



**HAL**  
open science

## Cold-source paradigm for steep-slope transistors based on van der Waals heterojunctions

D. Logoteta, J. Cao, Marco G. Pala, P. Dollfus, Y. Lee, G. Iannaccone

► **To cite this version:**

D. Logoteta, J. Cao, Marco G. Pala, P. Dollfus, Y. Lee, et al.. Cold-source paradigm for steep-slope transistors based on van der Waals heterojunctions. *Physical Review Research*, 2020, 2 (4), <10.1103/PhysRevResearch.2.043286>. <hal-03032703>

**HAL Id: hal-03032703**

**<https://hal.science/hal-03032703v1>**

Submitted on 16 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Cold-source paradigm for steep-slope transistors based on van der Waals heterojunctions

D. Logoteta<sup>1,\*</sup>, J. Cao<sup>2</sup>, M. Pala<sup>3,†</sup>, P. Dollfus<sup>3</sup>, Y. Lee<sup>4</sup>, and G. Iannaccone<sup>1</sup><sup>1</sup>*Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Via Girolamo Caruso 16, 56126 Pisa, Italy*<sup>2</sup>*School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*<sup>3</sup>*Université Paris-Saclay, Centre National de la Recherche Scientifique, Centre de Nanosciences et de Nanotechnologies, 91120 Palaiseau, France*<sup>4</sup>*Integrated Systems Laboratory, ETH Zürich, 8092 Zürich, Switzerland*

(Received 16 June 2020; revised 16 September 2020; accepted 13 October 2020; published 30 November 2020)

The availability of transistors capable of operating at low supply voltage is essential to improve the key performance metric of computing circuits, i.e., the number of operations per unit energy. In this paper, we propose a device concept for energy-efficient, steep-slope transistors based on heterojunctions of two-dimensional (2D) materials. We show that by injecting electrons from an isolated and weakly dispersive band into a strongly dispersive one, subthermionic subthreshold swings can be obtained, as a result of a cold-source effect and of a reduced thermalization of carriers. This mechanism is implemented by integrating in a metal-oxide-semiconductor field-effect transistor (MOSFET) architecture two different monolayer materials coupled through a van der Waals heterojunction, combining the subthermionic behavior of tunnel field-effect transistors (FETs) with the robustness of a MOSFET architecture against performance-degrading factors, such as traps, band tails, and roughness. A further advantage with respect to tunnel FETs is that only an *n*-type or *p*-type doping is required to fabricate the device. In order to demonstrate the device concept and to discuss the underlying physics and the design options, we study through *ab initio* and full-quantum transport simulations a possible implementation that exploits two recently reported 2D materials.

DOI: [10.1103/PhysRevResearch.2.043286](https://doi.org/10.1103/PhysRevResearch.2.043286)

## I. INTRODUCTION

In recent years, the number of operations per unit energy has emerged as the single most important performance metric for digital systems. This is true both in the domain of large data centers, which are essentially constrained by the total power consumed by the information technology equipment and by the back-room area (cooling plant and switchgear), and in the domain of distributed sensors for the Internet of things, which have extremely reduced energy budgets.

Major improvements in performance per unit energy require a change of paradigm with respect to the current metal-oxide-semiconductor field-effect transistor (MOSFET) technology, in which the supply voltage is intrinsically lower-limited by the impossibility to decrease the subthreshold swing (SS) of transistors below the thermionic limit of 60 mV/decade. This constraint is unavoidably associated with the thermal excitation of carriers within the high-energy Boltzmann tail of the electron distribution.

Reducing the supply voltage is of the utmost importance: In a complementary metal-oxide semiconductor (CMOS) architecture, energy per logic gate switch scales with the square of the supply voltage, if only dynamic power consumption is considered, and superquadratically, if the static power consumption is also taken into account [1].

Steep-slope transistors are a class of devices able to provide subthermionic SS and that therefore can be operated at lower supply voltages. Different physical mechanisms, including band-to-band tunneling (BTBT) [2], negative capacitance effects [3], impact ionization [4], and Mott transition [5], have been explored as bases for the operation of these devices.

In particular, tunnel field-effect transistors (FETs), in which the Boltzmann tail is filtered out via BTBT tunneling, have attracted much attention in the last decade, as devices with potentially sub-40 mV/decade SS, low control voltages, and short delay times [6]. However, despite the encouraging theoretical predictions, the experimental observation of subthermionic SS has proven extremely challenging. The switching process in tunnel FETs relies, indeed, on the modulation of the tunneling probability through a barrier, which is likely to be dominated by disorder-induced nonidealities, as traps [7], band tails [8], and roughness [9].

As a way to circumvent these issues, in this paper we propose an alternative design paradigm that exploits the tailoring opportunities offered by the two-dimensional (2D) materials platform. Within this approach, the subthermionic capabilities of the devices are still driven by an energy-filtering effect, as in tunnel FETs, but the switching is obtained by

\* Also at Université Paris-Saclay, Centre National de la Recherche Scientifique, Centre de Nanosciences et de Nanotechnologies, 91120 Palaiseau, France; logotetad@gmail.com

† marco.pala@c2n.upsaclay.fr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

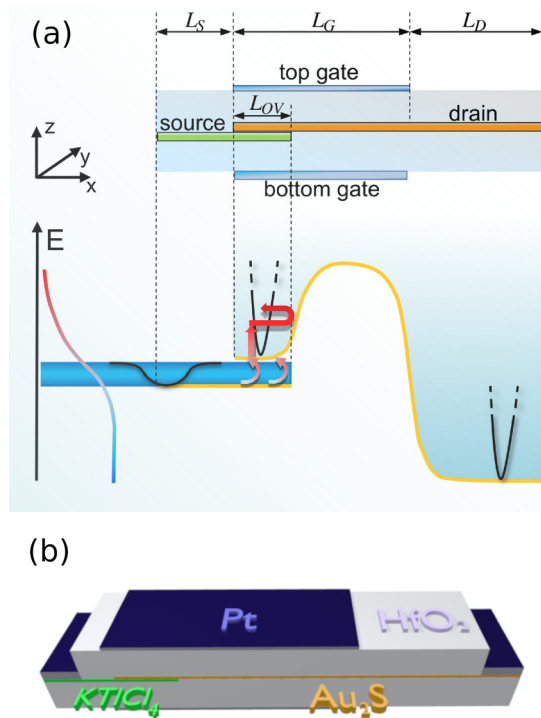


FIG. 1. (a) Sketch of the transistor structure (top) and of the corresponding band diagram (bottom) when the device is not conducting. Electrons are injected in the narrow band of the source monolayer, leaving essentially empty the high-energy Boltzmann tail. Next, electrons tunnel through a van der Waals heterojunction in the strongly dispersive monolayer implementing the channel and the drain, where they undergo a partial thermalization before being reflected back by the barrier (red arrows). (b) Sketch of the representative device studied in this paper, in which the source and channel-drain regions are implemented by monolayers of  $\text{KTiCl}_4$  and  $\text{Au}_2\text{S}$ , respectively.

modulating the height of a thermionic barrier, as in MOSFETs. According to previous studies [10], the operation of these devices is expected to be significantly less sensitive to the performance-degrading factors plaguing the tunnel FETs. As a further advantage, only an  $n$ -type ( $p$ -type) doping is needed for  $n$  ( $p$ ) devices, while a combination of  $n$  type and  $p$  type is always needed in tunnel FETs. We demonstrate this paradigm by numerically investigating the operation and the potential of a representative device, based on two recently reported 2D materials: the ternary  $\text{KTiCl}_4$  monolayer and the  $\text{Au}_2\text{S}$  monolayer in its  $\alpha$  phase.

