacuum to avoid spurious interactions with periodic images imposed by boundary conditions. The decomposition of the total effective dielectric response into contributions from the insulator, MoS₂, and the vdW gap is not directly accessible from DFT. Instead, it is inferred using classical electrostatics by modeling the system as a combination of capacitors in series or parallel (47), depending on the direction of the applied electric field. This approximation is justified by the spatial separation between components and the non-overlapping nature of their wavefunctions in typical slab geometries. Intrinsic dielectric properties of each region are determined from three separate DFT calculations within a common supercell of length L_{cell} , all incorporating a vacuum region: (i) the combined system comprising the insulator, vdW gap, and monolayer MoS₂ (yielding $\varepsilon_{c,\text{Ins-MoS}_2}$); (ii) the isolated insulator slab ($\varepsilon_{c,\text{Ins}}$); and (iii) the isolated MoS₂ layer ($\varepsilon_{c,\text{MoS}_2}$). In each case, ε_c represents the effective dielectric constant of the entire cell, including vacuum, whereas ε_{Ins} , $\varepsilon_{\text{MoS}_2}$, and ε_{vdW} denote the intrinsic dielectric constants of the respective material regions. The intrinsic value ε_{vdW} is then extracted by combining the computed effective constants through classical electrostatics.

Out-of-Plane Direction: The dielectric constant of each configuration is derived from the inverse-capacitance (series capacitor) model. For the *combined system* one can write:

$$\frac{L_{\text{cell}}}{\varepsilon_{c,\text{Ins-MoS}_2}^{\perp}} = \frac{L_{\text{Ins}}}{\varepsilon_{\text{Ins}}^{\perp}} + \frac{L_{\text{MoS}_2}}{\varepsilon_{\text{MoS}_2}^{\perp}} + \frac{L_{\text{vdW}}}{\varepsilon_{\text{vdW}}^{\perp}} + \frac{L_{\text{vac}}}{1} \quad (\text{S1})$$

Using $L_{\text{vac}} = L_{\text{cell}} - L_{\text{Ins}} - L_{\text{MoS}_2} - L_{\text{vdW}}$, one gets:

$$L_{\text{cell}} \left(\frac{1}{\varepsilon_{c,\text{Ins-MoS}_2}^{\perp}} - 1 \right) = L_{\text{Ins}} \left(\frac{1}{\varepsilon_{\text{Ins}}^{\perp}} - 1 \right) + L_{\text{MoS}_2} \left(\frac{1}{\varepsilon_{\text{MoS}_2}^{\perp}} - 1 \right) + L_{\text{vdW}} \left(\frac{1}{\varepsilon_{\text{vdW}}^{\perp}} - 1 \right) \quad (\text{S2})$$

For the *isolated insulator* and *isolated MoS₂*, respectively, one obtains:

$$L_{\text{cell}} \left(\frac{1}{\varepsilon_{c,\text{Ins}}^{\perp}} - 1 \right) = L_{\text{Ins}} \left(\frac{1}{\varepsilon_{\text{Ins}}^{\perp}} - 1 \right), \quad L_{\text{cell}} \left(\frac{1}{\varepsilon_{c,\text{MoS}_2}^{\perp}} - 1 \right) = L_{\text{MoS}_2} \left(\frac{1}{\varepsilon_{\text{MoS}_2}^{\perp}} - 1 \right) \quad (\text{S3})$$

Substituting eq. S2–eq. S3 into eq. S1:

$$L_{\text{cell}} \left(\frac{1}{\varepsilon_{c,\text{Ins-MoS}_2}^{\perp}} - \frac{1}{\varepsilon_{\text{Ins}}^{\perp}} - \frac{1}{\varepsilon_{c,\text{MoS}_2}^{\perp}} + 1 \right) = L_{\text{vdW}} \left(\frac{1}{\varepsilon_{\text{vdW}}^{\perp}} - 1 \right) \quad (\text{S4})$$

Solving for $\varepsilon_{\text{vdW}}^{\perp}$:

$$\varepsilon_{\text{vdW}}^{\perp} = \frac{1}{1 + \frac{L_{\text{cell}}}{L_{\text{vdW}}} \left(\frac{1}{\varepsilon_{c,\text{Ins-MoS}_2}^{\perp}} - \frac{1}{\varepsilon_{\text{Ins}}^{\perp}} - \frac{1}{\varepsilon_{c,\text{MoS}_2}^{\perp}} + 1 \right)} \quad (\text{S5})$$

In-Plane Direction: The in-plane response follows a volume-averaged (parallel capacitor) model.

For the *combined system*, one obtains:

$$L_{\text{cell}} \varepsilon_{c,\text{Ins-MoS}_2}^{\parallel} = L_{\text{Ins}} \varepsilon_{\text{Ins}}^{\parallel} + L_{\text{MoS}_2} \varepsilon_{\text{MoS}_2}^{\parallel} + L_{\text{vdW}} \varepsilon_{\text{vdW}}^{\parallel} + L_{\text{vac}} \quad (\text{S6})$$

Using $L_{\text{vac}} = L_{\text{cell}} - L_{\text{Ins}} - L_{\text{MoS}_2} - L_{\text{vdW}}$:

$$L_{\text{cell}} \left(\varepsilon_{c,\text{Ins-MoS}_2}^{\parallel} - 1 \right) = L_{\text{Ins}} \left(\varepsilon_{\text{Ins}}^{\parallel} - 1 \right) + L_{\text{MoS}_2} \left(\varepsilon_{\text{MoS}_2}^{\parallel} - 1 \right) + L_{\text{vdW}} \left(\varepsilon_{\text{vdW}}^{\parallel} - 1 \right) \quad (\text{S7})$$

Similarly, for the isolated cases:

$$L_{\text{cell}} \left(\varepsilon_{c,\text{Ins}}^{\parallel} - 1 \right) = L_{\text{Ins}} \left(\varepsilon_{\text{Ins}}^{\parallel} - 1 \right), \quad L_{\text{cell}} \left(\varepsilon_{c,\text{MoS}_2}^{\parallel} - 1 \right) = L_{\text{MoS}_2} \left(\varepsilon_{\text{MoS}_2}^{\parallel} - 1 \right) \quad (\text{S8})$$

Substituting into the expression yields:

$$L_{\text{cell}} \left(\varepsilon_{c,\text{Ins-MoS}_2}^{\parallel} - \varepsilon_{c,\text{Ins}}^{\parallel} - \varepsilon_{c,\text{MoS}_2}^{\parallel} + 1 \right) = L_{\text{vdW}} \left(\varepsilon_{\text{vdW}}^{\parallel} - 1 \right) \quad (\text{S9})$$

Solving for $\varepsilon_{\text{vdW}}^{\parallel}$:

$$\boxed{\varepsilon_{\text{vdW}}^{\parallel} = 1 + \frac{L_{\text{cell}}}{L_{\text{vdW}}} \left(\varepsilon_{\text{c,Ins-MoS}_2}^{\parallel} - \varepsilon_{\text{c,Ins}}^{\parallel} - \varepsilon_{\text{c,MoS}_2}^{\parallel} + 1 \right)} \quad (\text{S10})$$

Microscopic Approach: This section outlines the microscopic procedure used to extract the dielectric constant of a vdW gap from first-principles calculations. It complements the macroscopic perspective presented next by analyzing the spatial distribution of the dielectric response. The method involves evaluating the charge redistribution induced by a small external electric field applied along the out-of-plane (z) direction, as well as computing the resulting polarization profiles and effective electric fields. This approach provides local insights into the dielectric behavior across the slab (66, 67).