## II. DEVICE CONCEPT

A sketch of the proposed architecture is given in Fig. 1(a). The device consists of two monolayers, one of which acts as source, while the other implements the channel and the drain. The monolayers overlap over a length  $L_{OV}$ , which establishes a connection between them via a van der Waals heterojunction [11,12]. The current is controlled by a top and a bottom gate of the same length  $L_G$ , separated from the 2D

materials by two insulating layers. The source material and the drain are both  $n$ - or  $p$ -doped, according to the required device polarity. Overall, this architecture differs from that of a monolayer-based double-gate MOSFET only by the presence of a source injector coupled to the channel.

In order to explain the device operation, a sketch of the band diagram of an  $n$ -type transistor in the nonconducting regime is also displayed in Fig. 1(a). The source injector is designed to provide an energy-filtering effect able to cut off the thermal tail of the Boltzmann distribution of carriers (a so-called “cold-source” effect [13–16]). This effect can be obtained by exploiting isolated narrow bands, intrinsically present in the band structure of several 2D materials [17]. When the transport occurs in one of these bands, the electron energy is confined to a small window, and the Boltzmann tail remains essentially empty of carriers [see Fig. 1(a)].

By tunneling through the van der Waals heterojunction, electrons can reach the other monolayer, where they can undergo a partial thermalization driven by phonon absorption. However, if the electron-phonon scattering rates in this material are small enough, only a small percentage of electrons can acquire enough energy to overcome the potential barrier, and most of them will be reflected back [see Fig. 1(a)]. Within this assumption, the device basically works as a traditional FET with a reduced leakage thermionic current in the subthreshold regime and can thus exhibit subthermionic capabilities. Suitable materials to implement the channel-and-drain layer should thus have high phonon-limited mobility, usually associated with very small effective masses (and, therefore, low density of states) [18]. The transport in these materials occurs within strongly dispersive bands, the opposite of what is required in the source injector, where the high effective mass and density of states promote a strong electron-phonon scattering.

A suitable band alignment at the heterojunction is of critical importance to ensure good performance. Ideally, the conduction bands of the monolayers should be close enough or even overlapping, to favor the interlayer tunneling and ensure good current levels in the “on” state. In this respect, we notice that extending the top and bottom gates over the heterojunction, as shown in Fig. 1(a), provides a way to tune the band offset as a function of the difference between the work functions of the metals used for the two gates. Metal gates with different work functions (and held at the same voltage, as is supposed here) induce an electric field that makes the potential on the two monolayers different. If such a field is strong enough, sizable rearrangements of the band alignment can be obtained, as already experimentally demonstrated (see, e.g., Ref. [19] and references quoted therein).

We remark that the tunnel barrier, intrinsically included in the device structure, is not demanded to play any role in the switching process. Accordingly, the impact on the subthreshold performance of defects such as traps and roughness or disorder-induced band tails is expected to be much more limited with respect to tunnel FETs, and similar to that reported for MOSFETs [10]. If a proper design guarantees a sufficient transparency of the tunnel barrier, then the variability induced by such nonidealities, typically responsible for an increase in the tunneling probability, can also be expected to be significantly smaller.

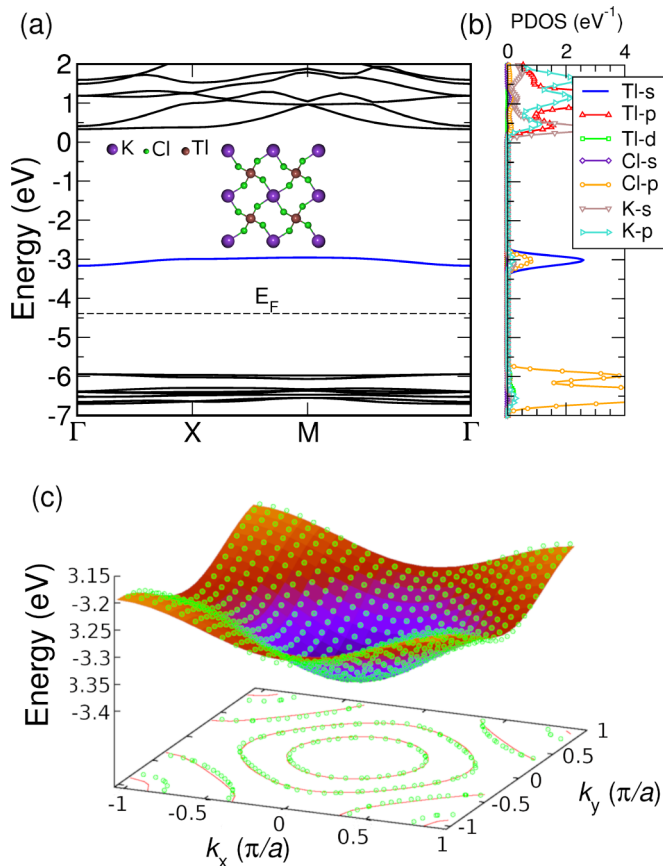


FIG. 2. (a) Band structure of the  $\text{KTiCl}_4$  monolayer. The dashed line denotes the Fermi level. (b) Corresponding density of states projected over the valence atomic orbitals. (c) Plot over the first Brillouin zone of the isolated band highlighted in blue in (a), compared with the approximation obtained within the adopted tight-binding model (green circles).

### III. REPRESENTATIVE DEVICE MODEL

In order to demonstrate the proposed paradigm in a practical case and to discuss in more detail the operation of this class of devices, we have numerically studied a possible implementation using two recently reported 2D materials that exhibit the required properties. The device is sketched in Fig. 1(b).

We use as source material the ternary  $\text{KTiCl}_4$ , reported as a thermodynamically stable monolayer, easily exfoliable from its parent 3D compound [17]. Its band structure is reported in Fig. 2(a), together with the projection of the density of states (PDOS) over the different atomic species [Fig. 2(b)]. The isolated band highlighted in blue, with a minimum in  $\Gamma$  and approximately 0.2 eV wide, is used for the electron transport. This band exhibits a strong contribution from the squared sublattice of Tl atoms and reflects the character of a simple tight-binding band deriving from an s-like orbital [see Fig. 2(b)].

As channel-and-drain material, we choose the synthetic  $\text{Au}_2\text{S}$  in its  $\alpha$  phase. This material exhibits a wide gap of 1 eV, which allows good electrostatic control over the current, and a very small, isotropic electron effective mass of  $0.06m_0$  [20]. Its theoretical phonon-limited mobility is very high and has

been estimated to be  $8.45 \times 10^4 \text{ cm}^2/(\text{V s})$  at room temperature [20].

The transport in the device is addressed by means of full-quantum simulations within the nonequilibrium Green's function formalism. The electron-phonon scattering is included within a deformation potential approximation as detailed in Appendix.

In the transport simulation framework, the  $\text{KTiCl}_4$  monolayer is described within a two-orbital tight-binding model. The Hamiltonian reads

$$H(\mathbf{k}) = \begin{pmatrix} \epsilon_0 + t_k & t_k \\ t_k & -\epsilon_0 + t_k \end{pmatrix},$$

where  $t_k = 2\gamma[\cos(k_x a_x) + \cos(k_y a_y)]$  and  $a_x = a_y = a_{\text{KTiCl}_4}$  is the  $\text{KTiCl}_4$  lattice constant. The lowest band of the model is used to describe the  $\text{KTiCl}_4$  doubly spin-degenerate isolated band of interest. The parameters  $\epsilon_0$  and  $\gamma$  have been set to  $-0.5$  and  $-0.025$  eV, respectively, in order to fit the density functional theory (DFT) band structure [see Fig. 2(c)]. These values also guarantee that the artificial upper band of the tight-binding model is located at energies high enough to be empty of carriers and uninfluential on the transport.

The  $\text{Au}_2\text{S}$  is described within an effective-mass model, which turns out to be a good approximation, due to the large separation ( $\approx 1.5$  eV) of the conduction-band minimum from the above-lying bands.

While the simplicity of the band structures makes it easy to accurately model the two uncoupled materials, the particular values of the incommensurate lattice constants,  $a_{\text{KTiCl}_4} = 6.577 \text{ \AA}$  and  $a_{\text{Au}_2\text{S}} = 5.74 \text{ \AA}$ , hinder a description of the heterojunction directly based on *ab initio* simulations. Such a description is commonly based on finding supercells for the top and bottom layers of approximately the same size and then applying a small amount of strain to make them commensurate [21,22]. Unfortunately, in the case at hand, the minimum size of the supercells compatible with reasonably low levels of strain is  $\sim 7a_{\text{KTiCl}_4} = 4.6 \text{ nm}^2$ , and it is thus too large to allow a full *ab initio* study. We are thus compelled to resort to a different, more approximate strategy. Accordingly, we modeled the coupled monolayers as effectively commensurate over a much smaller cell, of the same size as the  $\text{KTiCl}_4$  unit cell, and we described the interlayer coupling by a spatially constant hopping amplitude  $t_c$ . The value of  $t_c$  should be calibrated in such a way that the average interlayer tunneling probability per unit area is the same as in the real system. However, since the *ab initio* simulation of the system of coupled monolayers is not affordable, the actual value of  $t_c$  cannot be accurately predicted, and the best we can do is to locate it within a reasonable range. In the following, we assume  $t_c = 30 \text{ meV}$ . The choice of this particular value will be discussed at the end of Sec. IV.

In the same spirit, the band offset of the coupled monolayers is estimated by means of Anderson's rule [23]. The electron affinities of  $\text{KTiCl}_4$  and  $\text{Au}_2\text{S}$  are computed to be 4.8 and 4.47 eV, respectively (see Appendix for details), resulting in a gap of  $\sim 0.1$  eV between the conduction bands of the coupled  $\text{KTiCl}_4$ - $\text{Au}_2\text{S}$  system. However, as already mentioned, to enhance the interlayer coupling and prevent a significant degradation of the "on" current, a condition in which the bands overlap is preferable. In order to achieve

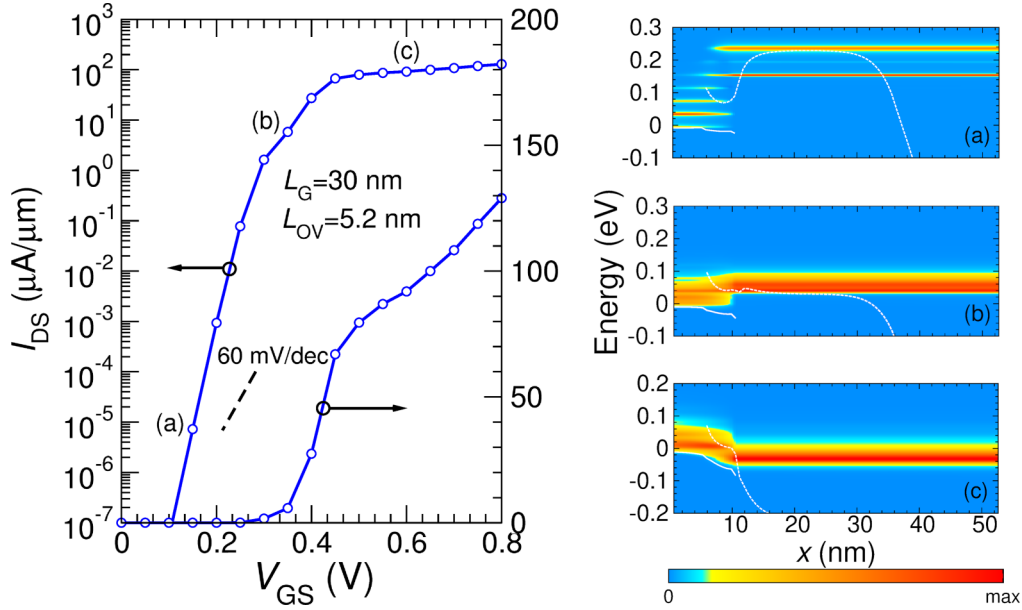


FIG. 3. Left: transfer characteristics of the  $\text{KTiCl}_4\text{-Au}_2\text{S}$  transistor for  $L_G = 30$  nm and  $L_{OV} = 5.2$  nm.  $V_{DS} = 0.4$  V. Right: Maps of the longitudinal current-density spectrum along the device for (a)  $V_{GS} = 0.15$ , (b)  $V_{GS} = 0.35$ , and (c)  $V_{GS} = 0.6$  V. The considered bias points are indicated by the labels “(a)”, “(b)”, and “(c)” on the transfer characteristics. The white solid and dashed curves denote the bottom of the two conduction bands involved in the transport; dec, decade; max, maximum.

this goal, we choose for the top and bottom gates two metals with quite different work functions  $\chi$ . The top gate is made of Pt ( $\chi = 6.3$  eV), while the bottom gate is made of Ag ( $\chi = 4.7$  eV). The vertical electric field arising between the gates downshifts the potential in the  $\text{Au}_2\text{S}$  monolayer with respect to that in the  $\text{KTiCl}_4$  one, enabling an overlap between the bands for all the regimes of operation of the device.