A homogeneous electric field E_{ext} is applied along the z -axis (fig. S1), inducing a redistribution of electronic charge. The corresponding difference in charge density is obtained as:

$$\rho_{\text{ind}}(z) = \rho^E(z) - \rho^0(z), \quad (\text{S11})$$

where $\rho^E(z)$ and $\rho^0(z)$ are the planar-averaged charge densities with and without the external electric field, respectively. The spatially varying polarization $P(z)$ is then computed by integrating the induced charge density:

$$P(z) = - \int_{z_0}^z \rho_{\text{ind}}(z') dz'. \quad (\text{S12})$$

The corresponding induced screening field $E_{\text{ind}}(z)$ and the total effective electric field $E_{\text{eff}}(z)$ are:

$$E_{\text{ind}}(z) = -\frac{1}{\varepsilon_0} P(z), \quad E_{\text{eff}}(z) = E_{\text{ext}} + E_{\text{ind}}(z). \quad (\text{S13})$$

The relative dielectric constant at each position z is then given by:

$$\varepsilon(z) = \frac{E_{\text{ext}}}{E_{\text{eff}}(z)} = 1 + \frac{P(z)}{\varepsilon_0 E_{\text{eff}}(z)}. \quad (\text{S14})$$

To mitigate atomic-scale oscillations in the dielectric profile, a Gaussian filter can be applied to $\rho_{\text{ind}}(z)$ and the derived quantities. The effective out-of-plane dielectric constant of the vdW gap is then computed as:

$$\boxed{\frac{1}{\varepsilon_{\text{vdW}}^{\perp}} = \frac{1}{L_{\text{vdW}}} \int_{z_1}^{z_2} \frac{1}{\varepsilon(z)} dz}, \quad (\text{S15})$$

where z_1 and z_2 define the boundaries of the vdW gap region, determined from atomic positions.

S2 Dielectric Thickness Dependence and Interfacial Effects

Dielectric materials at the nanoscale often deviate from bulk behavior due to finite-size effects, interfacial phenomena, and structural inhomogeneities. This section focuses on two key cases and the models describing them: (i) the layer-dependent dielectric constant observed in vdW materials, and (ii) dead layer effects typically associated with interfaces in high- κ dielectrics but also arising from grain boundaries, lattice defects, or other microstructural features. Both phenomena highlight the critical role of structural and interfacial effects in shaping the dielectric properties of nanoscale devices.

Thickness-Dependent Dielectric Constant in vdW Materials: vdW materials display distinctive dielectric behavior due to their layered structure and low-permittivity gaps between atomic planes. The effective dielectric constant often varies with the number of layers, especially in thin films, reflecting differences in polarization between interfacial and inner layers. This variation can be modeled as a series of dielectric regions: the atomic layers with similar permittivity, and the vdW gaps and interfacial layers, which differ from bulk properties. In ultrathin vdW structures, surface and interface contributions become significant, lowering the effective dielectric constant relative to bulk values. As the number of layers increases, this influence diminishes, and the dielectric constant approaches its bulk limit. This behavior is conceptually similar to the dead layer effect in thin films, where interfacial regions with reduced permittivity decrease the total capacitance.

The use of a series capacitance model to describe the out-of-plane dielectric constant in a layered vdW stack is illustrated in fig. S2(A), where the total thickness of an N -layer stack is given by: $t_{\text{total}} = t_{\text{SL}} + (N - 1)(t_{\text{slab}} + t_{\text{vdW}})$, where t_{SL} is the effective thickness of a single layer, t_{slab} the dielectric thickness near atomic planes, and t_{vdW} the vdW gap thickness. A common approximation is: $t_{\text{SL}} = t_{\text{slab}} + t_{\text{vdW}}$. The effective dielectric constant ϵ_{eff} follows:

$$\frac{t_{\text{total}}}{\epsilon_{\text{eff}}} = \frac{t_{\text{SL}}}{\epsilon_{\text{SL}}} + (N - 1) \left(\frac{t_{\text{slab}}}{\epsilon_{\text{slab}}} + \frac{t_{\text{vdW}}}{\epsilon_{\text{vdW}}} \right), \quad (\text{S16})$$

where ϵ_{SL} , ϵ_{slab} , and ϵ_{vdW} are the dielectric constants of the surface-influenced layer, the atomic-layer region, and the vdW gap, respectively. As N increases, interfacial effects become negligible, and ϵ_{eff} approaches the series combination of ϵ_{slab} and ϵ_{vdW} . Moreover, external factors such as strain can reduce the vdW gap thickness, enhancing polarization and increasing the dielectric constant, as shown in Fig. 2(e) of the manuscript.

Dead Layer Effects in Thin-Film Dielectrics: Interfacial dead layers commonly arise in high- κ dielectrics, particularly when integrated into ultrathin gate stacks. These regions near interfaces exhibit suppressed polarization and lower permittivity compared to the bulk (6, 68), as illustrated in fig. S2(B), leading to a degradation in effective dielectric response and reduced capacitance. Dead layers can also result from microstructural inhomogeneities such as grain boundaries, dislocations, or defects, which locally suppress dielectric screening. Although the focus here is on interfacial dead layers, other types can be similarly treated using a series capacitance model.

The overall stack is modeled as two low-permittivity interfacial dead layers of thickness t_{DL} (one on each side), sandwiching a central bulk region of thickness t_{bulk} and permittivity $\varepsilon_{\text{bulk}}$. The effective dielectric response of the entire stack is determined by the series combination of the bulk and dead-layer contributions. The effective permittivity of the dead layer is obtained from a graded inverse-permittivity profile. As demonstrated in *ab-initio* calculations (6), the inverse permittivity $\varepsilon^{-1}(z)$ within each dead layer can be assumed to vary linearly from vacuum ($\varepsilon = 1$ at $z = 0$) to the bulk value ($\varepsilon = \varepsilon_{\text{bulk}}$ at $z = t_{\text{DL}}$). This assumption yields:

$$\varepsilon_{\text{DL}}^{-1}(z) = \left(1 - \frac{z}{t_{\text{DL}}}\right) + \frac{z}{t_{\text{DL}}} \cdot \frac{1}{\varepsilon_{\text{bulk}}}. \quad (\text{S17})$$

The effective permittivity of the dead layer, $\varepsilon_{\text{DL}}^{\text{eff}}$, is then calculated by inverting the spatial average of $\varepsilon_{\text{DL}}^{-1}(z)$:

$$\varepsilon_{\text{DL}}^{\text{eff}} = \left[\frac{1}{t_{\text{DL}}} \int_0^{t_{\text{DL}}} \varepsilon^{-1}(z) dz \right]^{-1}. \quad (\text{S18})$$

Evaluating the integral leads to:

$$\varepsilon_{\text{DL}}^{\text{eff}} = \left[\frac{1}{2} \left(1 + \frac{1}{\varepsilon_{\text{bulk}}} \right) \right]^{-1} = \frac{2\varepsilon_{\text{bulk}}}{1 + \varepsilon_{\text{bulk}}}. \quad (\text{S19})$$

The parameter D is introduced to characterize the contribution of interfacial dead layers to the inverse capacitance per unit area in thin-film dielectrics. This effect is modeled by treating the dielectric stack as a series combination of the bulk and interfacial regions. The total effective permittivity $\varepsilon_{\text{tot}}^{\text{eff}}$ is expressed as:

$$\frac{t}{\varepsilon_{\text{tot}}^{\text{eff}}} = \frac{t_{\text{bulk}}}{\varepsilon_{\text{bulk}}} + \frac{2t_{\text{DL}}}{\varepsilon_{\text{DL}}^{\text{eff}}}, \quad (\text{S20})$$

where $t = t_{\text{bulk}} + 2t_{\text{DL}}$ denotes the total thickness. This expression can be rearranged to isolate the dead layer contribution (10):

$$\frac{t}{\varepsilon_{\text{tot}}^{\text{eff}}} = \frac{t}{\varepsilon_{\text{bulk}}} + D, \quad (\text{S21})$$

with

$$D = 2t_{\text{DL}} \left(\frac{1}{\varepsilon_{\text{DL}}^{\text{eff}}} - \frac{1}{\varepsilon_{\text{bulk}}} \right). \quad (\text{S22})$$

By substituting the expression for the effective permittivity of the dead layer from eq. S19, the parameter D becomes:

$$D = 2t_{\text{DL}} \left(\frac{1}{2} \left(1 + \frac{1}{\varepsilon_{\text{bulk}}} \right) - \frac{1}{\varepsilon_{\text{bulk}}} \right) = t_{\text{DL}} \left(1 - \frac{1}{\varepsilon_{\text{bulk}}} \right). \quad (\text{S23})$$

This formulation captures how regions with reduced permittivity near interfaces act in series with the bulk dielectric, thereby lowering the overall capacitance. As the total thickness t is scaled down, the relative impact of the dead layer increases. In the limiting case of an ideal bulk dielectric with infinite permittivity ($\varepsilon_{\text{bulk}} \rightarrow \infty$), the degradation parameter simplifies to $D \rightarrow t_{\text{DL}}$, providing a useful upper bound for the penalty imposed by interfacial dead layers in high- κ systems.

This model has been successfully applied to extract dead layer characteristics from experimental data. For instance, in the case of SrTiO₃ (STO), a bulk dielectric constant of 270 and a dead layer thickness of 1.61 Å were reported (10). Similarly, for HfO₂, fitting to thickness-dependent permittivity data yielded a bulk dielectric constant of 16.0 and a dead layer thickness of 1.18 Å, as shown in fig. S2(C).

The effective areal capacitance associated with both dead layers is given by:

$$C_{\text{DL}} = \frac{\varepsilon_0 \varepsilon_{\text{DL}}^{\text{eff}}}{2t_{\text{DL}}} = \frac{\varepsilon_0 \varepsilon_{\text{bulk}}}{t_{\text{DL}} (1 + \varepsilon_{\text{bulk}})}. \quad (\text{S24})$$

By substituting the earlier expression for $D = t_{\text{DL}} \left(1 - \frac{1}{\varepsilon_{\text{bulk}}} \right)$, the total dead-layer capacitance can be rewritten as:

$$C_{\text{DL}} = \frac{\varepsilon_0}{D} \cdot \frac{\varepsilon_{\text{bulk}} - 1}{\varepsilon_{\text{bulk}} + 1}. \quad (\text{S25})$$

This formulation emphasizes that the dead-layer capacitance is inversely proportional to the degradation parameter D , and is additionally suppressed by the polarization mismatch between vacuum and the bulk dielectric. The result quantitatively captures how interfacial suppression of the permittivity degrades the overall dielectric performance in ultrathin films. In the limiting case where $\varepsilon_{\text{bulk}} \rightarrow \infty$, the effective dead-layer capacitance approaches

$$C_{\text{DL}}^{\text{tot}} \rightarrow \frac{\varepsilon_0}{D}, \quad (\text{S26})$$

This sets an upper bound on the achievable capacitance limited solely by the interfacial degradation parameter D .

S3 Simplified Tsu–Esaki Model for Tunneling Current

The Tsu–Esaki formalism provides the foundation for rigorously modeling electron tunneling through a dielectric stack under elastic transport conditions. This model evaluates the total current as an energy-integrated product of the quantum mechanical transmission probability and the thermally broadened supply function. The general expression for the current density reads:

$$J = \frac{q}{2\pi^2\hbar} \int dE_z T(E_z) \left[\frac{m^*}{\hbar^2} \int_0^\infty (f_M(E_z + E_\perp) - f_{ch}(E_z + E_\perp)) dE_\perp \right], \quad (S27)$$

where $E = E_z + E_\perp$ is the total electron energy split into longitudinal and transverse components, and f_M, f_{ch} are the Fermi–Dirac distribution functions in the metallic gate and the channel, respectively. $T(E_z)$ is the transmission probability at energy E_z . Assuming parabolic, isotropic bands, the transverse integration can be performed analytically to give the supply function:

$$N(E_z) = k_B T \ln \left(\frac{1 + \exp [(\phi_M - E_z)/(k_B T)]}{1 + \exp [(E_{F,ch} - E_z)/(k_B T)]} \right)$$

where all energies are referenced from the vacuum level. Here, ϕ_M is the (negative) metal work function and $E_{F,ch}$ is the Fermi level of the channel. The total current becomes:

$$J = \frac{qm^*}{2\pi^2\hbar^3} \int T(E) N(E) dE, \quad (S28)$$

The subscript z is omitted, and E is used to denote the longitudinal energy for simplicity. In tunneling-dominated regimes, the supply function decays exponentially with increasing energy, so the dominant contributions arise from energies near the conduction band edge of the channel:

$$E = E_{c,ch} = \chi_{ch} - qV_g, \quad (S29)$$

where χ_{ch} is the electron affinity of the channel, and V_g is the applied gate voltage. Assuming the gate as the electrostatic reference, the channel band edge varies with the applied gate voltage. In typical operation, the conduction band edge of the channel lies several thermal energies (more than $4k_B T$) above both the channel Fermi level and the work functions of most metals used for contacts. This places the system in the non-degenerate regime, where occupation probabilities are sufficiently low to approximate the Fermi–Dirac statistics by a Maxwell–Boltzmann distribution. In addition, the channel-side occupation can be neglected to yield:

$$N(E) \approx k_B T \exp \left(\frac{\phi_M - E}{k_B T} \right) = k_B T \exp \left(\frac{\phi_M - \chi_{ch} + qV_g}{k_B T} \right) \approx k_B T \exp \left(-\frac{\phi_B}{k_B T} \right) \exp \left(\frac{qV_g}{k_B T} \right). \quad (S30)$$

where the effective barrier height is given by

$$\phi_B = \chi_M - \phi_{ch}. \quad (S31)$$

Assuming $T(E)$ is approximately constant over a thermal window $\Delta E \sim k_B T$, the current becomes:

$$J \approx \left(\frac{qm^*}{2\pi^2\hbar^3} k_B T \right) T(E) N(E). \text{ Thus,}$$

$$J \approx \underbrace{\frac{qm^*(k_B T)^2}{2\pi^2\hbar^3} \exp\left(-\frac{\phi_B}{k_B T}\right) \exp\left(\frac{qV_g}{k_B T}\right)}_{J_0} T(E_{c,ch}), \quad (S32)$$

where the transmission probability is evaluated at the energy corresponding to the conduction band edge of the channel, as defined in eq. S29. Using the trapezoidal barrier approximation, the transmission at this energy can be estimated as

$$T(E_{c,ch}) \approx \exp\left[-\frac{4\sqrt{2m^*}}{3\hbar q E_{ins}} \left((\Delta E + qV_g)^{3/2} - \Delta E^{3/2} \right)\right], \quad \Delta E = \chi_{ch} - \chi_{ins}, \quad E_{ins} = \frac{V_g}{t_{ins}}, \quad (S33)$$

with ΔE representing the band offset between the conduction bands of the channel and the insulator, and $E_{ins} = V_g/t_{ins}$ denoting the electric field across the insulator of thickness t_{ins} . The prefactor in eq. S32 resemble those of the classical Richardson equation for thermionic emission:

$$J = A^* T^2 \exp\left(-\frac{q\phi_B}{k_B T}\right), \quad \text{with} \quad A^* = \frac{qm^* k_B^2}{2\pi^2\hbar^3} \approx 120 \text{ A/cm}^2 \text{ K}^2 \quad \text{for } m^* = m_0. \quad (S34)$$

This similarity highlights that the Tsu–Esaki model extends the classical Richardson emission by incorporating sub-barrier tunneling through a transmission coefficient $T(E) \ll 1$. While the Richardson model assumes perfect transmission $T(E) = 1$ above the barrier, the Tsu–Esaki model explicitly accounts for quantum tunneling across the energy barrier. The gate voltage dependence enters through both the supply function and the barrier height, leading to an exponential modulation of the tunneling current.