We set the source length of the simulated device to  $L_S \simeq 5$  nm. This value is large enough to guarantee the stabilization of the potential at the left end of the simulation domain. Due to the much smaller density of states in  $\text{Au}_2\text{S}$  with respect to  $\text{KTiCl}_4$ , fulfilling the same condition on the drain end requires a larger value for  $L_D$ , which we set, accordingly, to  $\simeq 25$  nm. We also assume that the monolayers are separated from the top and bottom gates by 3-nm-thick layers of  $\text{HfO}_2$  and that the  $\text{KTiCl}_4$  monolayer and the  $\text{Au}_2\text{S}$  drain are  $n$ -doped with a concentration of  $6 \times 10^{13}$  and  $4 \times 10^{12}$   $\text{cm}^{-2}$ , respectively. The source-to-drain bias ( $V_{DS}$ ) is set to 0.4 V. The “off” state of the transistor is defined by setting the value of the current in this state to  $I_{off} = 10^{-4}$   $\mu\text{A}/\mu\text{m}$  [24], while the “on” state is obtained, as usual, at the gate voltage  $V_{GS} = V_{GS}^{off} + V_{DS}$ , where  $V_{GS}^{off}$  is the gate voltage in the “off” state.

The parasitic resistances associated with the source and drain contacts are neglected in our model. They are expected to mainly degrade the current close to and above the conduction threshold, leaving essentially unaltered the subthreshold current, due to negligible voltage drop across contacts. We remark that although the quest for an approach to obtain low-resistance contacts compatible with large-scale manufacturability is still ongoing, several strategies have already been proposed and demonstrated, including one-dimensional edge contacts [25], van der Waals contacts [26], and surface roughness engineering [27].

#### IV. RESULTS AND DISCUSSION

Figure 3 illustrates the transfer characteristics of the device for  $L_{OV} = 5.2$  and  $L_G = 30$  nm. The results confirm the subthermionic capabilities of the transistor, which is able to achieve a minimum subthreshold swing of 26 mV/decade. The “on” current amounts to  $I_{on} \approx 90$   $\mu\text{A}/\mu\text{m}$ , corresponding to an  $I_{on}$ -to- $I_{off}$  ratio of around  $10^6$ .

In order to gain more insights into the operation of the device, the spectrum of the longitudinal current for three different values of  $V_{GS}$  is also reported in Fig. 3. For  $V_{GS}$  close to the “off” state [Fig. 3(a)], electrons are mostly injected from the source contact close to the bottom of the  $\text{KTiCl}_4$  conduction band, and the strong phonon absorption causes them to rise at energies above the top of the barrier. This is in sharp contrast to the behavior usually observed in traditional MOSFETs, in which the injection of carriers directly occurs at energies closer to the top of the channel barrier. Significant spectral components at energies under the top of the barrier, related to source-to-drain tunneling processes, are also present. Both the strength of the electron-phonon interaction and the tunneling probability through the channel barrier are therefore important to determine the leakage current in the “off” state.

As  $V_{GS}$  increases and the channel barrier is pushed down [Fig. 3(b)], optical-phonon absorption gradually decreases and concentrates at the right edge of the heterojunction, where the electrons propagating in  $\text{KTiCl}_4$  face the monolayer boundary. This trend is expressed by a steplike feature in the current spectrum near  $x = 10$  nm. An analogous behavior can be observed close to the “on” state [Fig. 3(c)], where, however, the transport is dominated by phonon emission processes and the sign of the step is reversed. All the maps indicate that the electron thermalization outside the source

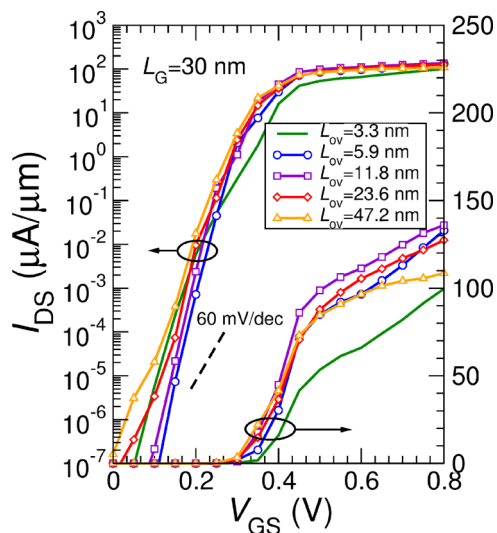


FIG. 4. Transfer characteristics of the transistor for  $L_G = 30$  nm and different values of  $L_{OV}$ .  $V_{DS} = 0.4$  V.

and the overlap region is negligible over the considered scale lengths.

Next, we consider the impact of the length of the overlap region on the device operation. Figure 4 reports the transfer characteristics obtained for different values of  $L_{OV}$ , while the interlayer current spectrum close to the “off” and “on” states for  $L_{OV} = 23.6$  nm is reported in Figs. 5(a) and 5(b), respectively.

Figure 5(a) shows that in the “off” state, the interlayer tunneling mostly occurs in narrow energy windows, corresponding to the thermionic and tunneling peaks of the longitudinal current. Spatially, the interlayer current follows an oscillating behavior, associated with the presence of quantum-well-like states. They result from the confinement enforced in the  $Au_2S$  by the ending of the monolayer on the source side and by the channel barrier on the other side. The weak degradation trend of SS as  $L_{OV}$  increases, visible in Fig. 4, reflects a modest enhancement of the electron thermal-

ization as the overlap region is enlarged, due to the increase in the number of electron-phonon scattering events.