In the case of the insulator–MoS₂ stacks examined in this work, the gate metal work function is taken as 5.2 eV and the MoS₂ electron affinity as 4.3 eV, resulting in a barrier height of $\phi_B = 0.9$ eV (eq. S31). To evaluate the tunneling current according to IRDS guidelines for 2035 nodes, a gate voltage of 0.6 V was assumed which yields

$$J_0 \approx 100 \text{ A/cm}^2 \quad \text{with} \quad \phi_B = 0.9 \text{ eV}, \quad V_G = 0.6 \text{ V}. \quad (S35)$$

Remark: While this simplified model yields a closed-form prefactor J_0 for estimating the minimum EOT (Section S5) and captures key scaling trends, it does not replace full quantum transport.

S4 Two-Band Model

When electrons tunnel through a wide-gap insulator, their wavefunctions decay exponentially in the classically forbidden energy range. This behavior is described by the complex band structure, which extends conventional band theory to complex wavevectors. Inside the gap, propagating Bloch states become evanescent states characterized by an imaginary wavevector component. The decay rate is quantified by the inverse decay length, defined as $\beta(E) = 2 \operatorname{Im}\{k(E)\}$.

A widely used approximation is the two-band Franz model (69), which assumes symmetric parabolic conduction and valence bands. It expresses $\beta(E)$ as:

$$\beta(E) = \frac{2}{\hbar} \sqrt{2m^* \frac{(E - E_v)(E_c - E)}{E_g}}, \quad (\text{S36})$$

where m^* is the tunneling effective mass, E_v and E_c are the valence and conduction band edges, and $E_g = E_c - E_v$ is the band gap. In terms of an energy offset ΔE from a band edge, eq. S36 becomes:

$$\boxed{\beta(\Delta E) = \frac{2}{\hbar} \sqrt{2m^* \Delta E \left(1 - \frac{\Delta E}{E_g}\right)}}, \quad (\text{S37})$$

with $\Delta E = E_c - E$ for electrons and $\Delta E = E - E_v$ for holes. This form clearly shows that the decay is minimal near the bandedges and has a maximum at midgap.

To account for different effective masses in the conduction (m_c) and valence (m_v) bands, the model can be generalized into an elliptic two-band approximation (70), yielding:

$$\beta(\Delta E) = \begin{cases} \frac{2}{\hbar} \sqrt{2m_v \Delta E \left(1 - \frac{\Delta E}{2E_{q,v}}\right)}, & \text{holes: } 0 \leq \Delta E = E - E_v \leq E_{q,v}, \\ \frac{2}{\hbar} \sqrt{2m_c \Delta E \left(1 - \frac{\Delta E}{2E_{q,c}}\right)}, & \text{electrons: } 0 \leq \Delta E = E_c - E \leq E_{q,c}, \end{cases} \quad (\text{S38})$$

where the transition energies are:

$$E_{q,v} = \frac{m_c}{m_c + m_v} E_g, \quad E_{q,c} = \frac{m_v}{m_c + m_v} E_g.$$

The decay reaches its maximum at the branch-point energy $E_q = E_v + E_{q,v} = E_c - E_{q,c}$.

For electron tunneling from SiO_2 into Si, the relevant barrier is the conduction band offset $\Delta E_{\text{SiO}_2} \approx 3.1$ eV (71, 72). Using $m_c \approx 0.42 m_0$, $m_v \approx 0.33 m_0$, and $E_g = 8.9$ eV for SiO_2 , the resulting inverse decay length from eq. S38 is:

$$\boxed{\beta_{\text{SiO}_2} \simeq 0.91 \text{ \AA}^{-1}.} \quad (\text{S39})$$

S5 Insulator Figure of Merit and Minimum Achievable EOT

As described in Section S3, the gate leakage current in a metal–insulator–semiconductor structure is strongly influenced by the transmission probability $T(E)$ through the insulator, which can be approximated using the WKB method:

$$T(E) \approx \exp\left[-\int_{z_1}^{z_2} \beta(z, E) dz\right], \quad (\text{S40})$$

where

$$\beta(z, E) = \frac{2}{\hbar} \sqrt{2 m^* \Delta E(z, E)} \quad (\text{S41})$$

is the local inverse decay length and $\Delta E(z, E)$ the energy barrier between the carrier energy E and the relevant band edge of the insulator: $\Delta E = E_{c,\text{ins}}(z) - E$ for electrons and $\Delta E = E - E_{v,\text{ins}}(z)$ for holes. The analysis that follows focuses on electrons; analogous relations apply for holes. In the flat-band limit, where the band edges are spatially uniform, $E_{c,\text{ins}}(z) \simeq E_{c,\text{ins}}$, the integral simplifies to

$$T(E) \approx \exp(-\beta(E) t_{\text{ins}}), \quad (\text{S42})$$

with $\beta(E) = 2\sqrt{2 m^* \Delta E}/\hbar$, $\Delta E = E_{c,\text{ins}} - E$ for electron tunnelling, and t_{ins} denotes the physical thickness of the dielectric. Because the Fermi–Dirac tail decays exponentially, the dominant contribution to the integral stems from electrons whose longitudinal energy lies close to the band edge of the semiconductor channel. The transmission is therefore evaluated at the channel conduction-band edge $E_{c,\text{ch}}$. As illustrated in fig. S3(B), the relevant barrier height is the conduction-band offset,

$$\Delta E = E_{c,\text{ins}} - E_{c,\text{ch}} = \chi_{\text{ch}} - \chi_{\text{ins}},$$

where χ_{ch} and χ_{ins} denote, respectively, the electron affinities of the channel and of the insulator. (For hole tunnelling the analogous quantity is $\Delta E = E_{v,\text{ch}} - E_{v,\text{ins}}$.) This band-offset ΔE is the height that enters the inverse decay length $\beta(E)$. Using the simplified Tsu–Esaki model in eq. S32, one can approximately evaluate the tunneling current as:

$$J \approx J_0 T(E_{c,\text{ch}}) = J_0 \exp(-\beta t_{\text{ins}}), \quad (\text{S43})$$

where J_0 includes prefactors and the supply term and the inverse of the decay length is given by

$$\beta = \frac{2}{\hbar} \sqrt{2m^* \Delta E} = \frac{2}{\hbar} \sqrt{2m^* (E_{c,\text{ins}} - E_{c,\text{ch}})} \quad (\text{S44})$$

To provide a more accurate reference for benchmarking tunneling suppression, one can utilize the decay constant using the complex band structure formalism described in eq. S37 or eq. S38.