In the “on” state [Fig. 5(b)], the interlayer tunneling principally occurs in the energy window between the bottom of the two bands and is mostly localized at the edges of the heterojunction. The spatial distribution of the interlayer current for different values of  $L_{OV}$  is plotted in Fig. 5(c). It can be noticed that as  $L_{OV}$  decreases, the current concentrates more on the right edge of the heterojunction. Actually, the tunneling probability is particularly high at this edge, and for  $L_{OV}$  that is not too small ( $\geq 5.9$  nm in Fig. 4), the increase in the current density at this point approximately compensates its decrease with  $L_{OV}$  in the rest of the overlap region. This explains the weak dependence of the current at large gate overdrives on  $L_{OV}$ , noticeable in Fig. 4. Overall, the device appears weakly sensitive to  $L_{OV}$  in any operation regime. This represents a relevant property, as  $L_{OV}$  is a design parameter that can be difficult to finely control in practical manufacturing processes.

To assess the scaling behavior of the device, we report in Fig. 6(a) the transfer characteristics for  $L_{OV} = 5.2$  nm and different values of  $L_G$ . The minimum value of SS over the  $V_{GS}$  window of interest and the  $I_{on}$ -to- $I_{off}$  ratio are plotted as a function of  $L_G$  in Fig. 6(b). Scaling down the gate length only influences the subthreshold regime and, particularly, degrades the subthreshold swing. To achieve subthermionic behavior, the gate length must be larger than 15 nm, while values above 30 nm improve only negligibly the performance of the device. The SS value of  $\approx 23$  mV/decade, obtained for  $L_G = 55$  nm, represents a lower limit for this device, set by the intrinsic phonon absorption. The pronounced dependence of the current on the gate length is due to the small effective mass of  $Au_2S$ , which results in a rapid increase in the source-to-drain tunneling probability as the gate length is reduced.

In order to assess the role played by the strength of the electron–optical-phonon coupling, we also investigated the limit of vanishing interactions between electrons and optical phonons in the case  $L_G = 55$  nm. The corresponding transfer characteristics [dashed curves in Fig. 6(a)] are

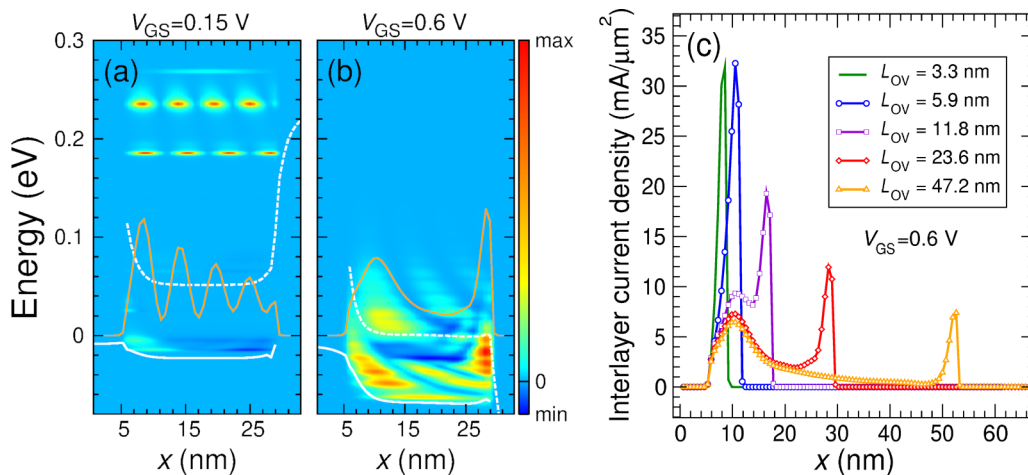


FIG. 5. Maps of the interlayer current spectrum for  $L_{OV} = 23.6$  nm and (a)  $V_{GS} = 0.15$  V or (b)  $V_{GS} = 0.6$  V. The orange curves superimposed on the maps are the plot (in arbitrary units) of the interlayer current as a function of the longitudinal position. The white curves are defined as in Fig. 3; min, minimum. (c) Interlayer current along the device at  $V_{GS} = 0.6$  V for different values of  $L_{OV}$ .

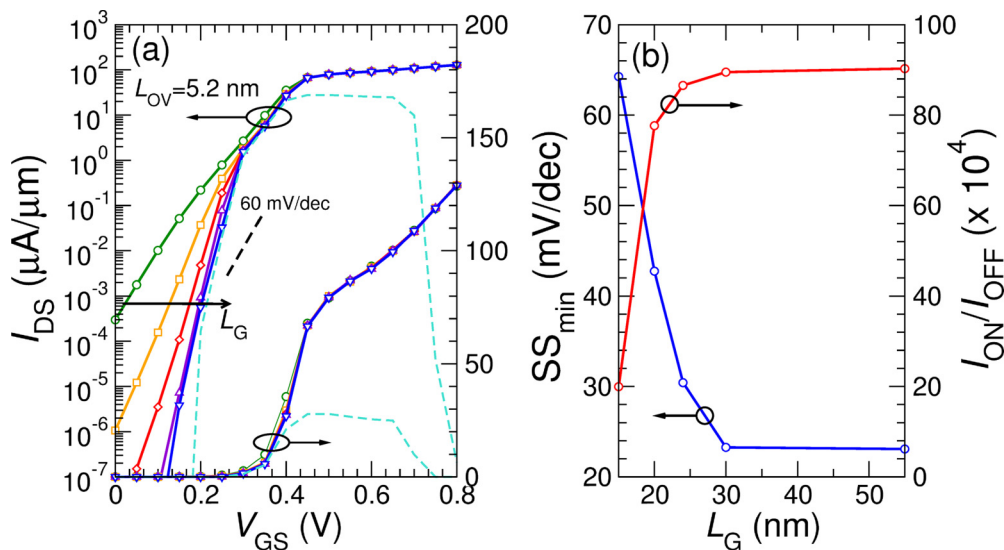


FIG. 6. (a) Transfer characteristics of the transistor for  $L_{OV} = 5.2$  nm and  $L_G = 15, 20, 25, 30,$  and  $55$  nm.  $V_{DS} = 0.4$  V. The dashed curves are obtained by taking into account only the elastic interaction between electrons and acoustic phonons. (b) Minimum of SS for  $I_{DS} \geq I_{off} = 10^{-4} \mu A/\mu m$  and  $I_{on-to-Ioff}$  ratio as a function of  $L_G$ .

obtained by only including in the simulation the interactions between electrons and elastic acoustic phonons. According to the previous analysis, lower values of SS are achieved, particularly for  $V_{GS} \lesssim 0.2$  V, namely, when the top of the channel barrier is above the highest energy of the  $KTiCl_4$  conduction band. Indeed, in these conditions the electrons injected by the source contact necessarily need to increase their energy by absorbing optical phonons in order to overcome the barrier. On the other hand, we notice that the current at large gate overdrives significantly decreases. The reason is that optical phonons enable further connections between high-density-of-state regions, therefore activating new high-conductive paths. The consequent enhancement of the

interlayer tunneling eventually dominates over the increase in the backscattering. The transfer characteristics with and without electron–optical-phonon interactions almost overlap for intermediate  $V_{GS}$  values, for which the energy redistribution of electrons driven by optical phonons is weaker [cf. map in Fig. 3(b)]. The sudden drop of the current for  $V_{GS} > 0.65$  V is an unrealistic artifact of the simulation, due to the strong backscattering of electrons elastically traveling in  $KTiCl_4$  from the down-bending potential profile.

As the dashed curve in Fig. 6(a) suggests, in general our model predicts that sub-1 mV/decade SSs could be in principle attained by using channel materials with small enough electron-phonon scattering rates and large enough band gaps

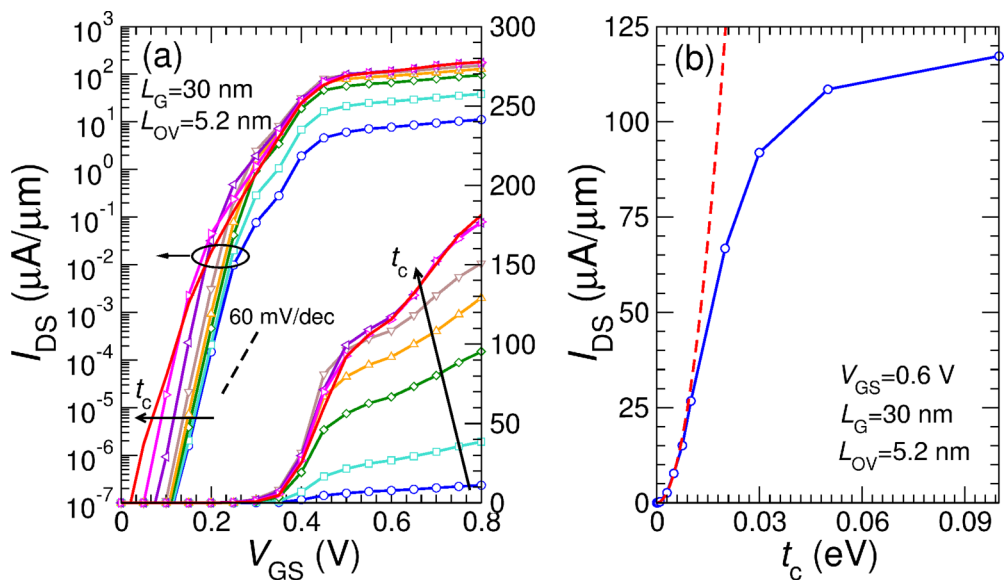


FIG. 7. (a) Transfer characteristics of the transistor for  $L_G = 30$  nm,  $L_{OV} = 5.2$  nm, and effective hopping values  $t_c = 5, 10, 20, 30, 50, 100, 150,$  and  $200$  meV.  $V_{DS} = 0.4$  V. (b)  $I_{DS}$  at  $V_{GS} = 0.6$  V as a function of  $t_c$ . The dashed curve is a parabolic fit for small values of  $t_c$ .

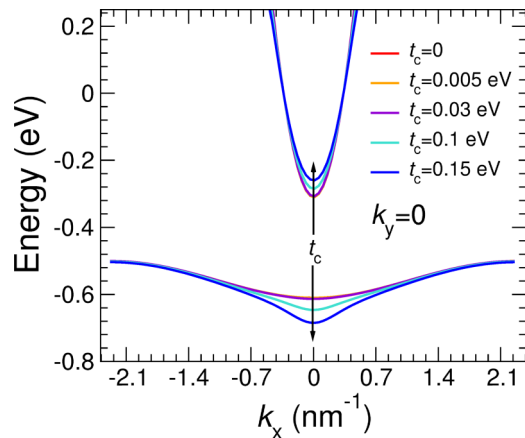


FIG. 8. Cross section at  $k_y = 0$  of the band structure of the  $\text{KTlCl}_4\text{-Au}_2\text{S}$  heterostructure for different values of  $t_c$ .

(to avoid interband tunneling between conduction and valence bands) and by properly extending the gate length in order to suppress the tunneling through the channel barrier. A favorable band alignment of the monolayers is furthermore needed to locate the smallest values of the SS in the  $I_{\text{DS}}$  window of interest for the operation of the device as a digital switch.

Finally, we explore the dependence of the device operation on the value of  $t_c$ . We report in Fig. 7(a) the transfer characteristics of the device for  $L_{\text{OV}} = 5.2$  nm,  $L_G = 30$  nm, and different values of  $t_c$ . Different choices of  $t_c$  mainly impact the current at large gate overdrives, with a behavior that can be better understood from Fig. 7(b), where the current at  $V_{\text{GS}} = 0.6$  V is plotted as a function of  $t_c$ . For small values of the coupling, the current increases quadratically with  $t_c$ , according to a first-order approximation of the interlayer current [28],  $I_{\text{int}} \propto t_c^2 \times D_{\text{T}} \times D_{\text{B}}$ , where  $D_{\text{T}}$  and  $D_{\text{B}}$  are the density of states on the top and bottom layers, respectively. For  $t_c > 100$  meV, the current saturates, indicating that it is not limited anymore by  $t_c$  itself, but by the low density of states of the  $\text{Au}_2\text{S}$ . This almost optimal regime is not far from that assumed in this paper by setting  $t_c = 30$  meV. Our results should thus be considered close to an upper limit of the performance potentially attainable by the device.

The value of  $t_c$  also modulates the band alignment between the monolayers, as shown in Fig. 8. The increasing separation of the bands as  $t_c$  increases results in the gradual shift of the transfer characteristics toward more negative  $V_{\text{GS}}$  visible in Fig. 7(a). On the other hand, it can be seen that the value of 30 meV for  $t_c$  is small enough to entail only slight modifications of the band structure of the coupled monolayers. It is thus compatible with weak interactions of van der Waals type.