The physical thickness is related to electrostatic scaling by expressing it in terms of the EOT:

$$t_{\text{ins}} = \frac{\varepsilon}{\varepsilon_{\text{SiO}_2}} \text{EOT}, \quad (\text{S45})$$

where ε is the dielectric constant of the insulator. Substituting, the current becomes:

$$J = J_0 \exp \left(-\beta \frac{\varepsilon}{\varepsilon_{\text{SiO}_2}} \text{EOT} \right). \quad (\text{S46})$$

It is convenient to normalise every material to thermally-grown SiO_2 by factoring out its WKB decay parameter expressed in eq. S39. The dimensionless figure of merit (FoM) is introduced using this reference:

$$\boxed{\text{FoM} = \frac{\varepsilon}{\varepsilon_{\text{SiO}_2}} \frac{\beta}{\beta_{\text{SiO}_2}} = \frac{\varepsilon}{\varepsilon_{\text{SiO}_2}} \sqrt{\frac{m^* \Delta E}{m_{\text{SiO}_2}^* \Delta E_{\text{SiO}_2}}}}, \quad (\text{S47})$$

so that the leakage current can be written in a compact form as

$$\boxed{J = J_0 \exp[-\beta_{\text{SiO}_2} \text{FoM EOT}].} \quad (\text{S48})$$

By construction, the FoM is normalized to unity for SiO_2 . Values larger (smaller) than 1 indicate superior (inferior) suppression of gate tunneling per unit EOT. This is because the FoM encapsulates both electrostatic enhancement, via the permittivity ε , and quantum tunneling suppression, via the inverse decay length β . Accordingly, the dimensionless FoM introduced in eq. S47 can also be refined using the two-band model as shown in Section S4.

Physically, the FoM quantifies how effectively a given insulator suppresses the tunneling current for each unit increase in EOT. It is proportional to the product of the relative dielectric constant ε and the square root of the product of the tunneling effective mass and the conduction band offset between the insulator and the channel material (e.g., MoS_2). This reflects the decay rate of the electron wavefunction across the dielectric. A larger FoM implies stronger suppression of gate leakage for a given EOT, thereby enabling more aggressive scaling of gate control while maintaining acceptable leakage limits. Rearranging eq. S47 yields the minimum achievable EOT under a fixed leakage current target $J = J_{\text{target}}$, where the IRDS roadmap specifies $J_{\text{target}} = 1.5 \times 10^{-2} \text{ A/cm}^2$ for low-power applications (2), as:

$$\boxed{\text{EOT}_{\min} = \frac{1}{\beta_{\text{SiO}_2} \text{FoM}} \ln \left(\frac{J_0}{J_{\text{target}}} \right)}. \quad (\text{S49})$$

Dead Layer Effects: Incorporation of dead layer effects is achieved via the effective permittivity model outlined in Section S2. By inserting the effective dielectric constant from eq. S21 into the EOT relation in eq. S45, the resulting expression is:

$$\text{EOT} = \left(\frac{t}{\varepsilon_{\text{bulk}}} + D \right) \varepsilon_{\text{SiO}_2}. \quad (\text{S50})$$

Solving for the physical thickness t yields:

$$t = \frac{\varepsilon_{\text{bulk}}}{\varepsilon_{\text{SiO}_2}} (\text{EOT} - D\varepsilon_{\text{SiO}_2}). \quad (\text{S51})$$

Upon substitution of this thickness into eq. S43, the current is given by:

$$J = J_0 \exp \left[-\beta \frac{\varepsilon_{\text{bulk}}}{\varepsilon_{\text{SiO}_2}} (\text{EOT} - D\varepsilon_{\text{SiO}_2}) \right]. \quad (\text{S52})$$

This can be rearranged as:

$$J = J_0 \exp \left[-\beta_{\text{SiO}_2} \text{FoM}_{\text{bulk}} (\text{EOT} - D\varepsilon_{\text{SiO}_2}) \right] \quad (\text{S53})$$

where the bulk FoM, in the absence of a dead layer, is defined as:

$$\text{FoM}_{\text{bulk}} = \frac{\varepsilon_{\text{bulk}}}{\varepsilon_{\text{SiO}_2}} \frac{\beta}{\beta_{\text{SiO}_2}}. \quad (\text{S54})$$

The minimum EOT corresponding to a fixed target current $J = J_{\text{target}}$ is obtained as:

$$\text{EOT}_{\text{total}}^{\min} = \text{EOT}_{\text{bulk}}^{\min} + \text{EOT}_{\text{dead}} = \frac{1}{\beta_{\text{SiO}_2} \text{FoM}_{\text{bulk}}} \ln \left(\frac{J_0}{J_{\text{target}}} \right) + D\varepsilon_{\text{SiO}_2}. \quad (\text{S55})$$

This final expression separates two key contributions: an ideal minimum EOT term determined by bulk material properties, and an additive penalty from dead layer effects. For example, although bulk SrTiO₃ (STO) has a dielectric constant exceeding 270 at room temperature, dead layers greatly reduce the effective response in ultrathin films, limiting the benefit of STO's high- κ permittivity. Using reported STO parameters from (10), with a minimal dead layer parameter of $D = 1.6 \text{ \AA}$, the increase in minimum achievable EOT is:

$$\text{EOT}_{\text{dead}}^{\text{STO}} = D \times \varepsilon_{\text{SiO}_2} = 1.61 \text{ \AA} \times 3.9 \approx 6.3 \text{ \AA}. \quad (\text{S56})$$

This offset imposes a hard limit on gate scaling with ultra high- κ dielectrics like STO.

Electronic Polarizability in Dead-Layer Models: In the previous section, it was assumed that the permittivity scales nearly inversely with thickness. This effect is expected to originate primarily from the ionic contribution to permittivity, which is degraded by interfacial effects. By contrast, the electronic permittivity is reasonably assumed to remain largely unaffected by thickness variation. Considering this behavior, the minimum achievable EOT in the dead layer model must be revised. It is assumed here that ε^{el} remains nearly constant, while the ionic component ε^{ion} follows a thickness-dependent dead layer model. The total effective permittivity is expressed as the sum of the electronic and ionic contributions: $\varepsilon_{\text{eff}} = \varepsilon^{\text{el}} + \varepsilon^{\text{ion}}$. According to the dead layer model, the ionic permittivity is modified as follows:

$$\frac{t}{\varepsilon^{\text{ion}}} = \frac{t}{\varepsilon_{\text{bulk}}^{\text{ion}}} + D \quad \Rightarrow \quad \varepsilon^{\text{ion}} = \frac{\varepsilon_{\text{bulk}}^{\text{ion}}}{1 + D\varepsilon_{\text{bulk}}^{\text{ion}}/t}. \quad (\text{S57})$$

For simplicity, the bulk permittivity $\varepsilon_{\text{bulk}} = \varepsilon^{\text{el}} + \varepsilon_{\text{bulk}}^{\text{ion}}$ and $\eta = \varepsilon^{\text{el}}/\varepsilon_{\text{bulk}}$, $1 - \eta = \varepsilon_{\text{bulk}}^{\text{ion}}/\varepsilon_{\text{bulk}}$ are defined. With these definitions, the effective permittivity becomes:

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{bulk}} \left(\eta + (1 - \eta) \frac{1}{1 + D(1 - \eta)\varepsilon_{\text{bulk}}/t} \right). \quad (\text{S58})$$

The corresponding physical thickness associated with the EOT_{\min} is given by:

$$t_{\min} = \frac{\ln(J_0/J_{\text{target}})}{\beta}, \quad \text{where} \quad \beta = \frac{\text{FoM}^{\text{bulk}}\beta_{\text{SiO}_2}}{\varepsilon_{\text{bulk}}}. \quad (\text{S59})$$

Substituting t_{\min} and β yields the minimum achievable EOT:

$$\text{EOT}_{\min} = t_{\min}\varepsilon_{\text{SiO}_2}/\varepsilon_{\text{eff}} = \frac{\ln(J_0/J_{\text{target}})}{\beta} \varepsilon_{\text{SiO}_2} \left[\varepsilon_{\text{bulk}} \left(\eta + \frac{1 - \eta}{1 + (1 - \eta)D\varepsilon_{\text{bulk}}/t} \right) \right]^{-1}. \quad (\text{S60})$$

After simplification, and using the definition of the bulk figure of merit,

$$\text{FoM}^{\text{bulk}} = \frac{\varepsilon_{\text{bulk}}\beta}{\varepsilon_{\text{SiO}_2}\beta_{\text{SiO}_2}}, \quad \text{EOT}_{\min}^{\text{bulk}} = \frac{\ln(J_0/J_{\text{target}})}{\text{FoM}^{\text{bulk}}\beta_{\text{SiO}_2}}, \quad (\text{S61})$$

the final compact expression becomes:

$$\text{EOT}_{\min} = \text{EOT}_{\min}^{\text{bulk}} \left[\eta + \frac{1 - \eta}{1 + (1 - \eta)D\varepsilon_{\text{SiO}_2}/\text{EOT}_{\min}^{\text{bulk}}} \right]^{-1}. \quad (\text{S62})$$

This general result reduces to the pure dead-layer model when $\varepsilon^{\text{el}} \rightarrow 0 \Rightarrow \eta \rightarrow 0$, and recovers the bulk permittivity model in the limit $D \rightarrow 0$.

vdW Gap Effects: To assess the impact of the vdW gap on leakage and scaling, the gate stack is modeled as two sequential tunneling barriers: (i) an insulator of thickness t_{ins} and relative permittivity ϵ_{ins} , and (ii) a vdW gap of thickness t_{vdW} and permittivity ϵ_{vdW} . Within the WKB approximation and assuming phase-coherent transport across the ultrathin insulator and vdW gap, the total tunneling current density through the stack can be approximated by the product of the transmission probabilities through the individual regions. Accordingly, the total current density is given by: $J \approx J_0 \exp(-\beta_{\text{ins}} t_{\text{ins}} - \beta_{\text{vdW}} t_{\text{vdW}})$, where β_{ins} and β_{vdW} are the inverse decay lengths in the insulator and the vdW gap, respectively. Electrostatic scaling is captured by relating the physical thicknesses to their EOT contributions via:

$$t_{\text{ins}} = \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{\text{ins}}} \text{EOT}_{\text{ins}}, \quad t_{\text{vdW}} = \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{\text{vdW}}} \text{EOT}_{\text{vdW}}. \quad (\text{S63})$$

Substituting into the tunneling expression and using the FoM from eq. S47 yields:

$$J = J_0 \exp \left[-\beta_{\text{SiO}_2} (\text{FoM}_{\text{ins}} \text{EOT}_{\text{ins}} + \text{FoM}_{\text{vdW}} \text{EOT}_{\text{vdW}}) \right], \quad (\text{S64})$$

where the following dimensionless FoMs is defined:

$$\text{FoM}_{\text{ins}} = \frac{\epsilon_{\text{ins}}}{\epsilon_{\text{SiO}_2}} \sqrt{\frac{m_{\text{ins}}^* \Delta E_{\text{ins}}}{m_{\text{SiO}_2}^* \Delta E_{\text{SiO}_2}}}, \quad \text{FoM}_{\text{vdW}} = \frac{\epsilon_{\text{vdW}}}{\epsilon_{\text{SiO}_2}} \sqrt{\frac{m_{\text{vdW}}^* \Delta E_{\text{vdW}}}{m_{\text{SiO}_2}^* \Delta E_{\text{SiO}_2}}}. \quad (\text{S65})$$

For the insulator, the conduction band offset is given by $\Delta E_{\text{ins}} = \chi_{\text{MoS}_2} - \chi_{\text{ins}}$. Tunneling through the vdW gap is modeled as vacuum tunneling, consistent with DFT results (Fig. 4(B)), using a barrier height $\Delta E_{\text{vdW}} = \chi_{\text{MoS}_2}$. Lacking lattice and bonding, the vdW gap imposes vacuum-like conditions, justifying the use of $m_{\text{vdW}}^* = m_0$. To determine the minimum achievable EOT under a fixed gate leakage target $J = J_{\text{target}}$, the tunneling exponent must satisfy:

$$\beta_{\text{SiO}_2} (\text{FoM}_{\text{ins}} \text{EOT}_{\text{ins}} + \text{FoM}_{\text{vdW}} \text{EOT}_{\text{vdW}}) = \ln(J_0/J_{\text{target}}). \quad (\text{S66})$$

Solving for EOT_{ins} gives:

$$\text{EOT}_{\text{ins}} = \frac{\ln(J_0/J_{\text{target}}) - \beta_{\text{SiO}_2} \text{FoM}_{\text{vdW}} \text{EOT}_{\text{vdW}}}{\beta_{\text{SiO}_2} \text{FoM}_{\text{ins}}}, \quad (\text{S67})$$

where the first term, $\text{EOT}_{\text{ins}}^{\min} = \ln(J_0/J_{\text{target}})/(\beta_{\text{SiO}_2} \text{FoM}_{\text{ins}})$, represents the minimum achievable EOT in the absence of a vdW gap; the total minimum achievable EOT is therefore given by:

$$\boxed{\text{EOT}_{\text{total}}^{\min} = \text{EOT}_{\text{ins}}^{\min} + \text{EOT}_{\text{vdW}} = \frac{\ln(J_0/J_{\text{target}})}{\beta_{\text{SiO}_2} \text{FoM}_{\text{ins}}} + \left(1 - \frac{\text{FoM}_{\text{vdW}}}{\text{FoM}_{\text{ins}}}\right) \text{EOT}_{\text{vdW}}.} \quad (\text{S68})$$

The second term represents the modification introduced by the vdW layer. It contributes negatively—i.e., beneficially—only when: $\text{FoM}_{\text{ins}} < \text{FoM}_{\text{vdW}}$. Substituting the FoM definitions into this inequality and using $m_{\text{ins}}^* = m_r m_0$, $\Delta E_{\text{vdW}} = \chi_{\text{MoS}_2}$, and $\Delta E_{\text{ins}} = \chi_{\text{MoS}_2} - \chi_{\text{ins}}$, the condition simplifies to:

$$m_r \left(1 - \frac{\chi_{\text{ins}}}{\chi_{\text{MoS}_2}}\right) < \left(\frac{\varepsilon_{\text{vdW}}}{\varepsilon_{\text{ins}}}\right)^2. \quad (\text{S69})$$

This inequality defines the regime where the inclusion of a vdW gap improves the leakage–EOT tradeoff by offering more effective tunneling suppression per unit electrostatic thickness than the insulator alone. To illustrate this condition, fig. S4(A) presents the boundary in the $(\varepsilon_{\text{ins}}, \chi_{\text{ins}})$ space beyond which a vdW gap reduces the minimum achievable EOT. A clear dependence of the normalized EOT modification, $\Delta \text{EOT}/\text{EOT}_{\text{vdW}}$, on the insulator permittivity for various representative parameters is illustrated in fig. S4(B).

vdW Gap Average EOT and FoM: For all studied insulator–MoS₂ stacks, an analytical model of the form $\varepsilon_{\text{vdW}}^{-1} = 1 - c/t_{\text{vdW}}$, derived in eq. S5 based on polarization enhancement from wavefunction overlap, was fitted to the extracted dielectric constants. For each insulator–MoS₂ interface, a best-fit parameter c was obtained individually. The average value across all systems was found to be $\bar{c} = 0.72 \text{ \AA}$, as summarized in table S1. Based on this model and the fitted \bar{c} , the average dielectric constant and EOT of the vdW gap were determined as $\bar{\varepsilon}_{\text{vdW}} \approx 2$ and $\overline{\text{EOT}}_{\text{vdW}} \approx 2.7 \text{ \AA}$, respectively.