## V. CONCLUSION

We have proposed a device concept for energy-efficient steep-slope FETs based on the combination of 2D materials with a weakly dispersive and a strongly dispersive band, interconnected through a van der Waals heterojunction. The low-dispersive material acts as the device source and provides

an intrinsic cold-source energy-filtering effect on the electrons injected by the contact. The electron thermalization is to a large extent prevented by the low electron-phonon coupling in the strongly dispersive material implementing the channel and the drain of the device. Despite the presence of two materials, the device architecture reflects that of a standard MOSFET, in which the current is modulated by a thermionic barrier, instead of a tunnel barrier, as in tunnel FETs. Thus, according to previous studies, the proposed architecture is expected to be significantly less sensitive to the nonidealities that strongly degrade the performance of tunnel FETs, such as traps, band tails, and roughness. In order to demonstrate the device concept, we studied the operation and the performance of a device based on a  $\text{KTlCl}_4$  and an  $\alpha\text{-Au}_2\text{S}$  monolayer. For a gate length of 30 nm and an equivalent gate oxide thickness of about 0.5 nm, this device can attain SSs close to 20 mV/decade and a current modulation of six orders of magnitude. On the other hand, it shows little sensitivity to the size of the overlap between the monolayers, a hardly controllable parameter in manufacturing processes. The scaling properties of the device are dominated by the quite rapid increase in the source-to-drain tunneling as the gate length is decreased, favored by the smallness of the  $\text{Au}_2\text{S}$  effective mass. Finally, we remark that the occurrence of weakly dispersive, isolated bands around the Fermi energy is not uncommon in already reported 2D materials. We mention, for instance, the case of several monolayers that are ferromagnetic or antiferromagnetic in their ground state, such as  $\text{CrBr}_2$ ,  $\text{CrI}_2$ ,  $\text{MnBr}_2$ ,  $\text{MnCl}_2$ ,  $\text{NiBr}_2$ ,  $\text{NiCl}_2$ ,  $\text{VOBr}_2$ , and  $\text{VOCl}_2$  [17]. Also, several 2D materials with remarkably small effective mass of electrons and holes, and sizable band gap, such as  $\alpha\text{-AsSb}$  [29] and  $\text{GeSe}$  [30], have been proposed. Together with the conjectured abundance and variety of thermodynamically stable monolayers still to be explored [17], these considerations suggest that many opportunities to implement and optimize the proposed architecture may be found.

## ACKNOWLEDGMENTS

This work is partially supported by the European Commission through the Horizon 2020 FET Open project QUEFORMAL (Contract No. 829035) and by the Italian MIUR through the PRIN project FIVE2D (Contract No. 2017SRYEJH\_001). J.C. acknowledges partial support from the Natural Science Foundation of Jiangsu Province under Grant No. BK20180456 and from the Key Laboratory for Information Science of Electromagnetic Waves (MoE) under Grant No. EMW201906.

## APPENDIX: COMPUTATIONAL METHODS

The DFT simulations were performed in a plane-wave basis by means of the QUANTUM ESPRESSO suite [31,32]. The  $\text{KTlCl}_4$  monolayer was simulated by using scalar relativistic ultrasoft pseudopotentials based on the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional, while the  $\text{Au}_2\text{S}$  monolayer was simulated by using fully relativistic norm-conserving pseudopotentials with a Heyd-Scuseria-Ernzerhof hybrid functional. In both cases, a  $12 \times 12 \times 1$  Monkhorst-Pack (M-P) k-point grid was used, with a

cutoff energy of 50 Ry. The van der Waals interactions were taken into account through DFT-D3 Grimme's corrections [33].

Electron affinities were computed according to the following definition: the difference between the energy in the vacuum and the energy of the lowest unoccupied energy level in the materials.

The dielectric permittivity tensors were computed within the framework of the density functional perturbation theory (DFPT) as implemented in the PH.X code (included in the QUANTUM ESPRESSO package).  $12 \times 12 \times 1$  and  $3 \times 3 \times 1$  M-P wave-vector grids were used for electrons and phonons, respectively, and the energy cutoff was set at 60 Ry. The results were then interpolated on finer  $80 \times 80 \times 1$  grids by means of the EPW code [34]. In order to extract from the simulation results, which include the effect of the vacuum layer needed to avoid interactions between the supercell replicas, the value of the dielectric tensor associated with the monolayers only, we resorted to the principle of the series and parallel capacitances, as described in Ref. [35]. For  $\text{KTiCl}_4$ , we found  $\epsilon_{\parallel} = 2.5$  and  $\epsilon_{\perp} = 2.9$ , for the in-plane and out-of-plane dielectric permittivity, respectively. For  $\text{Au}_2\text{S}$ , we found  $\epsilon_{\parallel} = 6.0$  and  $\epsilon_{\perp} = 9.1$ .

The deformation potential of the dominant acoustic- and optical-phonon branches in  $\text{KTiCl}_4$  was computed from the corresponding matrix elements. We found 0.25 eV and  $2 \times 10^8$  eV/cm for the acoustic and the optical deformation po-

tential, respectively. In the case of  $\text{Au}_2\text{S}$ , we used the acoustic deformation potential associated with the longitudinal modes provided in Ref. [20]. The interaction of electrons and polar optical phonons in  $\text{Au}_2\text{S}$  was modeled through an effective deformation potential of  $1.6 \times 10^8$  eV/cm, computed by fitting the scattering rate of the electrons with the dominant polar optical-phonon mode. The interactions with flexural optical phonons are neglected, as they are expected to be strongly suppressed by the embodiment of the monolayers into the gate stacks. Moreover, in the case of  $\text{Au}_2\text{S}$ , the coupling of electrons with flexural phonons is zero to the first order, due to the existence of a glide symmetry plane [36]. Both the scattering rates and the electron-phonon matrix elements were computed within the framework of the DFPT, by means of the codes PH.X and EPW.

To the purpose of the simulation of transport, the device was modeled as translationally invariant in the transverse  $y$  direction. Transport simulations were performed within the framework of the nonequilibrium Green's function formalism, by self-consistently solving the transport equations and the Poisson equation in the device cross section. A set of 17 transverse wave vectors was used to sample the Brillouin zone of  $\text{KTiCl}_4$ , of  $\text{Au}_2\text{S}$ , and of the  $\text{KTiCl}_4$ - $\text{Au}_2\text{S}$  heterostructure. The electron-phonon interactions were described within the self-consistent Born approximation, by assuming an elastic and dispersionless approximation for acoustic and optical phonons, respectively.