For electron tunneling through the vdW gap, a barrier height equal to the electron affinity of MoS₂, $\chi_{\text{MoS}_2} = 4.30 \text{ eV}$, and a shape factor of $\alpha = 0.8$ were assumed. The corresponding tunneling decay constant is:

$$\beta_{\text{vdW}} = \frac{2\alpha}{\hbar} \sqrt{2 m_0 \chi_{\text{MoS}_2}} = 1.7 \text{ \AA}^{-1}. \quad (\text{S70})$$

From these values, the FoM for the vdW gap is calculated as:

$$\boxed{\overline{\text{FoM}}_{\text{vdW}} = \frac{\bar{\varepsilon}_{\text{vdW}} \times \beta_{\text{vdW}}}{\varepsilon_{\text{SiO}_2} \times \beta_{\text{SiO}_2}} = \frac{2.03 \times 1.7}{3.9 \times 0.91} \approx 1,} \quad (\text{S71})$$

where $\beta_{\text{SiO}_2} = 0.91 \text{ \AA}^{-1}$ was obtained from eq. S39.

Finally, the minimum achievable total EOT, including the vdW gap, can be expressed as:

$$\text{EOT}_{\text{total}}^{\min} = \frac{8.79}{0.91 \times \text{FoM}_{\text{ins}}} + \left(1 - \frac{1}{\text{FoM}_{\text{ins}}}\right) 2.73 = \frac{7 \text{ \AA}}{\text{FoM}_{\text{ins}}} + 2.73 \text{ \AA}. \quad (\text{S72})$$

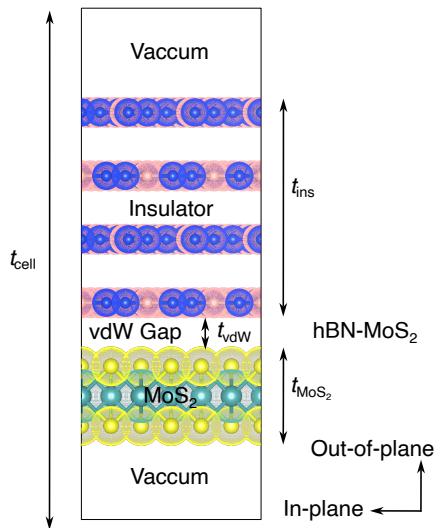


Figure S1: Simulation Cell Geometry for Macroscopic Dielectric Calculations. Schematic diagram of the simulation supercell consisting of an insulator, vdW gap, and MoS₂ layer, each surrounded by vacuum. The total length of the supercell is L_{cell} , with labeled thicknesses L_{Ins} , L_{vdW} , and L_{MoS_2} , indicating the lengths of the insulator, vdW gap, and MoS₂, respectively. Layer thicknesses are illustrative and not drawn to scale. This layered geometry is used in the microscopic and macroscopic dielectric analysis described in this and the following section.

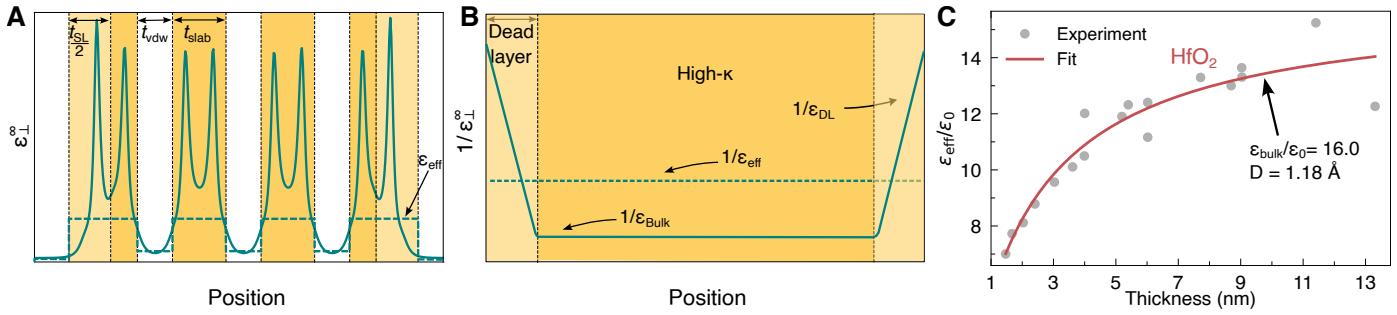


Figure S2: Dielectric Profile Variations Across vdW Materials and Dead Layers. (A) Schematic spatial profile of the dielectric constant along a vdW material, showing peaks around the atomic layers and minima between the layers corresponding to the vdW gaps. The region inside the material, labeled as the slab (filled in soft pink), represents the atomic layer plus the covalent radii of atoms at the interfaces and is characterized by an effective dielectric constant ϵ_{slab} . The vdW gap between slabs is shown in white and has an effective dielectric constant ϵ_{vdw} . Half-layers at the interfaces, which exhibit different polarizability compared to the inner layers, are indicated by darker pink regions. Together, these two half-layers form the single-layer (SL) region, characterized by an effective dielectric constant ϵ_{SL} . (B) Schematic of a high- κ insulator showing dead layers at both interfaces. The inverse dielectric function $1/\epsilon(z)$ transitions linearly from the bulk value to unity (vacuum-like) near the interfaces. (C) Thickness-dependent dielectric constant of HfO_2 . Experimental data from Ref. (49) are shown, along with fitted curves using the dead layer model from eq. S21.

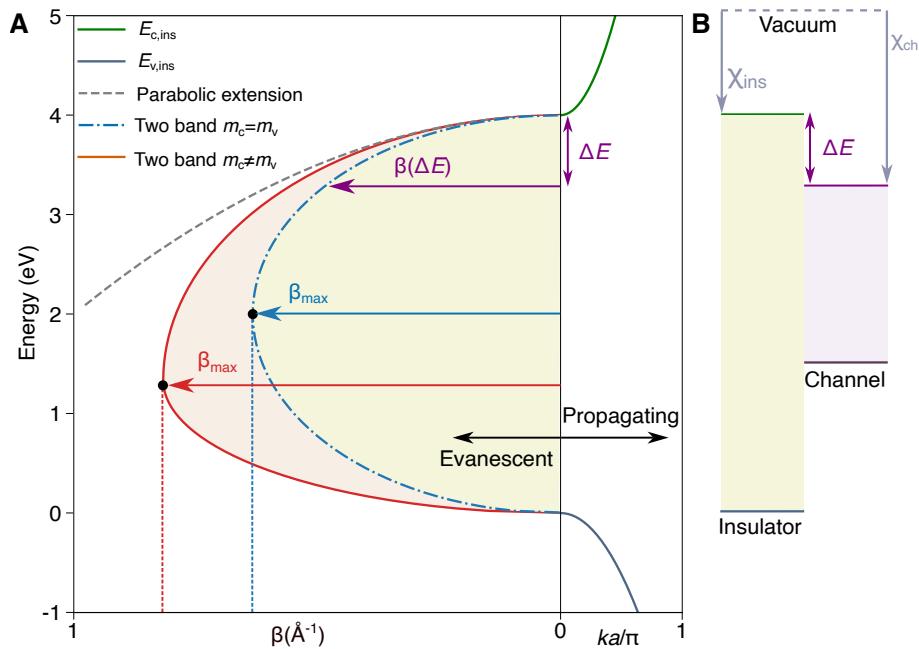


Figure S3: Complex Band Structure Models and Tunneling Decay Parameters. (A) Schematic comparison of three complex band-structure models. Right: real k -space dispersion for the conduction and valence bands. Left: corresponding inverse decay length $\beta(E)$. Dashed curve: single-parabolic continuation. Dash-dotted curve: symmetric two-band (Franz) model. Solid curve: asymmetric two-band model illustrating the branch-point shift that arises from unequal electron and hole effective masses. The purple curve $\beta(\Delta E)$ marks the inverse decay length evaluated at an energy offset ΔE measured from the conduction-band minimum. (B) Schematic band alignment at an insulator/semiconductor interface. The electron affinities, χ , of the two materials determine the conduction-band offset, ΔE , which sets the height of the tunneling barrier that electrons must overcome to traverse the insulator and reach the channel's conduction-band edge. The inverse decay length associated with this barrier height is shown as the purple curve in part (A).

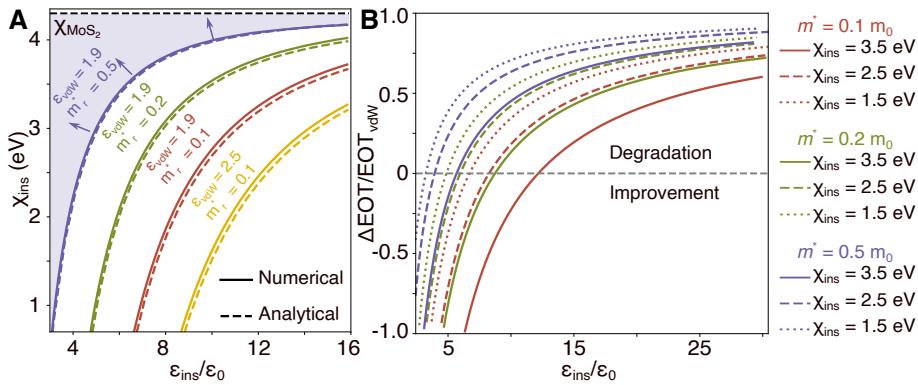


Figure S4: Analytical Boundaries and EOT Impact of vdW Gaps in Insulator Scaling. (A)

Analytical boundaries defined by eq. S69 for different values of m_r^* . Points in the region above each curve correspond to combinations of $(\epsilon_{\text{ins}}, \chi_{\text{ins}})$ where introducing a vdW gap reduces the minimum achievable EOT. Dashed curves: analytical model; solid curves: full numerical results. (B) $\Delta \text{EOT}/\text{EOT}_{\text{vdW}}$ vs. insulator dielectric constant for various χ_{ins} and m_{ins}^* . Negative values indicate improved EOT performance. $\epsilon_{\text{vdW}}/\epsilon_0 = 2$, $\alpha = 0.8$, and $\chi_{\text{MoS}_2} = 4.3$ eV.

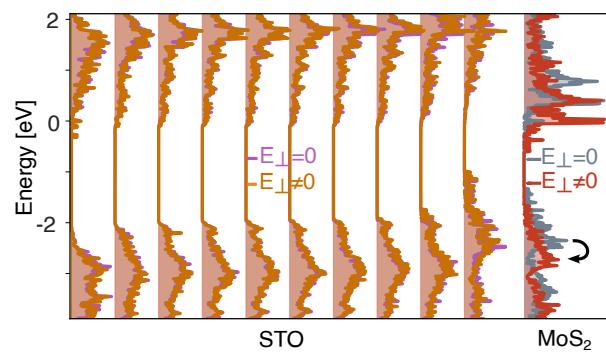


Figure S5: Layer-resolved density of states (DOS) of the STO-MoS₂ heterostructure. Layer-resolved DOS illustrating the interaction between sulfur atoms in MoS₂ and the interfacial layer of STO. In the absence of an electric field, the DOS in the middle of the STO slab remains largely unaffected, whereas pronounced variations emerge at the interface under an applied field due to reduced screening effects caused by the lower permittivity at the interface.

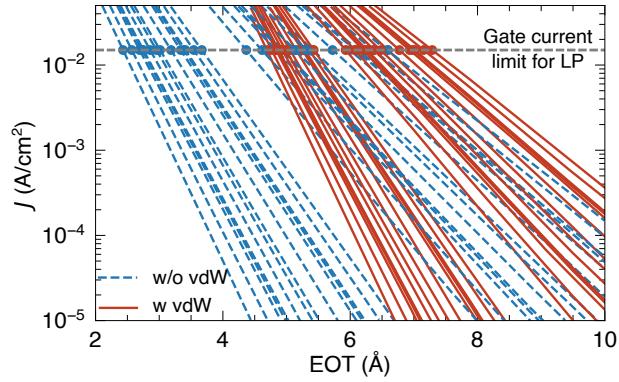


Figure S6: Determination of the minimum achievable EOT. The tunneling current is computed as a function of EOT for each MoS₂–insulator stack, using the corresponding material parameters (dielectric constant, effective mass, and conduction band offset). The minimum achievable EOT is defined as the point where the simulated current intersects the target leakage current. This analysis is conducted both without and with a vdW gap. The highlighted markers on each curve indicate the minimum EOT obtained for each parameter set. The figure shows representative results for CaF₂–MoS₂, using the parameter ranges provided in (18).

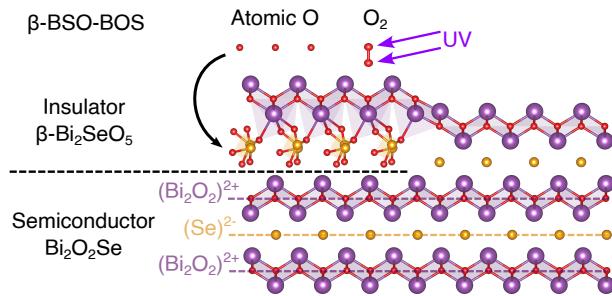


Figure S7: Schematic of a Zipper Interface in β -BSO-BOS Heterostructure. Schematic illustration of a zipper interface formed in the β -BSO-BOS heterostructure. The interface creates intermediate bonds that are stronger than vdW interactions but not fully covalent, effectively avoiding interfacial gaps. β -BSO is the native oxide of BOS.

Insulator	LaF ₃	LaOCl	Sb ₂ O ₃	CaF ₂	α -BSO	HfO ₂	STO	hBN	Average (Data)	Analytical (Model)
t_{vdW} (Å)	1.27	1.43	1.46	1.39	1.33	1.48	1.41	1.59	1.42	–
ε_{vdW}	2.03	1.89	2.18	1.72	2.11	2.10	2.66	1.79	2.06	2.03
EOT (Å)	2.44	2.95	2.61	3.14	2.45	2.74	2.06	3.47	2.73	2.73
c (Å)	0.64	0.67	0.79	0.58	0.70	0.77	0.88	0.70	0.72	–

Table S1: Parameters of vdW gaps in insulator–MoS₂ stacks. Average values across all listed material stacks are calculated as arithmetic means: gap thickness $\bar{t}_{\text{vdW}} = 1.42$ Å, permittivity $\bar{\varepsilon}_{\text{vdW}}/\varepsilon_0 = 2.06$, $\overline{\text{EOT}} = 2.73$ Å, and fitted coefficient $\bar{c} = 0.72$ Å. Analytical values are obtained from the dielectric fit model using the average c . The close agreement between the arithmetic averages and model predictions supports the validity of the dielectric approximation.