- 
- [1] T. N. Theis and P. M. Solomon, In quest of the next switch: Prospects for greatly reduced power dissipation in a successor to the silicon field-effect transistor, *Proc. IEEE* **98**, 2005 (2010).
- [2] A. C. Seabaugh and Q. Zhang, Low-voltage tunnel transistors for beyond CMOS logic, *Proc. IEEE* **98**, 2095 (2010).
- [3] S. Salahuddin and S. Datta, Use of negative capacitance to provide voltage amplification for low power nanoscale devices, *Nano Lett.* **8**, 405 (2008).
- [4] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer, Impact ionization MOS (I-MOS)—Part I: Device and circuit simulations, *IEEE Trans. Electron Devices* **52**, 69 (2005).
- [5] M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa, and Y. Tokura, Collective bulk carrier delocalization driven by electrostatic surface charge accumulation, *Nature (London)* **487**, 459 (2012).
- [6] H. Lu and A. Seabaugh, Tunnel field-effect transistors: State-of-the-art, *IEEE J. Electron Devices Soc.* **2**, 44 (2014).
- [7] M. Pala and D. Esseni, Interface traps in InAs nanowire tunnel-FETs and MOSFETs—Part I: Model description and single trap analysis in tunnel-FETs, *IEEE Trans. Electron Devices* **60**, 2795 (2013).
- [8] M. A. Khayer and R. K. Lake, Effects of band-tails on the subthreshold characteristics of nanowire band-to-band tunneling transistors, *J. Appl. Phys.* **110**, 074508 (2011).
- [9] F. Conzatti, M. Pala, and D. Esseni, Surface-roughness-induced variability in nanowire InAs tunnel FETs, *IEEE Electron Device Lett.* **33**, 806 (2012).
- [10] D. Esseni and M. Pala, Interface traps in InAs nanowire tunnel FETs and MOSFETs—Part II: Comparative analysis and trap-induced variability, *IEEE Trans. Electron Devices* **60**, 2802 (2013).
- [11] J. Na, Y. Kim, J. H. Smet, M. Burghard, and K. Kern, Gate-tunable tunneling transistor based on a thin black phosphorus-SnSe<sub>2</sub> heterostructure, *ACS Appl. Mater. Interfaces* **11**, 20973 (2019).
- [12] G. Iannaccone, F. Bonaccorso, L. Colombo, and G. Fiori, Quantum engineering of transistors based on 2D materials heterostructures, *Nat. Nanotechnol.* **13**, 183 (2018).
- [13] C. Qiu, F. Liu, L. Xu, B. Deng, M. Xiao, J. Si, L. Lin, Z. Zhang, J. Wang, H. Guo, H. Peng, and L.-M. Peng, Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches, *Science* **361**, 387 (2018).
- [14] D. Logoteta, M. G. Pala, J. Choukroun, P. Dollfus, and G. Iannaccone, A steep-slope MoS<sub>2</sub>-nanoribbon MOSFET based on an intrinsic cold-contact effect, *IEEE Electron Device Lett.* **40**, 1550 (2019).
- [15] E. G. Marin, D. Marian, M. Perucchini, G. Fiori, and G. Iannaccone, Lateral heterostructure field-effect transistors based on two-dimensional material stacks with varying thickness and energy filtering source, *ACS Nano* **14**, 1982 (2020).
- [16] F. Liu, Switching at Less Than 60 mV/Decade with a 'Cold' Metal as the Injection Source, *Phys. Rev. Appl.* **13**, 064037 (2020).
- [17] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, and N. Marzari, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds, *Nat. Nanotechnol.* **13**, 246 (2018).

- [18] M. S. Lundstrom, *Fundamentals of Carrier Transport* (Addison-Wesley, Reading, MA, 1990).
- [19] Y. Liu, N. O. Weiss, X. Duan, H. C. Cheng, Y. Huang, and X. Duan, Van der Waals heterostructures and devices, *Nat. Rev. Mater.* **1**, 16042 (2016).
- [20] Q. Wu, W. W. Xu, D. Lin, J. Wang, and X. C. Zeng, Two-dimensional gold sulfide monolayers with direct band gap and ultrahigh electron mobility, *J. Phys. Chem. Lett.* **10**, 3773 (2019).
- [21] A. Szabo, S. Koester, and M. Luisier, *Ab-initio* simulation of van der Waals MoTe<sub>2</sub>-SnS<sub>2</sub> heterotunneling FETs for low-power electronics, *IEEE Electron Device Lett.* **36**, 514 (2015).
- [22] J. Cao, D. Logoteta, S. Özkaya, B. Biel, A. Cresti, M. G. Pala, and D. Esseni, Operation and design of van der Waals tunnel transistors: A 3-D quantum transport study, *IEEE Trans. Electron Devices* **63**, 4388 (2016).
- [23] R. L. Anderson, Germanium-gallium arsenide heterojunctions [Letter to the Editor], *IBM J. Res. Dev.* **4**, 283 (1960).
- [24] International Roadmap for Devices and Systems, IRDS (2017), More Moore Tables, <https://irds.ieee.org/roadmap-2017>.
- [25] L. Wang, I. Meric, P. Y. Huang, Q. Gao, Y. Gao, H. Tran, T. Taniguchi, K. Watanabe, L. M. Campos, D. A. Muller, J. Guo, P. Kim, J. Hone, K. L. Shepard, and C. R. Dean, One-dimensional electrical contact to a two-dimensional material, *Science* **342**, 614 (2013).
- [26] Y. Wang, J. C. Kim, R. J. Wu, J. Martinez, X. Song, J. Yang, F. Zhao, A. Mkhoyan, H. Y. Jeong, and M. Chhowalla, Van der Waals contacts between three-dimensional metals and two-dimensional semiconductors, *Nature (London)* **568**, 70 (2019).
- [27] S. Banerjee, L. Cao, Y. S. Ang, L. K. Ang, and P. Zhang, Reducing Contact Resistance in Two-Dimensional-Material-Based Electrical Contacts by Roughness Engineering, *Phys. Rev. Appl.* **13**, 064021 (2020).
- [28] C. Delerue and M. Lannoo, *Nanostructures: Theory and Modeling*, 1st ed. (Springer-Verlag, Berlin, 2004).
- [29] P. Zhao, J. Li, W. Wei, Q. Sun, H. Jin, B. Huang, and Y. Dai, Giant anisotropic photogalvanic effect in a flexible AsSb monolayer with ultrahigh carrier mobility, *Phys. Chem. Chem. Phys.* **19**, 27233 (2017).
- [30] Y. Hu, S. Zhang, S. Sun, M. Xie, B. Cai, and H. Zeng, GeSe monolayer semiconductor with tunable direct band gap and small carrier effective mass, *Appl. Phys. Lett.* **107**, 122107 (2015).
- [31] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [32] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo *et al.*, Advanced capabilities for materials modelling with quantum ESPRESSO, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [33] S. Grimme, J. Antony, S. Ehrlich, and 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**, 154104 (2010).
- [34] S. Poncé, E. Margine, C. Verdi, and F. Giustino, EPW: Electron-phonon coupling, transport and superconducting properties using maximally localized Wannier functions, *Comput. Phys. Commun.* **209**, 116 (2016).
- [35] A. Laturia, M. L. Van de Put, and W. G. Vandenberghe, Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: From monolayer to bulk, *npj 2D Mater. Appl.* **2**, 6 (2018).
- [36] M. V. Fischetti and W. G. Vandenberghe, Mermin-Wagner theorem, flexural modes, and degraded carrier mobility in two-dimensional crystals with broken horizontal mirror symmetry, *Phys. Rev. B* **93**, 155413 (2016).